【A6】Delphiテクニカルセッション

EMBARCADERO TECHNOLOGIES

DEVELOPER CAMP

Delphi 2009ではじめるUnicodeアプリケーション - 既存コード移行のポイント -

有限会社 エイブル
富永 英明

アジェンダ

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- Unicode の基礎知識
- 旧プロダクトからの移行 – Delphi 2007 へ –
- Delphi 2009 と Unicode
- 旧プロダクトからの移行 – Unicode アプリケーション –
- 資料

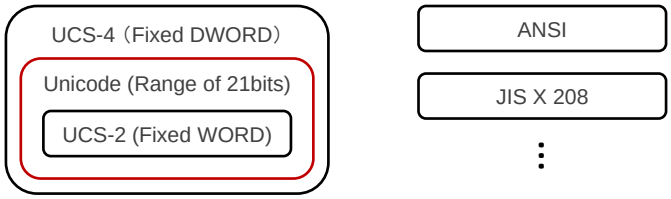
本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

EMBARCADERO TECHNOLOGIES

DEVELOPER CAMP

Unicode の基礎知識

- 文字セット

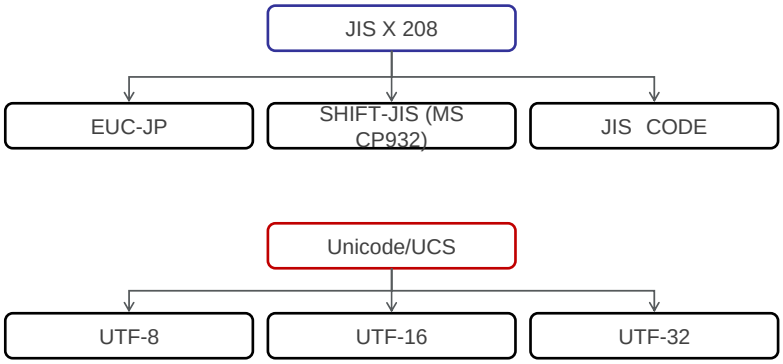


本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

Unicode の基礎知識

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- 文字エンコーディング

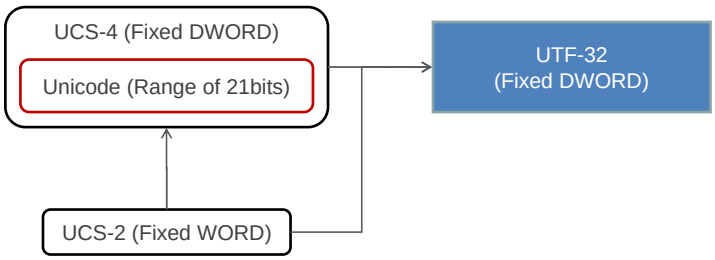


本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

Unicode の基礎知識

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- UTF-32

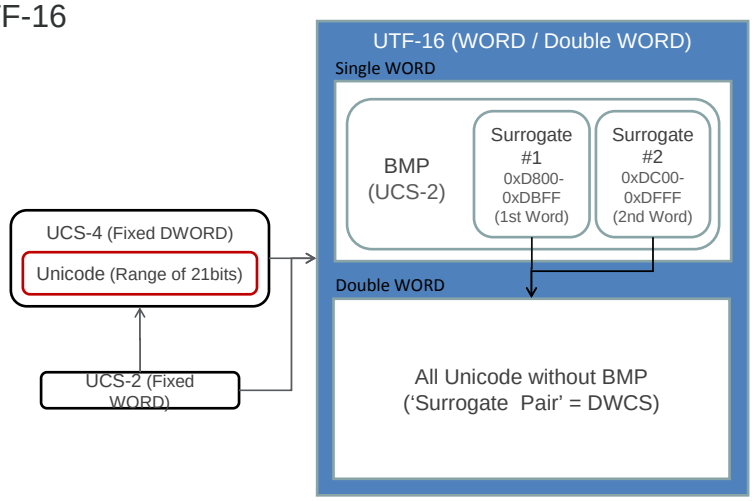


本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

Unicode の基礎知識

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- UTF-16

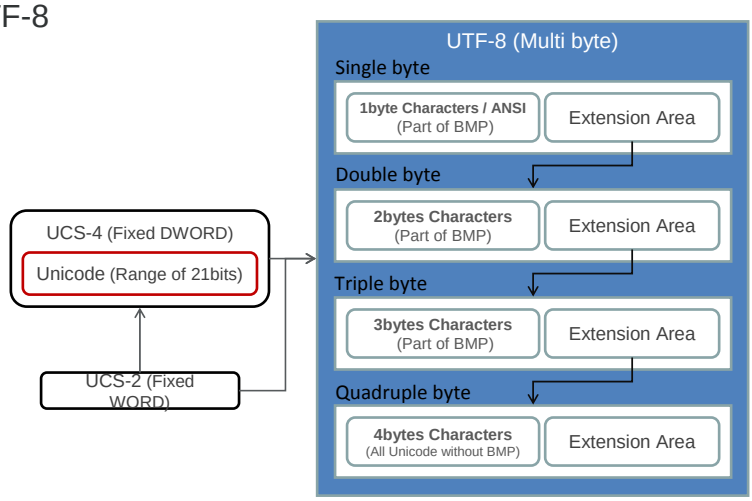


本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

Unicode の基礎知識

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- UTF-8



本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

Unicode の基礎知識

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- プレーン

Unicode

Plane #0 (BMP=UCS-2) U+0000-U+FFFF			
Plane #1 (SMP) U+10000-U+1FFFF	**Plane #2 (SIP)** U+20000-U+2FFFF	Plane #3 (reserved) U+30000-U+3FFFF	Plane #4 (reserved) U+40000-U+4FFFF
Plane #5 (reserved) U+50000-U+5FFFF	Plane #6 (reserved) U+60000-U+6FFFF	Plane #7 (reserved) U+70000-U+7FFFF	Plane #8 (reserved) U+80000-U+8FFFF
Plane #9 (reserved) U+90000-U+9FFFF	Plane #10 (reserved) U+A0000-U+AFFFF	Plane #11 (reserved) U+B0000-U+BFFFF	Plane #12 (reserved) U+C0000-U+CFFFF
Plane #13 (reserved) U+D0000-U+DFFFF	**Plane #14 (SSP)** U+E0000-U+EFFFF	**Plane #15 (PUP #A)** U+F0000-U+FFFFF	**Plane #16 (PUP #B)** U+100000-U+10FFFF

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

Unicode の基礎知識

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- サロゲートペア
  - UTF-16 で U+10000 以降のコードポイントの文字は 2ワード で表す。
  - SHIFT-JIS のように、第 2 文字構成要素が 1 ワード文字の範囲と重複する事はないため、文字検索やデリミタの判別は容易となる。
  - 第 1 文字構成要素と第 2 文字構成要素の範囲も重複しない。
  - UTF-8 / UTF-32 にサロゲートペアの概念はないが、Windows の Unicode は UTF-16 なので、サロゲートペアを考慮せずに変換した場合に影響が出る。例えば、UTF-8 の 4バイト文字は U+10000 以降のコードポイントの文字を表すので、4バイト文字が不正になってしまう。
  - サロゲートペアを考慮しないと、Unicode 1.0 (UCS2) アプリケーションしか作れない。
  - サロゲートペアで表される文字は コードポイントからすれば、Unicode全体の 16/17 となる。
  - 詳細は http://dn.codegear.com/jp/article/38811 で。

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

Unicode の基礎知識

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- 結合文字列
  - “見た目の1文字” を複数のコードポイントで表すのが結合文字列。
  - 結合文字列は “基底文字” + “結合文字” の並び。
  - 例えば、基底文字 “か” と 結合文字“ ゛” の 組み合わせで ”が” という結合文字列となる。
  - これとは別に、”が” という単独のコードポイントの文字も存在する。これは合成文字とよばれる。
  - 結合文字列 から 合成文字 へ変換する事を “合成”、逆に合成文字 から 結合文字列 へ変換する事を “分解” という。
  - 10を超えるコードポイントで構成される結合文字列が存在する。
  - 結合文字列の ”が” と 合成文字の ”が” は見た目では判断できない。
  - 文字列検索の際には、結合文字列 と 合成文字 を同一視しなくてはならない場合があり、前処理として “正規化(Normalize)” を行う必要がある。
  - 詳細は http://dn.codegear.com/jp/article/38816 で。

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

Unicode の基礎知識

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- その他の注意点
  - エンコードによってはファイルの先頭に BOM(バイト・オーダー・マーク) と呼ばれるバイトオーダを調べるためのマーカが付加される事がある。
  - U+FEFF (“ZERO WIDTH NO-BREAK SPACE”) のように、幅0 の文字が存在する。
  - Unicode では 文字の半角/全角サイズを文字構成要素数では判断できない。

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

Unicode の基礎知識

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- Windows と Unicode
  - Win9x の Unicode は UCS2
  - NT系の Windows の Unicode は UTF-16
  - Vista では 何も考えずに サロゲートペアを扱えるが、2000 / XP では多少の環境設定が必要となる。
  - “JIS X 0213:2004” の範囲の Unicode しか扱わないのであれば、フォントに “メイリオ” または IPAフォント(http://ossipedia.ipa.go.jp/ipafont/) が使える。
  - 単一のフォントで収録された文字が多いのは “ MingLiU ” または “SimSun” 。
  - Office 2000 以降を持っているのであれば、“Arial Unicode MS” が使える。これは Unicode 2.1 規格のすべての文字を網羅する。
  - Vista と “JIS X 0213:2004” についての 資料は “JIS X 0213:2004 / Unicode 実装ガイド (Microsoft)” http://go.microsoft.com/fwlink/?LinkId=78800 がある。

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

旧プロダクトからの移行
– Delphi 2007 へ -

旧プロダクトからの移行 – Delphi 2007 へ –

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- Delphi 1
  - → Delphi 7 または Delphi 2007 へ
  - 32bit 化
- Delphi 2 ～ Delphi 5
  - → Delphi 7 へ
  - コンポーネントの修正 (dsgnintf → DesignIntf)
  - ネットワークコンポーネントの置換 (NetManage)
- Delphi 6 ～ Delphi 7
  - → Delphi 2007 へ
  - CLXコードの排除
- Delphi 2005 ～ Delphi 2006 (Turbo Delphi 含む)
  - → Delphi 2007 へ
  - XP / Vista 対応 (XPMan / VistaAltFix / FastMM の削除)

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

旧プロダクトからの移行 – Delphi 2007 へ –

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- DsgnIntf (～ Delphi5)
  - Dsgnintf を uses しているコンポーネントは、DesignIntf / DesignEditors を uses するように変更しなくてはならない。

```
// [旧]
uses
  ..., DsgnIntf,...

// [新]
uses
  ..., DesignIntf, DesignEditors, VCLEditors, RTLConsts, ...
```

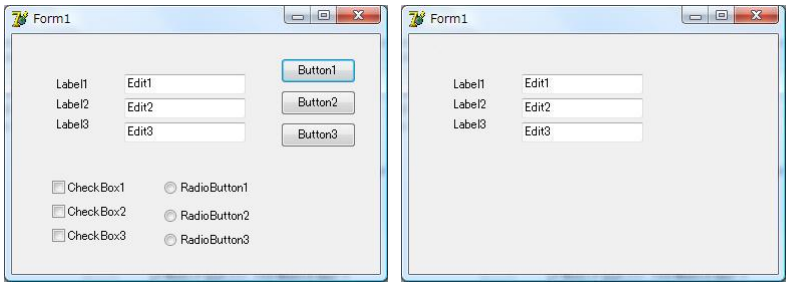
- FastMM (Option)
  - BDS2006 で採用された新しいメモリマネージャと同等のメモリマネージャ
  - http://sourceforge.net/project/showfiles.php?group_id=130631 からDL可能。
  - 詳細は以下のURLで
    http://dn.codegear.com/jp/article/33624
    http://dn.codegear.com/jp/article/33696

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

旧プロダクトからの移行 – Delphi 2007 へ –

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- Vista Alt Fix (～ BDS 2006)



Delphi 2007 以前のプロダクトでは、テーマに対応させたアプリケーションをVistaで動作させた場合、単にAltキーを押下するだけで、コントロールの一部が正しく描画されなくなってしまう。

これを回避するのが、”VistaAltFix (http://cc.codegear.com/item/24282)”。

※Delphi 2007 / Delphi 2009 には不要。

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

旧プロダクトからの移行 – Delphi 2007 へ –

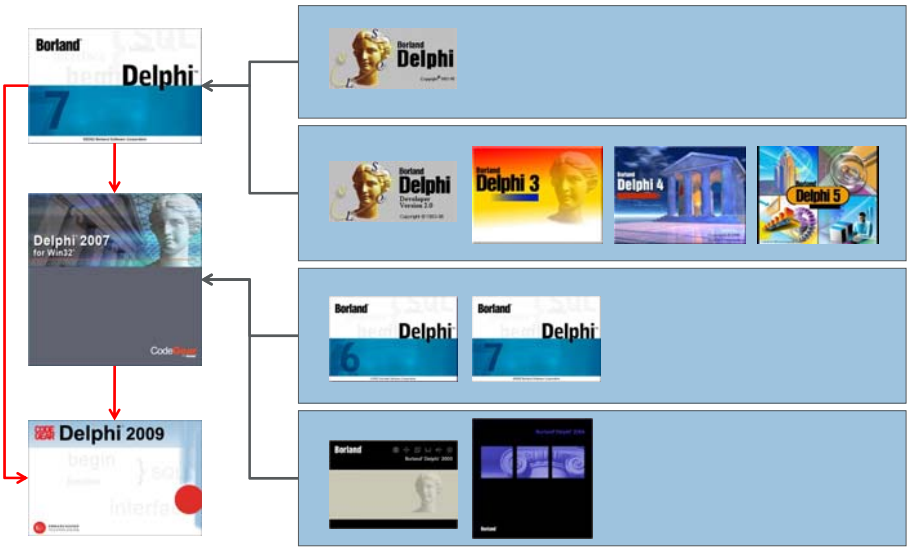
EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- Delphi 2007 でやるべき事
  - コンポーネント (パッケージ) / ライブラリの整備
  - XP / Vista 対応
  - BDE の代替
  - レポートツールの精査
  - （できれば）Delphi 2009 互換機能の実装
- ステップアップグレードのススメ
  - 16bit Windows 専用 → Delphi (1)
  - Win9x対応最終版 → Delphi 7
  - Ansi対応最終版 → Delphi 2007
  - Unicode 対応版 → Delphi 2009
  - 必要に応じて、Delphi 2009 からのバックポート

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

旧プロダクトからの移行 – Delphi 2007 へ –

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP



本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

Delphi 2009 と Unicode

Delphi 2009 と Unicode

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

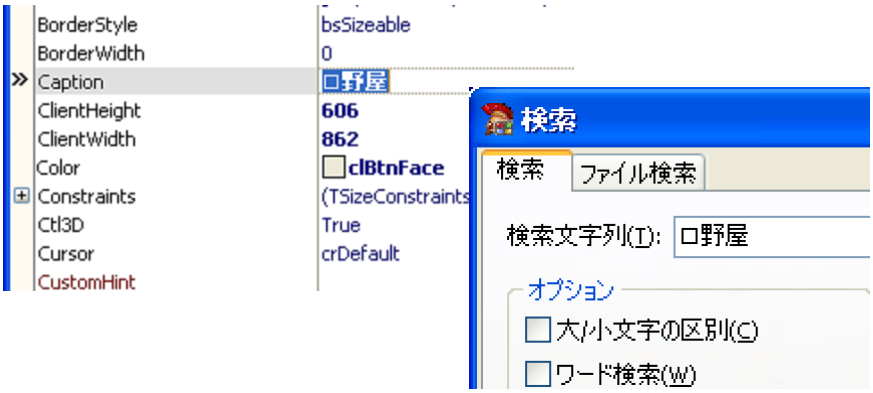
- IDE (1)
  - コードエディタ から オブジェクトインスペクタ に至るまで、完全に Unicode 化されている。サロゲートペアも正しく表示できる。

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

Delphi 2009 と Unicode

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- IDE (2)
  - Windows 2000 / XP ではシステムフォント “Tahoma” の影響で、サロゲートペアを正しく表示できない。

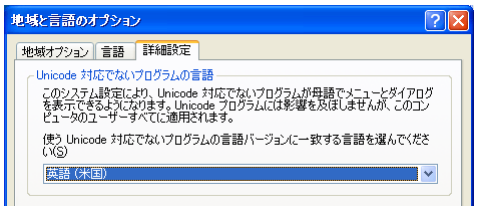


本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

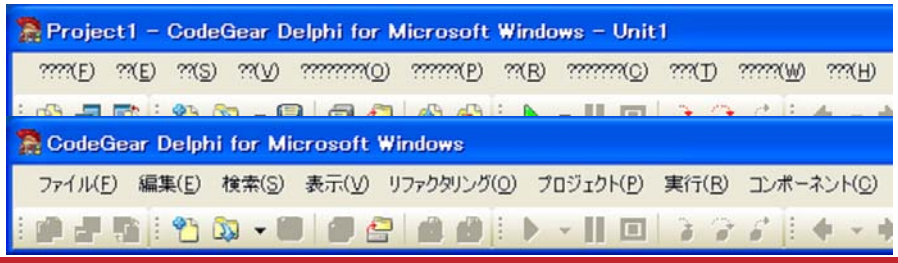
Delphi 2009 と Unicode

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- IDE (3)
  - Unicode化により、IDE がシステムのコードページに影響されなくなった。



上: Delphi 2007
下: Delphi 2009



本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

Delphi 2009 と Unicode

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

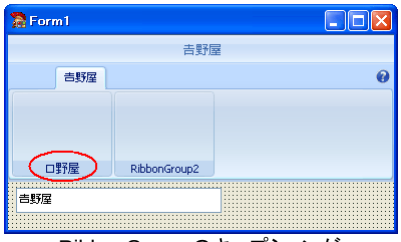
- RTL
  - RTL は Unicode を扱うための関数/クラスが強化されている。
  - ファイル入出力には TEncoding クラス を利用可能(後述)。
  - RTL は 基本的に W系API を呼び出す。

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

Delphi 2009 と Unicode

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- VCL
  - VCL も Unicode をサポートするが、IDE 同様、一部の VCL でシステムフォントの影響を受けてしまい、2000 / XP では正しくフォントを設定していても、サロゲートペアを正しく表示できない事がある。



RibbonGroup のキャプションが
口で表示されている(設計時/QCあり)

Button の Hint が
口で表示されている。

  - Hint や Form の Caption は [デスクトップ | プロパティ | デザイン | 詳細設定] にてフォントを変更する事により対処可能。

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

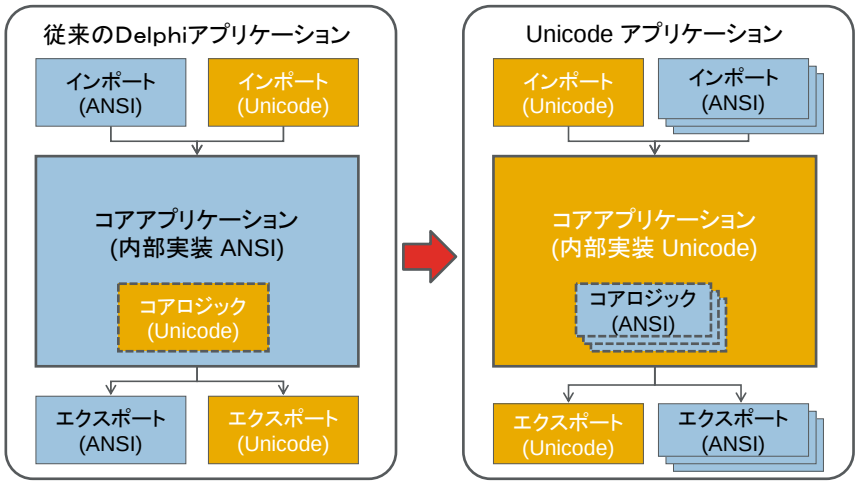
EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

旧プロダクトからの移行
– Unicode アプリケーション –

旧プロダクトからの移行– Unicode アプリケーション –

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- ANSIアプリケーション と Unicodeアプリケーション

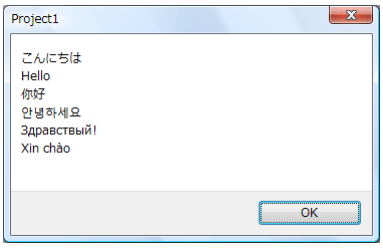


本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

旧プロダクトからの移行 – Unicode アプリケーション –

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- アプリケーションを Unicode化することのメリット
  - 多言語を同時に表示可能。



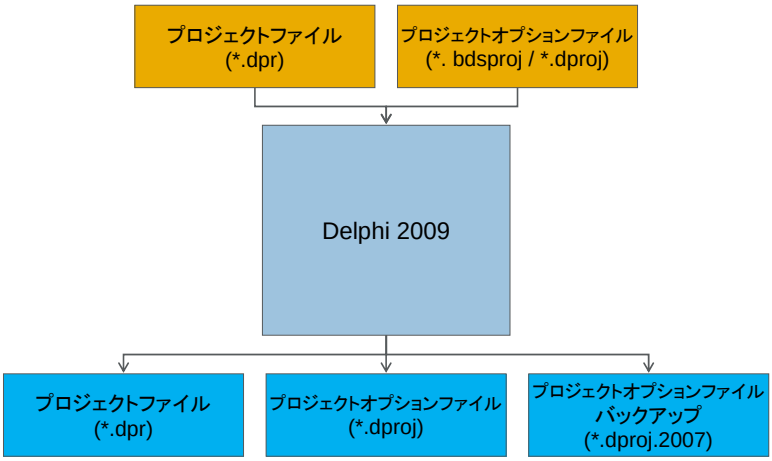
  - 日本でよく使われる文字に限定しても、利用可能な文字が増える。
    © ® ™ ㎥ 髙 etc…
  - 入出力ファイルに多くのANSIコードページが利用可能。
  - アプリケーションのローカライズが比較的簡単に。
  - 煩雑なマルチバイト処理から解放される。

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

旧プロダクトからの移行 – Unicode アプリケーション –

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- プロジェクトファイルの移行 (自動)

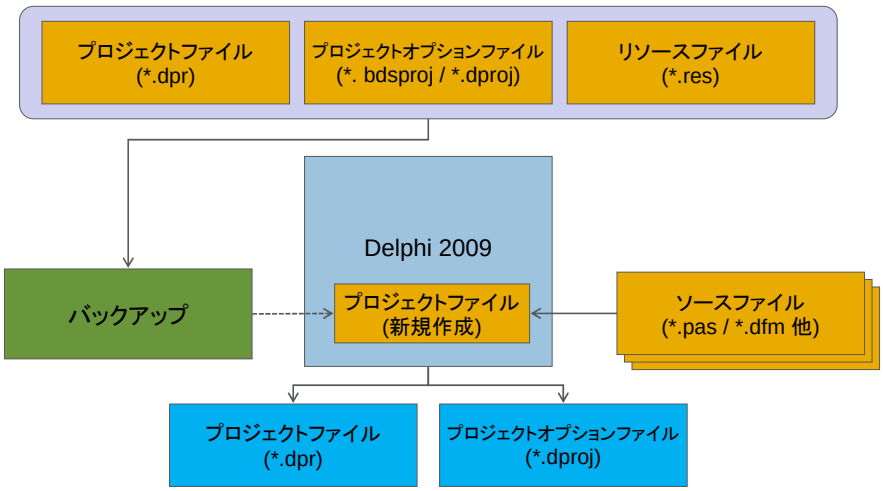


本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

旧プロダクトからの移行 – Unicode アプリケーション –

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- プロジェクトファイルの移行 (手動)



本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

旧プロダクトからの移行 – Unicode アプリケーション –

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- 新しいプロジェクトで不要なユニット / コンポーネントの除去
  - XPMan
    プロジェクトオプションで指定
  - FastMM
    標準のメモリマネージャと同等
  - VistaAltFix
    Delphi 2007 からは不要
- 新しいプロジェクトファイルに指定できるもの
  - ReportMemoryLeaksOnShutdown (BDS 2006 ～)
    アプリケーション終了時にメモリリークをレポート
  - Application.MainFormOnTaskbar (Delphi 2007 ～)
    メインフォームをタスクバーへ格納
  - Application. DefaultFont (Delphi 2009)
    “ParentFont := True” の フォーム 用デフォルトフォントの指定
  - Screen. MessageFont (Delphi 2009)
    メッセージボックス用デフォルトフォントの指定

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

旧プロダクトからの移行 – Unicode アプリケーション –

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- 文字欠損の検索
  - “暗黙的文字列キャスト(W1057)”
    “暗黙的文字列キャストによるデータ喪失の可能性(W1058) ”
    “set 式で WideChar がバイト文字に変換されました(W1050)”
    をワーニングからエラーに昇格させる。

```
{$WARN IMPLICIT_STRING_CAST ERROR}
{$WARN IMPLICIT_STRING_CAST_LOSS ERROR}
{$WARN WIDECHAR_REDUCED ERROR}
```

- Set式 を CharInSet() で置き換える。

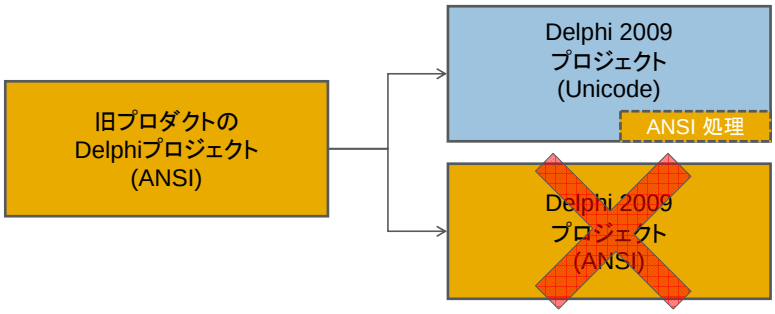
```
if (S[i] in ['A'..'Z']) then
    ↓
if CharInSet(S[i], ['A'..'Z']) then
```

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

旧プロダクトからの移行 – Unicode アプリケーション –

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- ANSI文字列の部分使用
  - ANSI文字列で内部実装してはいけない。局所的な使用に留めるべき。
  - String は AnsiString に置換。
  - 文字列操作関数の引数と戻り値、関数内の String は RawByteString に。
  - PChar は PAnsiChar に置換。
  - AnsiStrings 名前空間を uses した際の副作用の把握。



本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

旧プロダクトからの移行 – Unicode アプリケーション –

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- メモリリーク検出とデバッグ
  - 移行するプロジェクトには、
    “ReportMemoryLeaksOnShutdown := True;“
    を指定し、メモリリーク箇所を調べる。
  - 経験上、PChar 等のメモリ確保/解放部分にバグが潜んでいる事が多い。
- コントロール文字列の4桁化
  - 2桁のコントロール文字列は、意図しない ANSI->Unicode 変換を避けるためにすべて4桁化する (#$0D#$0A → #$000D#$000A)。
- 入出力するファイルのチェック
  - SHIFT-JIS で なくてはならないのか？Unicode でいいのか？
  - Iniファイルは それを利用する Unicode アプリケーションに合わせて、Unicode(UTF-16) で入出力するようにすべき。

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

- TEncoding (1)
  - TEncoding を利用して Unicode形式ファイルを入出力可能。

```
procedure TForm1.Button1Click(Sender: TObject);
var
  SL: TStringList;
begin
  SL := TStringList.Create;
  try
    SL.Add('あいうえお');
    // ファイル保存
    SL.SaveToFile('C:¥Unicode.txt', TEncoding.Unicode);            // UTF-16LE
    SL.SaveToFile('C:¥UTF-16BE.txt', TEncoding.BigEndianUnicode);  // UTF-16BE
    SL.SaveToFile('C:¥UTF-8.txt', TEncoding.UTF8);                 // UTF-8
    SL.SaveToFile('C:¥UTF-7.txt', TEncoding.UTF7);                 // UTF-7
  finally
    SL.Free;
  end;
end;
```

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

- TEncoding (2)
  - TEncodingクラスを利用できるメソッドは以下の通り。
    - ComCtrls.TOutlineNode.WriteNode
    - ComCtrls.TOutlineNode.LoadFromFile
    - ComCtrls.TOutlineNode.LoadFromStream
    - ComCtrls.TOutlineNode.SaveToFile
    - ComCtrls.TOutlineNode.SaveToStream
    - ComCtrls.TOutlineNode.WriteNode
    - Classes.TStrings.LoadFromFile
    - Classes.TStrings.LoadFromStream
    - Classes.TStrings.SaveToFile
    - Classes.TStrings.SaveToStream
    - Classes.TStringStream.Create
    - Classes.TStreamReader.Create
    - Classes.TStreamWriter.Create

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

- TEncoding (3)
  - TEncoding.GetEncoding() を利用する際にはインスタンスの破棄が必要。
  - CP_UTF8 を指定する事によって BOM なし UTF-8ファイルを扱える。

```
procedure TForm1.Button1Click(Sender: TObject);
var
  SL: TStringList;
  Enc: TEncoding;
begin
  SL := TStringList.Create;
  Enc := TEncoding.GetEncoding(CP_UTF8); // UTF-8(CP_UTF8/CP65001)
  try
    SL.Add('あいうえお');
    SL.SaveToFile('C:¥UTF-8N.txt', Enc);
  finally
    Enc.Free;
    SL.Free;
  end;
end;
```

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

- 移行後のブラッシュアップ
  - 今時の Unicode アプリケーションを作るのなら、それなりのブラッシュアップが必要。
  - フォント変更UIの追加
    - すべての Unicode 文字を網羅したフォントは存在しない。
    - アプリケーションで指定したフォントを相手が持っているとは限らない。
  - サロゲートペアへの対応
    - サロゲートペアは特殊文字ではない。
    - 詳細は http://dn.codegear.com/jp/article/38811 で。
  - 結合文字列への対応
    - “文字列検索” や ”データベースへのデータ格納” の処理を行う際には、前処理として正規化(Normalize)を行う必要がある。
    - 詳細は http://dn.codegear.com/jp/article/38816 で。

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

EMBARCADERO TECHNOLOGIES

DEVELOPER CAMP

資料

資料

EMBARCADERO TECHNOLOGIES

DEVELOPER CAMP

- Unicode 関連
  - Delphi Unicodeワールド パートI
    http://dn.codegear.com/jp/article/38781
  - Delphi Unicodeワールド パートII
    http://dn.codegear.com/jp/article/38698
  - Delphi Unicodeワールド パートIII
    http://dn.codegear.com/jp/article/38699
  - Delphi 2009 と Unicode : Part I
    http://dn.codegear.com/jp/article/38781
  - Delphi 2009 と Unicode : Part II
    http://dn.codegear.com/jp/article/38783
  - Delphi 2009 と Unicode : Part III
    http://dn.codegear.com/jp/article/38786
  - Delphi 2009 と Unicode : 番外編 (サロゲートペア)
    http://dn.codegear.com/jp/article/38811
  - Delphi 2009 と Unicode : 番外編 (結合文字列)
    http://dn.codegear.com/jp/article/38816
  - The Unicode Consortium
    http://unicode.org/

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

資料

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- Delphi 2009 関連
  - CodeGear Delphi 2009 および C++Builder 2009 のインストール ノート
    http://dn.codegear.com/jp/article/38484
  - CodeGear Delphi 2009 および C++Builder 2009 のリリース ノート
    http://dn.codegear.com/jp/article/38489
  - Delphi/C++Builder/RAD Studio2009 のアンインストール
    http://support.codegear.com/jp/article/38836
  - Delphi 2009 機能評価ガイド
    http://dn.codegear.com/jp/article/38817
  - Delphi 2009 と 文字列型
    http://dn.codegear.com/jp/article/38791
  - Delphi 2009 のIDE設定
    http://homepage1.nifty.com/ht_deko/tech020.html
  - リファレンスマニュアル
    http://docs.codegear.com/
  - Quality Central (Delphi 2009 関連)
    http://qc.codegear.com/wc/qcmain.aspx?search=1&proj=10&vers=12.0
  - QualityCentral の Tips
    http://dn.codegear.com/jp/article/38793

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

資料

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- その他
  - Delphi の製品情報
    http://www.geocities.jp/ht_deko/Delphi/index.html
  - InstallAware を使って配布モジュールを作成する
    http://dn.codegear.com/jp/article/34383
  - InstallAware に関する Tips
    http://homepage1.nifty.com/ht_deko/tech023.html
  - 今更ながら BDE (Borland Database Engine)
    http://homepage1.nifty.com/ht_deko/tech024.html
  - MECSUtils
    http://cc.codegear.com/item/26061
  - MECSUtils リファレンス
    http://homepage1.nifty.com/ht_deko/tech021.html
  - Vista Alt Fix
    http://dn.codegear.com/jp/article/38791
  - FastMM4
    http://sourceforge.net/project/showfiles.php?group_id=130631
  - QuickReport Standard
    http://www.quickreport.co.uk/stanlegacy.html

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。

資料

EMBARCADERO TECHNOLOGIES
DEVELOPER CAMP

- MEMO

本文書の一部または全部の転載を禁止します。本文書の著作権は、著作者に帰属します。