

## 中文外字服務使用說明

### 壹、前言

本服務提供各內碼間的轉換功能，其中支援的內碼包含 BIG5、DOH、EUC、UTF-16LE，以中推會元件為核心，完善處理各系統間交換資訊產生的內碼轉換問題，但基於因為 Web Service 先天上的限制，本服務並不適合使用在大量的資料轉換。

### 貳、Web Service 使用說明

(一) WSDL 測試路徑 (以實際介接所提供之 IP 為主)

<http://203.65.17.143:8080/CharacterSet/PortTypeBndPort?WSDL>

(二) WSDL 帶入參數及輸出結果

**表1 帶入參數說明**

參數	型別	說明
sourcetype	string	來源的內碼編碼格式，包括 BIG5、DOH、EUC 與 UTF-16LE。
desttype	string	欲轉出的內碼編碼格式，包括 BIG5、DOH、EUC 與 UTF-16LE。
convertcontent	string	欲轉換的內容。

**表2 輸出參數說明**

參數	型別	說明
contentresult	string	轉換後的內容。

## 參、其他注意事項

### (一) Java 字元集設定規則

- 1、如果資料來源是資料庫，且資料庫字元集為 UTF-8 時，資料以字串型態讀入程式時為 UTF-16 字串，以 UTF-16LE 方式來處理。
- 2、如果資料來源是文字檔案，無論檔案的編碼方式為何，讀入程式時皆為 UTF-16 字串。

針對文字檔案輸入/輸出欄位不同之內碼，BufferedReader 與 BufferedWriter 的字元集之設定說明如下表。

**表3 文字檔案輸入/輸出內碼之字元集設定說明**

文字檔案輸入 <charset-string>	文字檔案輸出 <charset-string>	字元集
BIG5	BIG5	BIG5
UTF-16LE/UTF-8	UTF-16LE/UTF-8	UTF-16LE
MS950	ISO-8859-1/MS950	EUC
MS950	MS950	DOH

以 EUC 轉為 BIG5 內碼為例，說明如下：

=====

```
string source = "";
BufferedReader BR = null;
try {
    //讀取 EUC 編碼的文字檔，此時的 Character Set 為 MS950
    BR = new BufferedReader(new BufferedReader(
        new InputStreamReader(
```

```
        new FileInputStream("D:\\Output_MS950.txt"), "MS950"));
    source = BR.readLine();
    BR.close();
} catch (FileNotFoundException Fe) {
    System.out.println(Fe.getMessage());
} catch (IOException e) {
    System.out.println(e.getMessage());
}

//呼叫轉碼服務(Web Service)
InputWebServiceService ConvertTest = new InputWebServiceService();
CharacterSet.PortType service = ConvertTest.getPortTypeBndPort();
try{
    //將轉碼的結果輸出到文字檔，此時的 Character Set 為 BIG5。
    BufferedWriter bw = new BufferedWriter(
        new OutputStreamWriter(
            new FileOutputStream("D:\\Output_BIG5_Line.txt"), "BIG5"));
    //呼叫服務
    bw.write(service.operation("EUC", "BIG5", source));
    bw.close();
} catch (UnsupportedEncodingException e)
    System.out.println(e.getMessage());
} catch (IOException IOe)
    System.out.println(IOe.getMessage());
}
```

=====

- 3、在 Java 的定義中，BIG5 字元集是不包括 BIG5 造字，只有標準的 13,051 字。所以資料內容如果包含造字，一定要用 MS950 取代 BIG5，否則造字會變成問號。

- 4、以文字檔案輸入與輸出時，如果文字檔案的編碼為 UTF-8 時，要用 UTF-8 取代 UTF-16LE。