



tsss001 Feb 12, 2019 10:04 PM

[ㇿ-龠](2E80~2FD5)のコードポイントの文字

This question has been **Answered**.

クライアントから頂いたデータを元にレイアウト作業を行っていて、スクリプトを絡めて作業する中でコードポイント2E80~2FD5の範囲の漢字が混じっていることに気がつきました。

具体的には[黑白](9ed2,767d)が[黑白](2fca,2f69)になっているなど。

クライアントからのデータはwin cs6のai,indd,およびexcel、

こちらはmacのcs6なのですがこれらの文字が何かの拍子に勝手に変換されるって事はありませんか？

また、これらの文字を自動で置き替える事は可能でしょうか？

今回は数が少なかったのでfind=[黑白] replace=[黑白]の様なマッピングを作成してスクリプトにて対処しましたが。



Correct Answer

by monokano on Feb 12, 2019 10:54 PM

もしかして、PDFから文字を抽出したりしていませんか？ PDFなら場合によってはありえるかもしれません。

これらの文字を自動で置き替える事は可能でしょうか？

「CJK部首補助+康熙部首 → 普通の漢字」のマッピングテーブルは以前作ったことがあるので、よかったら利用してみてください。タブ区切りのUTF-8です。

[CJK部首補助+康熙部首→普通の漢字.csv.zip](#) - Box



monokano Feb 12, 2019 10:54 PM (in response to tsss001)



Correct Answer Re: [ㄣ-龠](2E80~2FD5)のコードポイントの文字

もしかして、PDFから文字を抽出したりしていませんか？ PDFなら場合によってはありえるかもしれません。

これらの文字を自動で置き替える事は可能でしょうか？

「CJK部首補助+康熙部首 → 普通の漢字」のマッピングテーブルは以前作ったことがあるので、よかったら利用してみてください。タブ区切りのUTF-8です。

[CJK部首補助+康熙部首→普通の漢字.csv.zip](#) - Box

4 people found this helpful

Actions

Edit

Delete

Report Abuse

Liked (1)



tsss001 Feb 12, 2019 10:58 PM (in response to monokano)

Re: [ㄣ-龠](2E80~2FD5)のコードポイントの文字

ありがとうございます。

返答のタイミングがズレましたが、まさにPDFからでした。

助かります。マッピングテーブル活用させていただきます。

Actions

Edit

Delete

Report Abuse

Like (0)



tsss001 Feb 12, 2019 10:57 PM (in response to tsss001)

Re: [ㄣ-龠](2E80~2FD5)のコードポイントの文字

すみません。自己レスです。

いろいろ確認していたところ、PDF(パワーポイントから作成)をアクロバットで開き、文字をコピーした時に

一部の文字が「白が白白」などとダブって変換され、その部分の頭の文字の文字コードが変換されているようでした。

ちなみにアクロバットからテキスト書き出しした場合、こちらもまともに出力されなかったのですが、

ダブリと文字コードが勝手に変更される現象は起きていないようでした。

Actions

Edit

Delete

Report Abuse

Like (0)

