

文字化けへの対処

文字化けとは、通常の文字で書かれていたはずのものが全然違う文字に変換されて読めない状態のことを言います。

皆さまも、ウェブサイトを開覧していて、ページ内のすべての文字が、例えば「譚◆ュ怜喧緹代@緹ヲ緹◆U緹◆」や「◆◆紘◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆障◆◆」などと表示されてしまった経験があるかと思います。

「文字コード」のページで解説した通り、これは、文書を作成した人のコンピュータ環境の文字コードセットと、それを読む人のコンピュータ環境で理解した文字コードセットが一致していないために起きる現象です。

文字化けは、Web の閲覧時、受信した電子メール、電子メール等に添付されていた Word などの文書、データをダウンロードした時の CSV ファイル、などで発生します。ここでは、それぞれの場合に応じた対処方法を解説します。

Web 閲覧時の文字化け

Web ブラウザで閲覧しているページが文字化けすることがあります。この場合の文字化けは、Web コンテンツを提供している側の責任、つまり、Web ブラウザがそのページがどの文字コードで書かれているかを明確に宣言していないからであって、閲覧している側に問題があるわけではありません。

多くの場合、そのページが古い手法で書かれていた時に生じます。

現在では多言語を同じページ内に表記する必要性が高まっているため、多言語に対応できる UTF-8 によってテキストを書く(保存する)ように推奨されていて、Web ブラウザ側も、一般的には、文字コードが明示されていない場合(デフォルトエンコード)は、UTF-8 で書かれていると見なすようになっていきます。Web ページがシフト JIS (Shift_JIS) や EUC-JP で書かれているに、ページの先頭にその文字コードを使用している旨の宣言がないと、文字化けしてしまうのです。

【便利知識】

文字化けに遭遇した時に、Web ブラウザがインターネットエクスプローラー（IE）であれば、右クリックして表示されるショートカットメニュー上の「エンコード」をクリックすると、日本語（シフト JIS）を選んだり、さらに「その他」から日本語（EUC）や各国の言語用の文字コードを選んで切り替えることができます。



IE11 の文字コード変換の仕方

【便利知識】

Web ブラウザが Safari（サファリ）や Firefox（ファイヤフォックス）の場合は、設定機能でデフォルトの文字コードを設定できるようになっています。

Web ブラウザが Windows10 で標準の Edge（エッジ）や、Google が提供する Chrome（クローム）の場合は、残念ながらこのエンコードの機能が標準では用意されていません。それだけ、UTF-8 の文字コードで書かれたページが普及したからでしょう。

Chrome にはエンコードのための「テキストエンコーディング」という名称の拡張機能（プラグイン）が公開されていますが、上級者向けのテクニックとなりますので、インターネットエクスプローラーなどの別のブラウザでリトライする方が简便でしょう。

受信メールの文字化け

受信メールの文字化けには、メールの題名や本文全体が文字化けしている場合と、一部の文字が文字化けしている場合の 2 通りがあります。

パソコンで受信したメールの一部分に「＝」や「？」の記号が入ってしまうなどの文字化けは、「機種依存文字」や「携帯電話の絵文字」が文字化けしたことが多いです。一般的に、これらの文字はパソコン用のメールソフトでは正しく表示されません。「機種依存文字」や「携帯電話の絵文字」などによる文字化けは、メールソフトの設定では解決できないため、これらの文字を使わないようにするしかありません。

メールの題名や本文全体が文字化けをしてしまうのは、Web 閲覧の場合と同様、文字コードの違いによります。

Web ページの文字コード UTF-8 が推奨されていますが、日本語対応のメールソフト（メーラー）は過去からの経緯もあって、いまだにシフト JIS が標準になっています。

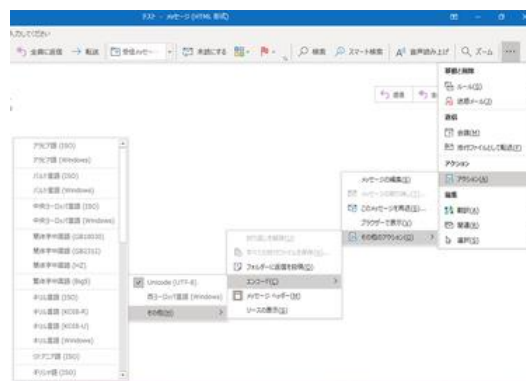
NTTドコモ、au、ソフトバンクなどの携帯電話会社のメールも、ベースはシフト JIS で、さらに絵文字や丸付き数字などの特殊文字を特定の場所に割り当てています。

一般的に、メールソフトには文字コードを変換して再表示する「エンコード」の機能が備わっています。

Outlook の場合

Microsoft Office の Outlook では、バージョンによって若干操作が異なりますが、「アクション」⇒「その他のアクション」の中に、「エンコード」があり、文字コードを選択できます。

右図は、Outlook2016 の例です。



Outlook2016 のエンコード

Windows メールの場合

Windows 10 などに標準的に備わっている Windows メールでは、バージョンによって若干操作が異なりますが、「表示」⇒「エンコード」で、文字コードを選択できます。

Gmail の場合

Gmail にはかつては「文字化けを修復」という機能がありましたが、現在はこの機能がなくなってしまいました。

ですので、上述の「Web 閲覧時の文字化け」で記載した、ブラウザのエンコード機能を利用して、文字コードを変換するしかありません。

Word 文書等の文字化け

Word、Excel、PowerPoint などの Office 文書での文字化けの原因には、以下のようなことが挙げられます。

1. ソフトウェアのバージョンの違いや互換性による不具合

ワードを含む Office ソフトは数年単位で大きくバージョンアップします。新しいバージョンで作成したファイルを古いバージョンのソフトで開くと、不具合を起こして文字化けを起こす事があります。

2. 文字コードを正しく認識していない

例えば、「メモ帳」や「OpenOffice」などで作成した文書をワードで開くと、文字コードの違いや、文字コードの混在等の原因で文字化けする事があります。また、海外支店から日本へ送信されたメールに添付された Word 文書を開くと文字化けしたなども、使用されている文字コードを正しく認識できなかったことによります。

3. コピーミスなどによるファイル破損

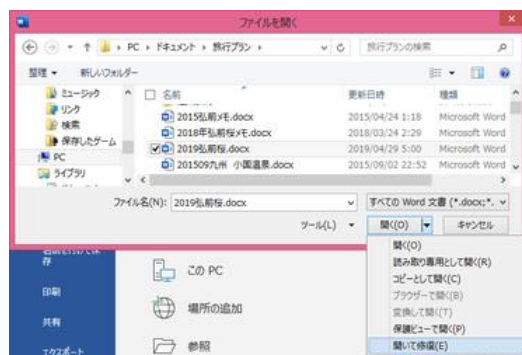
もし、自分で作成したエクセルデータや、日本語版の Office ソフトを使っているにも関わらず、文字化けする場合、ファイルが破損している可能性があります。

「開いて修復」

Word、Excel、PowerPoint には、「開いて修復」というリカバリー方法があります。バージョン違いなどによる文字化けには、この方法が有効です。

ですので、まずはこの方法を試してみることをお勧めします。

具体的には、Word を立ち上げ、「ファイル」タブから、「開く」⇒「参照」で「ファイルを開く」ダイアログボックスを出し、対象の文書を選んでから、右下の「開く」の右側の「▼」ボタンをクリックして「開いて修復」を選びます。



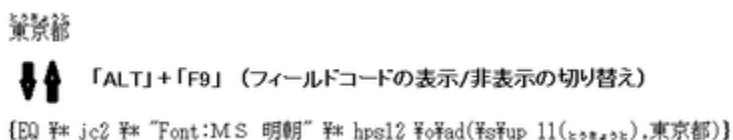
Word の「開いて修復」

Excel では「開いて修復する」、PowerPoint では「アプリケーションの自動修復」という呼び名に変わりますが、ほぼ同様の操作です。

【便利知識】

Word 文書で「フィールドコード」を表示した場合には、一見文字化けのように見えてしまいます。

例えば、漢字にルビを振った状態で「Alt」キーと「F9」キーを押すと、フィールドコードが表示されて、あたかも文字化けしたようになりますが、再度「Alt」キーを押しながら「F9」キーを押すと、元のルビ付きの漢字に戻ります。



CSV ファイルの文字化け

CSV ファイルをダブルクリックで Excel で開くと、文字化けすることがあります。

この原因は、その CSV ファイルの文字コードが BOM(Byte Order Mark) 無しの UTF-8 になってるからです。Excel も UTF-8 に対応してますが、BOM が付いていない場合にはその判断ができず、Shift_JIS だと思って開こうとするので文字化けしてしまいます。

【便利知識】

この場合の対処方法として簡便なのは、一旦「メモ帳」のソフトで CSV を開いて「上書き保存」をする方法です。これにより、BOM 付きの UTF-8 形式で保存されますので、次回から Excel でも文字化けせずに開くことができます。

PDF の文字化け

PDF で保存すれば、どのような環境でも同じように表示されるはず…ですが、ごく稀に PDF で保存したとしても文字化けなどのトラブルが起こる場合があります。

OS や Chrome などのブラウザに標準装備されている PDF ビューアで PDF を開いた場合、文字化けや表示エラー等が起こることがあります。そのような時は、その PDF をダウンロードして、Adobe 社の「Adobe Acrobat Reader DC(もしくは Adobe Reader)」で閲覧しましょう。

PDF 文書には通常その文書で使われているフォントが埋め込まれています(プロパティのフォントで「埋め込みサブセット」と表示されます)。もし、PDF 文書にフォントが埋め込まれておらず、システムがそのフォントをサポートしていない場合は、表示や印刷において文字化けが発生します。

PDF ファイルにテキストとして組み込まれている文字情報の全部または部分的に取り出して、Word などに書き出すことができます。最も簡単な方法はコピー&ペーストで、大抵の場合は問題なく文字だけを取り出すことができます。

もしも元のファイルに組み込まれたテキストに特殊なフォントが使われていたり、フォントのデータが正しく埋め込まれていないか、編集された際に破壊された場合には、正しく表示されていても文字化けを起こすことがあります。

【便利知識】

PDF ファイルや web ページからテキストをコピーし、Word に貼り付けると、行ごとに改行されてしまい、見栄えが悪いということがあります。短い文章であれば、手で不要な改行を消していけば良いのですが、長い文章では億劫です。そのような時は、「置換」の機能を使って、不要な改行を自動的に探して、まとめて削除(文字無しに置き換え)することができます。