入門者のための FAQ

Frequently Asked Question

今月の回答者
砂原秀樹、菊地宏明

【アドハイザー】砂原秀樹
奈良先端科学技術大学院大学
情報科学センター助教授
WIDE プロジェクト・ボードメンバー

このコーナーでは、みなさんから寄せられたインターネットに関する質問や疑問についてお答えしていきます。日頃からわからないなあと思っている疑問、困っていることなどありましたらどんなことでもけっこうですから質問を編集部までお寄せください。宛先は ip-faq@impress.co.jp です。電子メールでの回答はできませんのでご了承ください。

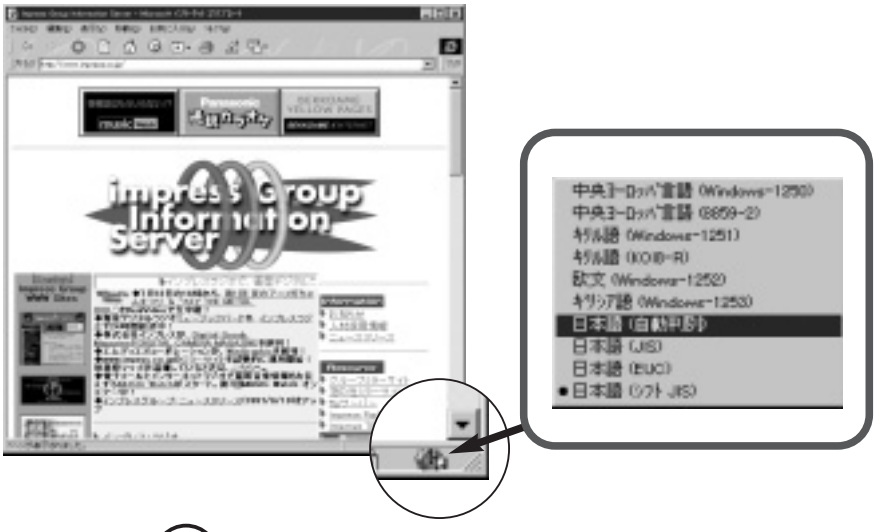
ブラウザーの表示言語を選択（設定）するところで、「日本語（自動判別）」とあるのに、その下に「日本語（JIS）」、「日本語（EUC）」、「日本語（シフトJIS）」まであるのはなぜですか？ 自動判別できるのであれば、そのほかには必要ないように思いますが…。文字が化けるたびにいちいちここを設定するのはとても面倒です。

（林加奈子さん）

A. ブラウザーの設定と、その設定を反映した動作はブラウザーの種類やバージョンによって変わるので一概にいえません。そこでこんな実験を行ってみました。テキストエディターで1ページ分のHTMLファイルを作ります。その中にJIS、EUC、シフトJISの漢字コードで書かれた文字を入れておきます。このファイルをブラウザーで表示させてみましょう。（右ページ参照）

ネットスケープコミュニケーターでは、言語表示で自動判定／シフトJIS／EUCの3つの選択肢があります。まず、EUCを選択すると、EUCコードの文字だけがうまく表示され、シフトJISコードの文字は一部に文字化けが生じました。JISコードはまったく読めません。次に、シフトJISを選択す

漢字コードの設定方法



マイクロソフト・インターネットエクスプローラ（上）と、ネットスケープコミュニケーター（下）の漢字コード設定画面。エクスプローラは、ステータスバーの右下にあるマークをクリックして設定する。コミュニケーターの場合は、一度設定したあともう一度設定画面で「Set Default Encording」を選択して設定を有効にするのを忘れずに。

ると、シフトJISコードの文字だけが読み取れ、あとのJISコードとEUCコードは読み取れなくなりました。最後に自動判定を指定して、表示させると、ページ先頭のJISコードの文字とシフトJISは読め、EUCコードだけが読めません。ここでわかることは利用するコンピュータのシステムのネイティブ漢字コードは変換なしに読めることと、ただ、これでは自動判定を選んでもどの漢字コードでも読めるわけではないらしいということです。自動判定はJISコードだけを変換するのかという疑問が残ります。

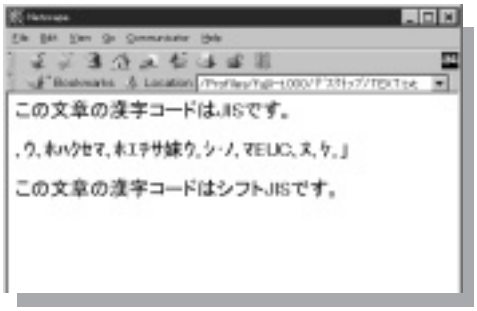
そこで、再度エディターでHTMLファイルを修正し、それぞれの漢字コードが現れる順番を変えてみてください。すると、ページの先頭にくる漢字コードの種別によって自動的に漢字コードが判定されるのがわかりました。そして、いったん判別されると、途中で漢字コードが変わっても、追従しない仕様になっていました。

つまり、特定の漢字コードだけでページが書かれているのならば、自動判定で十分対応できますが、ページの一部に異なる漢字コードがあるとその部分の表示が乱れるのです。その部分を表示させるために、個別の漢字コードが選択できたほうがいいのです。もちろん同じページ内でも随時漢字コードを判定しながら表示させることは技術的にはできるのでしょうが、効率的ではないので、そのようなWWWブラウザーはないのでしょう。　　（菊地宏明）

実験に使ったテキスト（HTMLファイル）

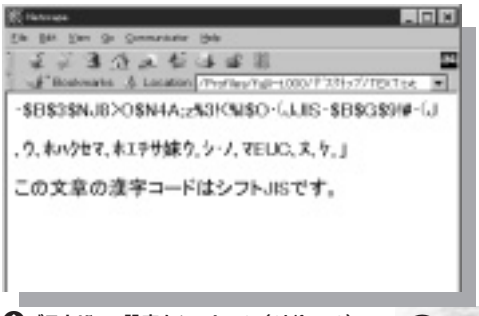
この文章の漢字コードはJISです。
JISコード
この文章の漢字コードはEUCです。
EUCコード
この文章の漢字コードはシフトJISです。
シフトJISコード

❶ 自動判別に設定した場合



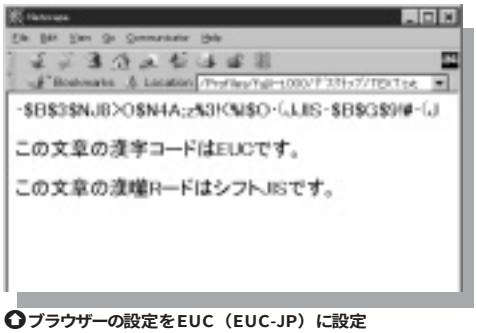
ブラウザーの設定を自動判別（Auto-Detect）に設定した。JISとシフトJISの文章を読むことができた。

❷ シフトJISに設定した場合

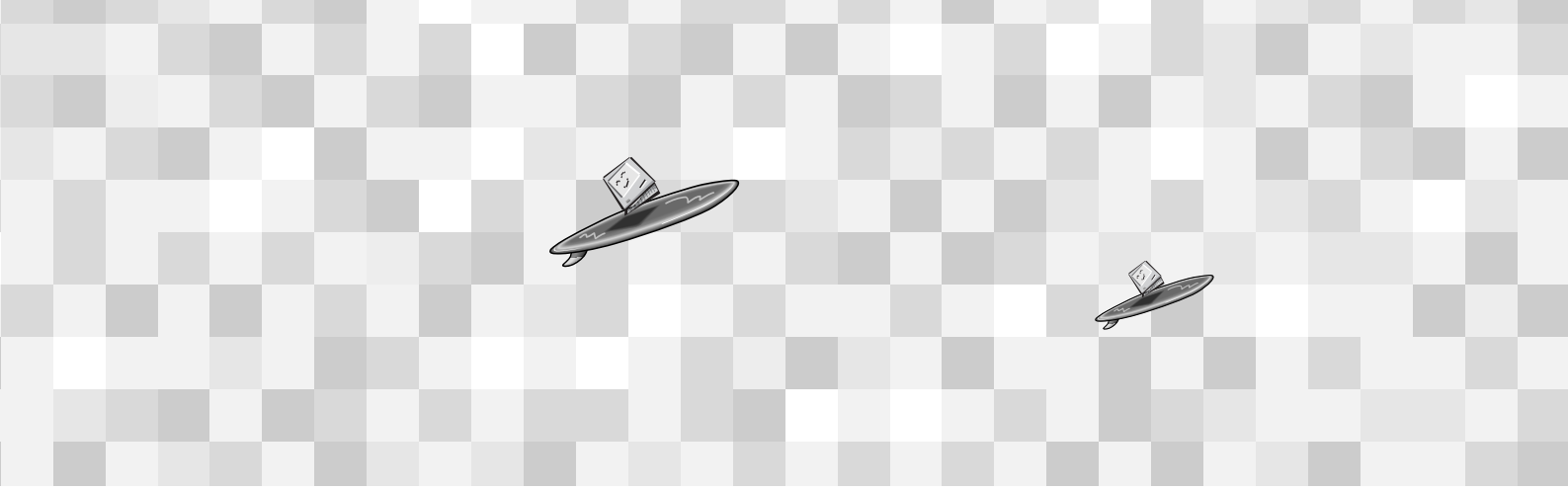


ブラウザーの設定をシフトJIS（Shift_JIS）に設定した。シフトJISの文字以外は文字化けして読めなかった。

❸ EUCに設定した場合



ブラウザーの設定をEUC（EUC-JP）に設定した。EUC以外の文章では文字化けした。シフトJISも一部に文字化けが見られた。



最近、解凍用のツールやソフトはよく出回っています。圧縮用のものも、あるにはあるんですが、どうも不満です。たとえば、100Kバイトのデータなら、10分の1の10Kバイトくらいまで圧縮できるものはないものでしょうか。もちろん、100Kバイトを1Kバイトにできるなら文句なしですが、そんなことは可能なんでしょうか？

（関根政実さん）

A. ディスクの容量はいくらあっても足りないですから、データを100分の1にできる仕組みがあったとしたら便利ですよね。さて、この話を厳密にするには、結構面倒な「数学」にしばらくお付き合いいただかないとならないのですが、限られたスペースの中ではなかなか難しいので、ここでは概念を簡単に説明したいと思います。実は情報にも「量」という概念があるのです。たとえば、

abbaab

という文字列があったとします。これをASCII文字列として表すと、1文字を8ビットで表現し、それが6文字ありますから、48ビットで表現されることになります。

しかし、この文字列をよく見ると「a」と「b」の2文字しか使っていませんから、1文字を8ビットで表現（これは256種類の文字を表現できる）することは非常にもったいないわけです。そこで「a」を0、「b」を1とそれぞれ1ビットずつで表現することにすれば、なんと6ビットで表現できてしまうのです。ただし、0が「a」、1が「b」に割り当てられたという情報を記録しておかなければなりません。そのために約16ビット程度のスペースが必要となります（ASCII文字「a」、「b」と0、1の対応表が格納されることになり、ASCII文字2文字分強のスペースが必要となる）。したがって、20数ビットでこの情報は表現できるということになります。

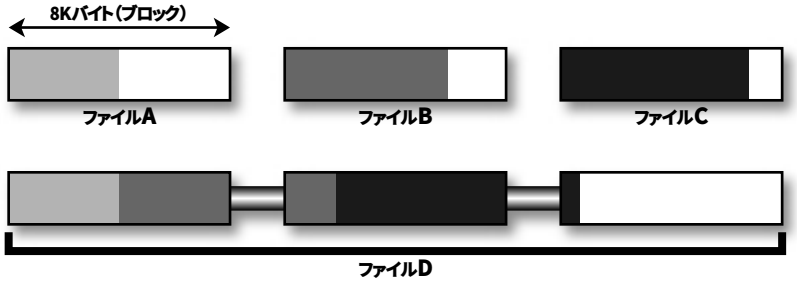
ところが、これ以上効果的に情報を表現しようとしても、なかなか難しいのです。たとえば、「ab」を0、「ba」を1に割り当て、010の3ビットで表現しようとした場合でも、ASCII文字列「ab」、「ba」と0、1との対応表のために32ビット程度のスペースが必要となり、全体で40ビット程度になってしまいます。

このように、情報を表現するための最低限必要なスペース（ビット数）のことを「情報の量」と呼んでいます。そしてこの「量」は情報（データ）によって異なります。実際の圧縮操作においては、ここで紹介した方法より複雑な仕組みを用いてできるだけ効果的に圧縮しようとしているのですが、どんなにすばらしい圧縮の方式（アルゴリズム）であっても、この「情報の量」よりも小さくすることができないわけです。また、すでに圧縮されたデータを、さらに圧縮ツールを用いて小さくするということができないのも同一の理由によります。

ですから、データをどの程度に圧縮できるのかという問題は、圧縮のアルゴリズムの効果もさることながら、そのデータの「情報の量」がどの程度かということに依存しているのです。基本的に以下のようなデータは圧縮しやすいと考えられます。

❶使われている文字の種類が少ない（バイナリーデータの場合は、1バイトずつに分割したときの各バイトのパターンの種類が少ない）。
❷使われている文字（各バイトのパターン）に偏りがある（たとえば「a」の文字が非常に多く利用されている）。
❸同一の文字列（バイトパターン）が何度も登場する。

複数のファイルをまとめる「アーカイブ」の仕組み



実際にデータが格納されているのはブロックの一部だけなので、残り部分は無駄になる（ファイルA～C）。これらを1つの大きなファイル（ファイルD）にすれば連続領域になるので、最後のブロック以外はいっぱいに使われることになる。

このためバイナリーデータの圧縮率は低く、テキストデータの圧縮率は比較的高いわけです。また、ASCII文字列で表現された数字だけが格納されているテキストデータならば、場合によっては100分の1の圧縮率を実現できるかもしれません。

というわけで、一般にどんなデータでも10分の1にしてしまうということは非常に難しいでしょう。ただし、圧縮ツールも使い方次第では圧縮率以上に効果的に利用できます。これは、多くのオペレーティングシステムで採用されているファイル格納の仕組みが理由となっています。

基本的に、ファイルはブロックと呼ばれるものの組み合わせでデータが格納されます。このブロックは大きさが決まっており、たとえば8Kバイトの大きさの領域に必要なデータを格納していくようになっているのです。したがって、10バイトの大きさしかないファイルでも、7.5Kバイトの大きさのファイルでも8Kバイトのブロック1つずつを利用するようになります。ですから、多くのファイルが利用しているブロックには多くの空きがあり、ディスク上は隙間だらけになっているのです。圧縮ツールの多くがアーカイバー（複数のファイルをまとめるツール）と組み合わされているのはまさに、これが理由なのです。つまり、単にデータを圧縮するだけでなく、複数のファイルを1つの大きなファイルにまとめてしまうことで、ブロックの隙間をできるだけなくそうとしているわけです。「DriveSpace」や「Stacker」といったディスク圧縮ツールも同様の概念で構成されています。

ですから、適当な単位でファイルをまとめながら圧縮をしたほうがより効果的であるわけです。これは、圧縮によってデータをコンパクトにするだけでなく、ファイルが実際にディスク上で消費するスペースも節約できるためです。

ところで、データを10分の1や100分の1にする圧縮アルゴリズムがないわけではないのです。これまでの話では、圧縮において元の情報は損なわないことが前提となっていました。つまり、圧縮して復元すると完全に元のデータになっていることが要求されていたわけです。これが「可逆圧縮」と呼ばれるものです。

しかし、場合によっては完全に元のデータに復元する必要のない場合もあります。たとえば、この回答も結構長く文章を書いていますが、

できません。

ということが本質であり、それだけで十分である場合もあります。このように、元の情報を認識できる程度に、不要であると思われる部分を取り除きながら圧縮する（つまり、元の情報の「量」を削る）方法もあります。これが「不可逆圧縮」です。

圧縮率を変えて保存したJPEG形式の画像データ

▲368,762バイトのJPEG形式の画像。画像の劣化はまったくといっていいほど気にならない。

▲左の画像を17,504バイトまで圧縮。画像の劣化が著しい。

完全の元の情報には戻りませんが、たとえば画像などでは人間が認識することができない情報までも格納しておく必要は少ないわけです。その代表がJPEGというアルゴリズムです。画像の質を落とすことで情報を削って圧縮するようになっており、圧縮率を指定して圧縮操作を行います。さすがに100分の1にしてしまうと元の画像との差が大きくなりますが、10分の1程度の圧縮なら、よく見ないと判別できない場合が多いようです。

とはいえ、不可逆圧縮が利用できる場合は限定されているため、多くの場合は可逆圧縮を用いることになります。現在の多くの圧縮ツールで用いられているアルゴリズムは、数学的には限界に非常に近づいているようです。ですから、これ以上圧縮率を劇的に変化させることは難しいのではないでしょうか。それよりも、圧縮ツールをうまく使いこなすことが、ディスクを効果的に使うためのポイントになるのではないかと思います。

（砂原秀樹）

以前、デジタル電話機というものを売っているのを見たことがあるのですが、普通の電話機と比べてとても高い値段でした。これっていったい何者なのでしょうか？私はISDNを引いていますが、なにかメリットのある商品なんでしょうか？ デジタル電話機について教えてください。　（樋口啓一さん）

A. デジタル電話機はINSネット64のデジタルインターフェイスを備えた電話機です。通常はDSUに接続するためのS/T点インターフェイスを持ち、バス型配線のコネクターに接続することができます。受話器から送られる音声をデジタル化して送り出すため、音質の劣化が少なく高品位の音声を扱うことができます。TA経由でアナログ電話をつないだ場合、音声はアナログデータとしてTAに送られ、TAでISDNに対応したデジタルデータに変換されます。ということは、デジタル電話機より1段階余分なステップを踏み音声劣化につながります。街のグレーのISDN公衆電話（通称：グレ電）が音質がよいのはこのためです。

ISDNの通話サービスにフレックスホンなどがあります。3者間での同時通話を行ったり、相手を切り替えながら話したりできます。また、かかってきた電話を第三者に転送する機能などもあります。デジタル電話機には、これらの機能をサポートする機能キーを備えていたり、サブアドレスを指定するボタンが装備されていたりします。アナログ電話機には専用のキーがないため、フレックスホンをサポートするTAにつないでも「＊」、「＃」キーを使って呪文のようなコマンドを入力する必要が生じます。ですから、デジタル電話機は機能キーを装備していることで、操作性は高いといえます。

では、オプションサービスを使わなければ差はないのでしょうか？ アナログ電話にはグレ電のような「発信ボタン」がありませんが、グレ電と家庭用のデジタル電話機には、電話番号をダイヤルした後に押す発信ボタンがあります。この発信ボタンを押すと瞬時に相手を呼び出してくれます。しかし、TAにはこのボタンがなく、ダイアルが終了したことをタイマーで監視しています（TAによっては、最後に決められたボタンを押すと発信ボタンの代わりになるものもあります）。この機能がないと、ダイアル後しばらく待たないといけないのです。

デジタル電話機の価格は以前とくらべると格段に安くはなりましたが、TA並みの低価格化が起こっていないので「アナログ電話＋TA」の組み合わせに比べて高いのは否めません。ぜひとも低価格製品が出回ってほしいものです。

（菊地宏明）

▶▶▶▶▶▶▶▶ 主なデジタル電話機 ◀◀◀◀◀◀◀◀

T-1000

デジタル電話機の先駆機といえるNTTの「T-1000」。現在は製造されていない。当時の価格は298,000円。

S-1000

現在もNTTで販売しているデジタル電話機「S-1000」。「T-1000」とくらべると価格が格段に下がって29,800円。

ALION（HD-1000）

TAとDSUを内蔵したデジタル電話機「ALION（HD-1000）」にはPHSにもなる子機1台が付属。岩崎通信機（株）から198,000円で発売。

［インターネットマガジン　バックナンバーアーカイブ］ ご利用上の注意

このPDFファイルは、株式会社インプレスR&D（株式会社インプレスから分割）が1994年～2006年まで発行した月刊誌『インターネットマガジン』の誌面をPDF化し、「インターネットマガジン　バックナンバーアーカイブ」として以下のウェブサイト「All-in-One INTERNET magazine 2.0」で公開しているものです。

http://i.impressRD.jp/bn

このファイルをご利用いただくにあたり、下記の注意事項を必ずお読みください。

- 記載されている内容（技術解説、URL、団体・企業名、商品名、価格、プレゼント募集、アンケートなど）は発行当時のものです。
- 収録されている内容は著作権法上の保護を受けています。著作権はそれぞれの記事の著作者（執筆者、写真の撮影者、イラストの作成者、編集部など）が保持しています。
- 著作者から許諾が得られなかった著作物は収録されていない場合があります。
- このファイルやその内容を改変したり、商用を目的として再利用することはできません。あくまで個人や企業の非商用利用での閲覧、複製、送信に限られます。
- 収録されている内容を何らかの媒体に引用としてご利用する際は、出典として媒体名および月号、該当ページ番号、発行元（株式会社インプレスR&D）、コピーライトなどの情報をご明記ください。
- オリジナルの雑誌の発行時点では、株式会社インプレス R&D（当時は株式会社インプレス）と著作権者は内容が正確なものであるように最大限に努めましたが、すべての情報が完全に正確であることは保証できません。このファイルの内容に起因する直接的および間接的な損害に対して、一切の責任を負いません。お客様個人の責任においてご利用ください。

このファイルに関するお問い合わせ先
株式会社インプレスR&D
All-in-One INTERNET magazine 編集部
im-info@impress.co.jp