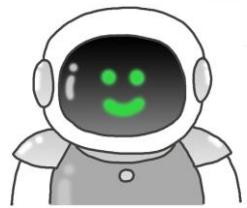


プログラミング教室のテクノロ



プログラミングの 世界の歩き方

「文字コード」



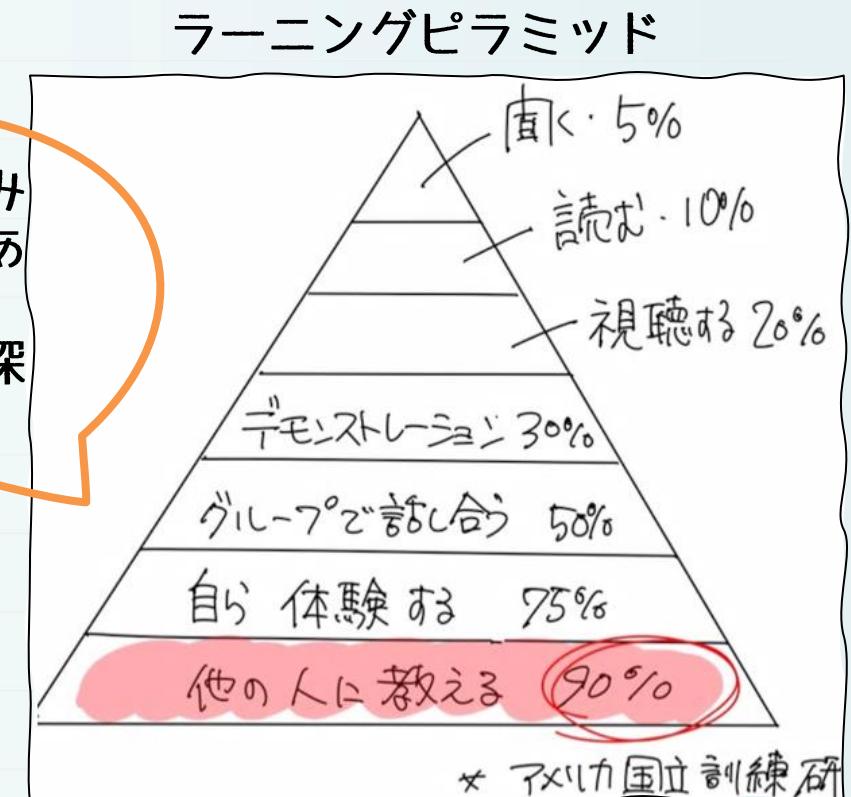
プログラミングの世界を歩こう！

プログラミングの世界を知るにはその世界で使われていることばを知ることが大切だ。

これはプログラミングに限らず、スポーツでも音楽でも何かを習得するには、その世界のことばを知ることから始まるよ。



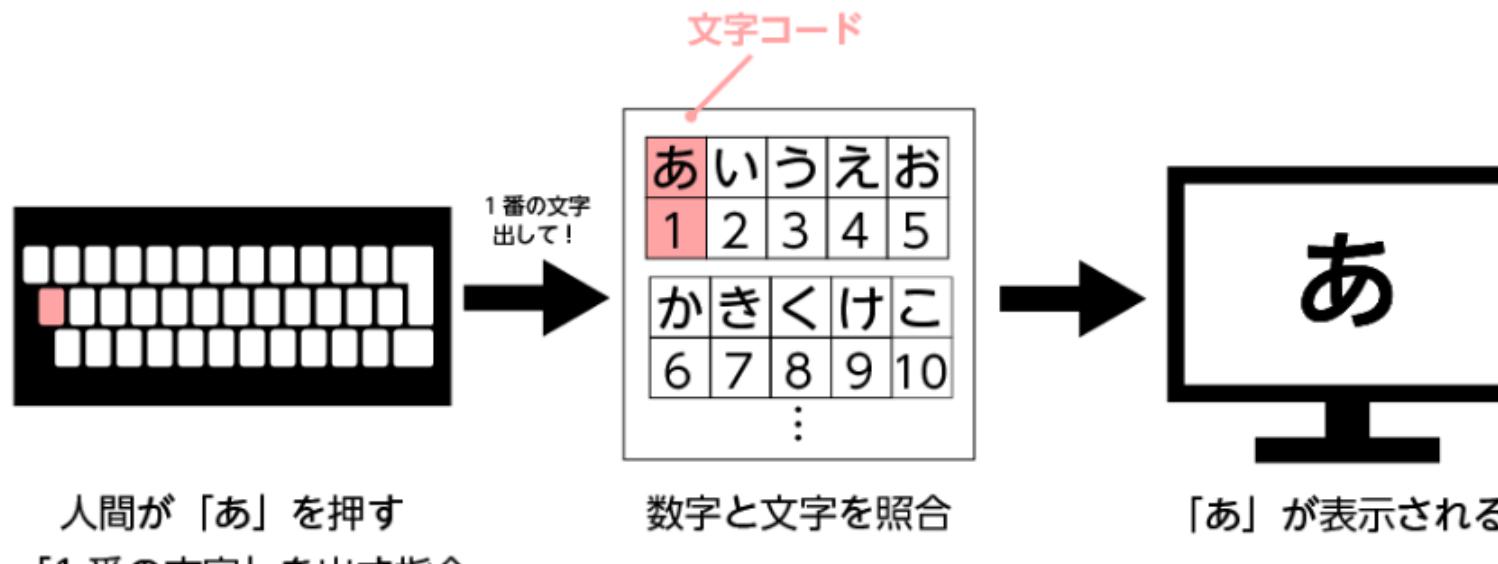
学んだことはみんなに教えてあげよう！
もっと理解が深まるよ！



文字コードって？

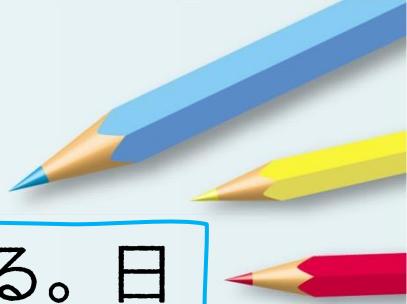


文字コードとは、文字をコンピュータで処理するために文字の種類に番号を割り振ったもの。文字の「背番号」にあたるもの。



コンピュータは「ひらがな」や「漢字」を理解しているわけではない。コンピュータが理解できる言葉は数字（0と1）だけ。つまり、コンピュータ上でひらがなや漢字などを扱うためには、文字に番号を振って対応させる必要がある。

文字コードの種類



ASCII、EUC-JPなど、文字コードにはさまざまな種類がある。日本でよく使われる文字コードは主に2種類（Shift_JISとUnicode）あることを知っておこう。



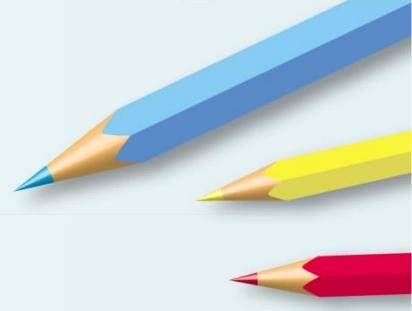
Shift_JIS

Unicode

Shift_JISは日本語に特化した文字コードで「ひらがな」や「漢字」、「カタカナ」などを扱うことができる。WindowsのOSで使われている。

Unicodeは、世界中の文字すべてを網羅する（ことを目標にした）文字コード。現在使われている文字はもちろん、歴史上で使われていた文字も登録されている。

文字コードの種類（参考）



ASCII（アスキー）

米国規格協会によって定められた、基本的な文字コード。含まれる文字はアルファベットと数字といいくつかの記号のみ。1文字を7ビットであらわす。

EBCDIC（エビシディック）

IBM社が定めた文字コードで、8ビットを使って1文字を表す。大型の汎用コンピュータで使われている。

EUC（イーユーシー）

拡張UNIXコードとも呼ばれ、UNIXというOS上でよく使われる日本語文字コード。

文字化けって？



文字化けとは、コンピュータで文字を表示する際に、正しく表示されない現象のこと。

例えば、ひらがなの「あ」は、UTF-8で「E3 81 82」、Shift_JISで「82 A0」、EUC-JPで「A4 A2」と、文字コードによって異なるコード値になっている。

UTF-8をUTF-8で表示

あいうえお 0 1 2 3
a b c X Y Z
アイウオ アイウオ
文字化けパターン

UTF-8をShift_JISで表示

縺ゅ>縺◆.・.縺◆? 撲シ托シ抵シ?
◆?スゑス? ◆?ク?ケ?]
◆?ア?イ?ウ?I?オ ◆?ア?イ?ウ?I?縁ウ
譚◆1怜喧縺代工縁リ縕シ縕ウ

UTF-8をEUC-JPで表示

？◆?◆?◆?◆?◆?◆?◆?◆? 錔◆錔◆錔◆
錔◆
譚◆譚◆譚◆ 錔酣執錔◆
譚縁讐譚鰯輯譚◆ 譚秀譚◆終◆?

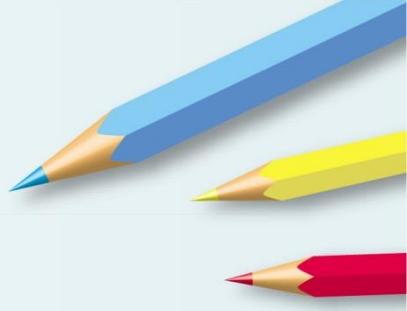
文字	あ
EUC-JP	A4 A2

⇒

文字	、	「
Shift_JIS	A4	A2

文字化けの原因は主に、テキストを「作成した時」と「開いたとき」で文字コードが違うと起こる。

×モ



プログラミング教室の テクノロ



なまえ：