

# もじ 文字コードとUTF-8

★ 文字を数字であらわす ★

コンピュータは文字も数字で記録する。  
世界中の文字を表せるのがUnicodeとUTF-8。

## 1 文字コードとは

文字に番号。

- 文字に番号をつける
- 例 A = 65
- コンピュータは数字で記録



## 2 Unicode

世界の文字を1つに。

- 世界中の文字を1つに
- 絵文字もふくむ
- 「コードポイント」で表す



## 3 UTF-8

可変長の方式。

- 1～4 バイトで表す
- 英数字は1バイト
- 日本語は約3バイト・絵文字は4バイト



## 4 文字数とバイト数

数え方に注意。

- 見た目の1文字とバイトは別
- 絵文字は数え方に注意
- UTF-8が世界標準

たとえば「😊😊」は見た目は2文字



4バイト × 2文字 = 8バイト



UTF-8 =  
世界中の文字を  
表せる方式