

【マルチメディア文化論】

文字データの世界 (半角文字)
2008年度 第5回
太田 信宏



1. コンピュータ内部のデータ表現

コンピュータはあらゆる情報を0と1だけで表現する。

あらゆる情報とは ……

文字、図形、画像、音声、動画のこと
---> 「マルチメディア情報」

文教大学女子短期大学部



2. 0と1の原理

0または1を入れる箱を考える。

この箱の数(けた数)を増やしていくと…

1桁の箱	2桁の箱	3桁の箱	
■	■ ■	■ ■ ■	
0	00	000	
1	01	⋮	
2通り	10	⋮	→ ?通り
	11	⋮	
	4通り	⋮	
		111	さらに、4桁以上の ケースを考えてみる。



3. 箱の桁数と表現できる種類の関係

■	1桁の箱	→	2通り
■ ■	2桁の箱	→	4通り
■ ■ ■	3桁の箱	→	8通り
■ ■ ■ ■	4桁の箱	→	16通り
⋮	⋮		⋮
■ ■ ■ ■ …… ■	n桁の箱	→	2 ⁿ 通り

では、8桁だと何通り? $2^8 = 2 \times 2 \times 2 \cdots \times 2 = 256$ 通り

4. 半角文字の種類

半角文字の種類(数)を考える

英字(大)	A~Z	26個
英字(小)	a~z	26個
数字	0~9	10個
カナ文字	ア~ン	約50個
記号	+ - * ? など	約50個

→ 全部で162個

箱を何個用意すればよい? ■ ■ ■ ■ ……

7桁だと → $2^7 = 128$ 個 …… 足りない
8桁だと → $2^8 = 256$ 個 …… 足りる(余る)

5. ビットとバイト

◇1桁の箱
■ → ビット(bit)という。

◇8桁の箱
1 1 0 0 1 1 0 1 → バイト(byte)という。

1バイト=8ビット という関係が成り立つ。

bit → 記憶の最小単位 (1個の0または1)
byte → 半角1文字を記憶する単位 (8個の0と1)

6. 文字コードのしくみ

◇コード表の見方を理解しよう。

半角文字（JISコード）の例

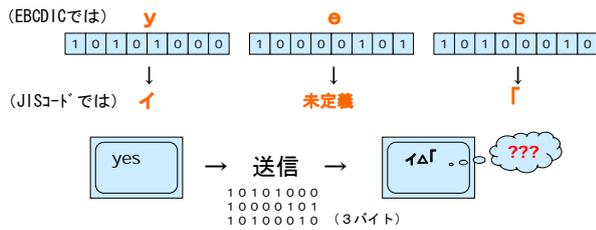
A →	0 1 0 0 0 0 0 1	すべての半角文字は ↓ 並びの異なる 8桁の0と1 で表現される。
B →	0 1 0 0 0 0 1 0	
a →	0 1 1 0 0 0 0 1	
5 →	0 0 1 1 0 1 0 1	
* →	0 0 1 0 1 0 1 0	
ア →	1 0 1 1 0 0 0 1	

7. いろいろな文字コード

- ① JISコード（日本工業規格）
Japanese Industrial Standards
日本で使用されている標準コード（別紙プリント）
- ② ASCIIコード（アスキー）
American Standard Code for Information Interchange
アメリカの標準コード
（JISコードから半角カナを除いたものと、ほぼ同じ）
- ③ EBCDICコード（エビシディック）
Extended Binary Coded Decimal Interchange Code
米IBM社が独自に採用したコード（すでに過去のもの）

8. 文字化けする理由

EBCDICのコンピュータで作成した「yes」
という文字をJISコードで受信すると・・・



→ 文字化けが発生してしまう
(0と1の並びが変わったわけではない)

(参考) [上山清二先生のホームページ](http://www.infonet.co.jp/ueyama/ip/) キャラクターコード
<http://www.infonet.co.jp/ueyama/ip/>