

情報の圧縮

2006年度司書教諭講習
第2回
2006年 8月11日(金) 2時限目

1

thema

データ圧縮

2

データ圧縮

1. 圧縮とは何か
2. 圧縮解凍の原理
3. 起原と歴史

3

圧縮とは何か



4

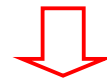
圧縮とは

1. データ圧縮
2. 圧縮率

5

データ圧縮

100kbyte のデータ



50kbyte のデータ

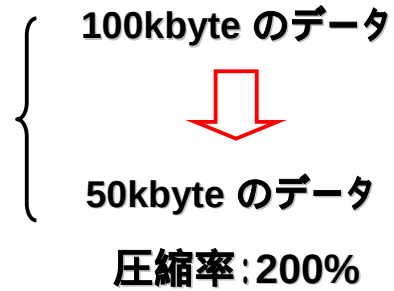
6

圧縮率

元のデータ量を、どれくらい小さいデータ量に圧縮できるかの割合

7

圧縮率



8

圧縮率の考え方

元のサイズの何%に圧縮できたかを示す場合もある

9

1.2. 圧縮の御利益

1. メディアに大量のデータを記録できる
2. データの転送速度が速くなる

10

圧縮法の原理

データ圧縮とは

- 元のデータを別の符号に変換することにより
- 全体のバイト数を減らす方法

11

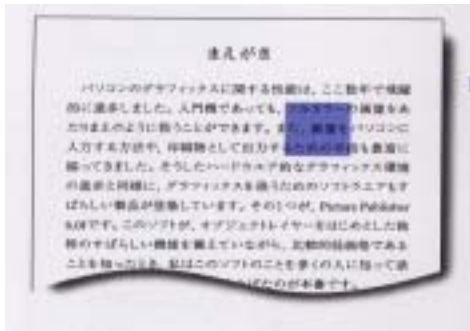
圧縮の方法

元データが持つ

1. 冗長性(繰り返し性)
 2. 法則性(予測可能性)
 3. 不要性
- という性質を利用して圧縮する

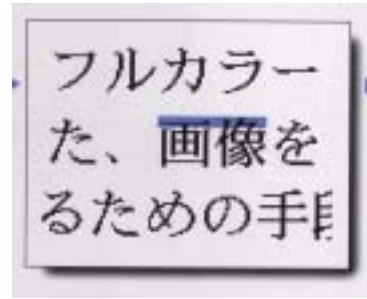
12

圧縮の例



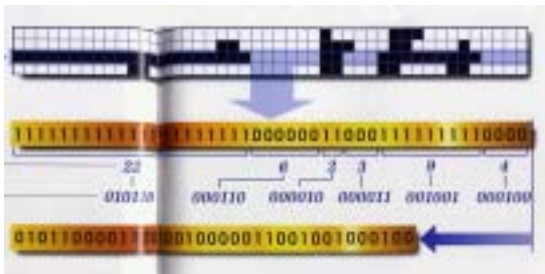
13

圧縮の例



14

圧縮の例 (Run Length法)



15

起原と歴史

1. モールス符号
2. ハフマン法
3. Lempel-Ziv法

16

モールス符号

1. 文字の出現頻度に応じて符号を割り当てる
2. モールス符号の規則

17

出現頻度

1. 高頻度の文字に短い符号
2. 低頻度の文字に長い符号

18

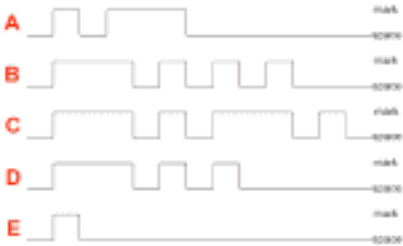
規則

1. 構成は「長点」,「短点」,「空白」の3要素から成立
2. 「短点」の長さ1に対し,「空白」1,「長点」3の長さ
3. 文字間の空白は短点の3倍
4. 単語間の空白は短点の7倍

モールス符号の構成

短点の構成	-							E
長点の構成	—							T
1 符号の構成	—	-	—	-				C
2 符号の構成	—	-	—	-	-	-	-	CQ
2 語の間隔	-	-		-	-	-	-	I arm

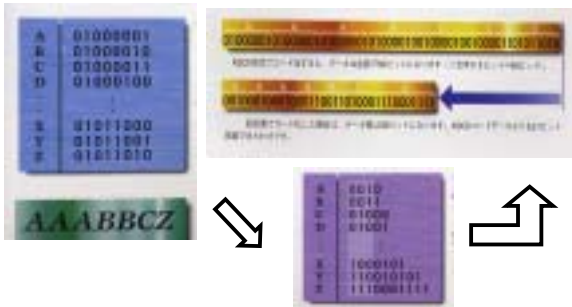
モールス符号



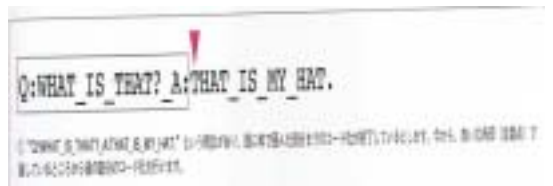
符号の生起確率

文字	生起確率	モース符号
A	0.0642	, -
B	0.0127	- . . .
C	0.0218	- , - .
D	0.0317	- . .
E	0.1031	,
Z	0.0005	- . . -

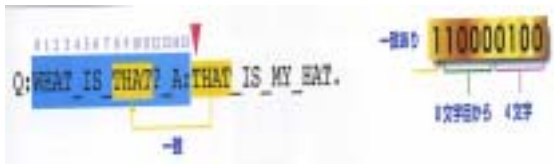
ハフマン法



Lempel-Ziv法



Lempel-Ziv法



25

Lempel-Ziv法



26

Lempel-Ziv法



27

Lempel-Ziv法



28