



情報処理システム論 (22)



⋮

もうちょっと圧縮の話

- 圧縮の利点
 - 時間の節約
 - 磁気テープの読み書き
 - 通信
- モデムに組み込まれた圧縮方式
 - MNP-5
 - V.42bis

•
•
•

MNP-5

- MNP: Microcom Networking Protocol
- 最大圧縮率: 2倍
- 適応型Huffman符号による圧縮

•
•
•

V.42bis

- ITU標準
- 最大圧縮率:4倍
- 辞書ベース圧縮方式

⋮

辞書ベース圧縮方式

- パソコン用圧縮ツールにも利用
 - ARC
 - PKZIP
 - LHarc
 - ARJ

-
-
-

例

- A good example of how dictionary based compression
- *1/1 822/3 674/4 1343/60 928/75 550/32
173/46 421/2*
 - by Random House Dictionary of the English Language, Sec. Ed.

圧縮率

- 辞書の指定は一語あたり20bit
 - 2200ページ(12bit)×256見出し(8bit)
 - 全部で160bit
- ASCII文字で表現すると43バイト
 - 全部で $8 \times 43 = 344$ bit
- 約半分の長さに圧縮

⋮

辞書をどうするのか

- 静的辞書
 - 同じ物を最初から両側で持つ
 - 変更は加えない
- 適応的辞書
 - 辞書を書き換えながら(追加しながら)
 - 新出の単語は最初だけそのまま出力

⋮

・
・
・

移動窓圧縮方式

- 移動窓
 - 数千文字の大きさ
- 先読みバッファ
 - 10～100文字程度
- 電話帳や郵便番号簿のような似た内容が繰り返すような場合に適合

-
-
-

例

窓

先読みバッファ

When there are many words to be compressed in the look ahead



ここ(6文字目)からの3文字と一致し、空白文字が続く

(6,3,' ')

窓の移動

there are many words to be compressed in the look ahead buf

1はないから...

(0,0,'1')

-
-
-
-
-
-
-
-

•
•
•

特許の問題

- 暗号方式にも特許がからむ
- ソフトウェアに関する特許
 - 米国：198x 年から可能になった
 - 日本：原則としては認められていないが...