

平成 10 年度 プロジェクト研究中間発表

**「文字コード自動認識機能付き
ファイル・ビューアの開発」**

指導教官 大岩 幸太郎 教官

**情報教育コース
0710300 永岩 里恵**

【目次】

- 1.はじめに
- 2.シフトJISコードの成り立ち
- 3.日本語EUCコード体系
- 4.ファイルビューア
- 5.これからの研究
- 6.参考文献・参考サイトURL

1. はじめに

日本語の符号化方式

符号化

文字を数値に対応させる処理

→ 対応する数値によって個別の文字を特定することが可能

項目	日本語	英語
バイト数	2バイト	1バイト
コードの種類	JIS シフトJIS 日本語EUC	ASCII EBCDIC

3.日本語EUCコード体系

EUCコード

マルチバイトコードをサポート
日本語の符号化専用ではなく
日本語や、それ以外の言語を
処理するため開発された

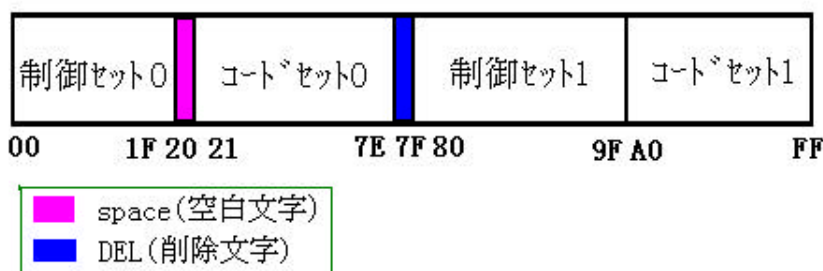


図:EUCのコード範囲と位置

EUCは4つのコードセットで構成

コードセット	第1バイト	第2バイト	第3バイト	日本語EUC
0	21-7E	-	-	ASCII/JIS ローマ字
1	A0-FF	A0-FF	-	JISコード
2	8E	A0-FF	-	半角片仮名
3	8F	A0-FF	A0-FF	補助漢字

表:EUCのコードセット

8E・・・SS2(特殊制御文字)

8F・・・SS3(特殊制御文字)

【日本語EUCコード】

- ・ 可変長コード(圧縮フォーマット)
→ 一般的
- ・ 固定長コード(2バイト固定長フォーマット)

表:EUC 圧縮フォーマットの仕様

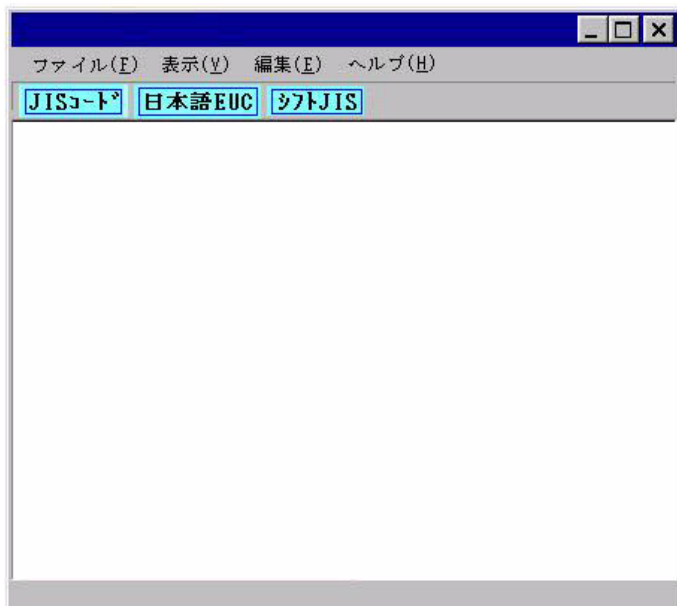
コードセット0 (ASCII/JIS ローマ字) バイト範囲	21-7E
コードセット1 (JIS コード) 第1バイト範囲 第2バイト範囲	A1-FE A1-FE
コードセット2 (半角片仮名) 第1バイト範囲 第2バイト範囲	8E A1-DF
コードセット3 (補助漢字) 第1バイト範囲 第2バイト範囲 第3バイト範囲	8F A1-FE A1-FE

→ 全てのコードセットが
2バイトで表現

表:2バイト固定長フォーマットの仕様

コードセット0 (ASCII/JIS ローマ字) 第1バイト範囲 第2バイト範囲	00 21-7E
コードセット1 (JIS コード) 第1バイト範囲 第2バイト範囲	A1-FE A1-FE
コードセット2 (半角片仮名) 第1バイト範囲 第2バイト範囲	00 A1-DF
コードセット3 (補助漢字) 第1バイト範囲 第2バイト範囲	A1-FE 21-7E

4. ファイルビューア



ファイル

- ・開く
- ・保存
- ・終了

表示

- ・テキスト表示
- ・ダンプ表示
- ・文字化けの修復

編集

コピー

ヘルプ

説明

5. これからの研究

- ・コード変換プログラムを動かす
- ・ファイルビューアの基本的操作の実現
 - ・開く
 - ・終了
 - ・保存

6.参考文献

- (1) 芝野 耕司 編著
『JIS漢字字典』
1997年 財団法人 日本規格協会

- (2) Ken Lunde 著
『日本語情報処理』
1995年 ソフトバンク株式会社

参考サイトURL

1.日本語と文字コード

<http://www.kanzaki.com/docs/jcode.html>

2.文字コードの国際規格について

<http://turbine.kuee.kyoto-u.ac.jp/FAQ/kanji-code.html>

3.インターネットメールの注意点

<http://www02.so-net.ne.jp/~hat/imap/cover.html>