

平成10年度 プロジェクト研究構想発表

「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
コース 情報教育
学籍番号 0710300
氏名 永岩 里恵

目 次

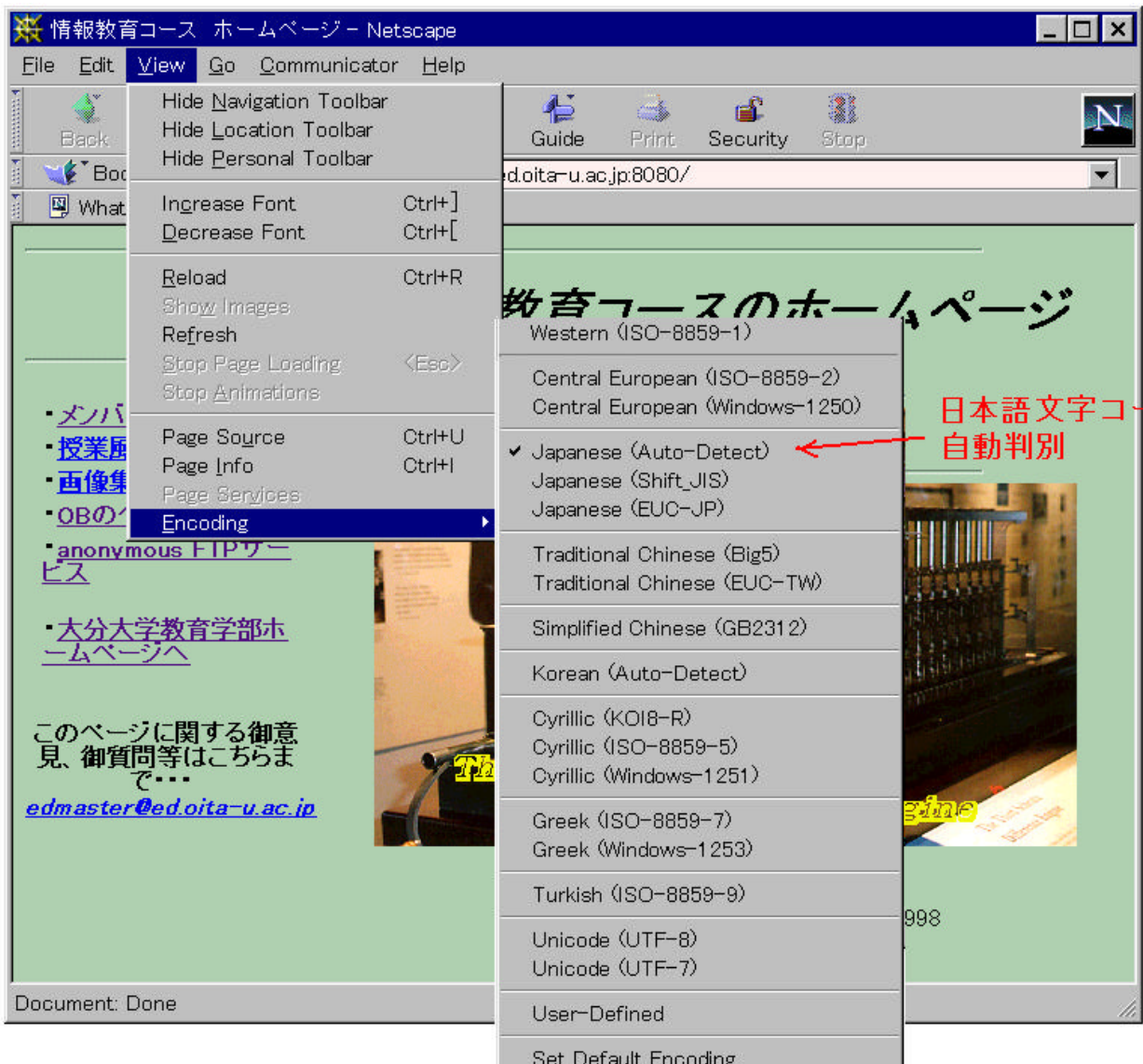
- 1 .はじめに
- 2 .文字コードのしくみと種類
- 3 .ESC シーケンス
- 4 .ファイル・ビューア
- 5 .参考文献
- 6 .参考サイト

1.はじめに

WWWブラウザソフト:

英語、日本語、韓国語等の多言語をカバー

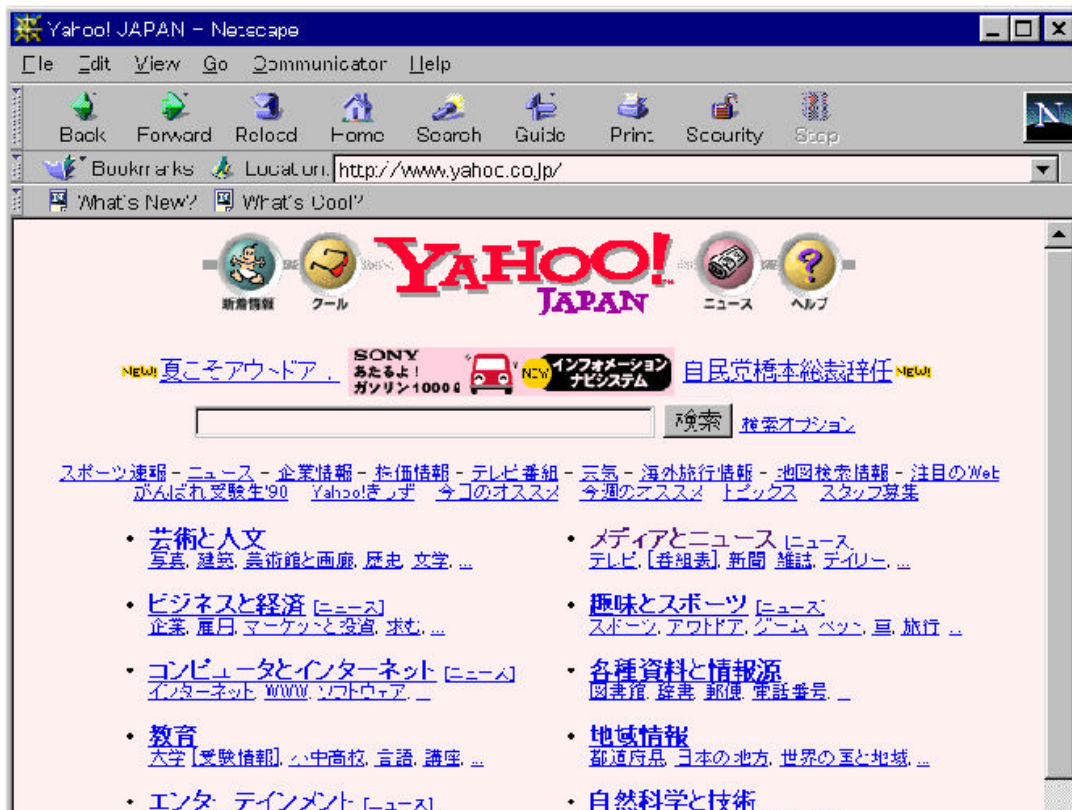
さまざまな文字に対応するため文字コードの自動判別機能が付いている



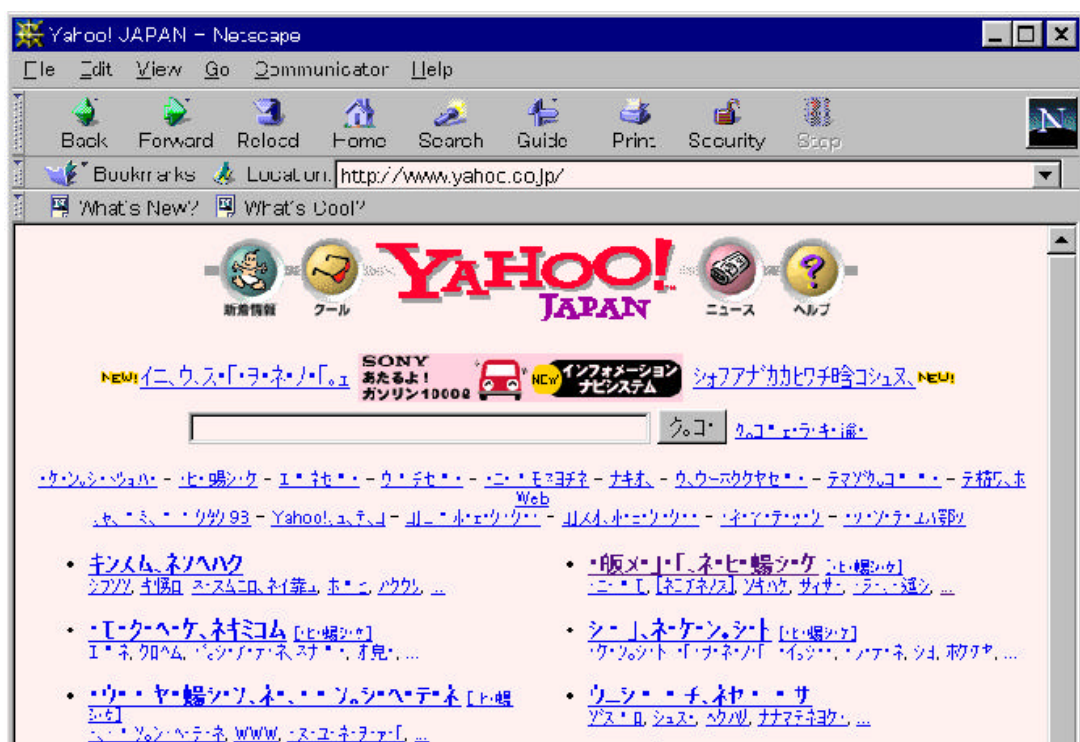
文字化け

データ通信時に、画面上などに意味不明の文字が表示されたり、文字が欠落する現象

(例) Yahooのページ



(図 1) 通常画面



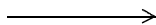
(図 2) 文字化けが生じた画面

mule でのメールのやりとりでも文字化けが起こることがある。

(例) 電子メール

```
TELNET - oita-cc
接続(C) 編集(E) ターミナル(T) ヘルプ(H)
Date: Fri, 17 Jul 1998 17:53:29 +0900
From: nagaiwa rie <nagaiwa>
Return-Path: <nagaiwa>
To: nagaiwa
Subject: sample
Content-Type: text
Content-Length: 139

大分大学へのアクセス方法
大分バス：トキハデパート前 1 番のりば
JR：豊肥線 敷戸駅下車。 徒歩20分。
```



```
TELNET - oita-cc
接続(C) 編集(E) ターミナル(T) ヘルプ(H)
Content-Length: 139

ツ醋ヱツ邊リ、ホ・「・ッ・サ・ケル・ヒ。
ツ醋ヱ・ミ・ケ・ア・ネ・ユ・マ・ヌ・ム。シ・ネチー。。」アネ、ホ、熙ミ。フ・イオ
「ハ」メ。アヒユネ銷。。ノクアアリイシヨ。。。ナヲ釜イ「ハヤ。」

かんざき : Mule: RMAIL (RMAIL)
No following nondeleted message
```

せっかくのメールが読めない

文字化けを修復するURL :

<http://www.kanzaki.com/docs/jis-recover.html#MIME>

→ 文字コード設定可能ファイル・ビューア

・バイナリファイルの表示可能

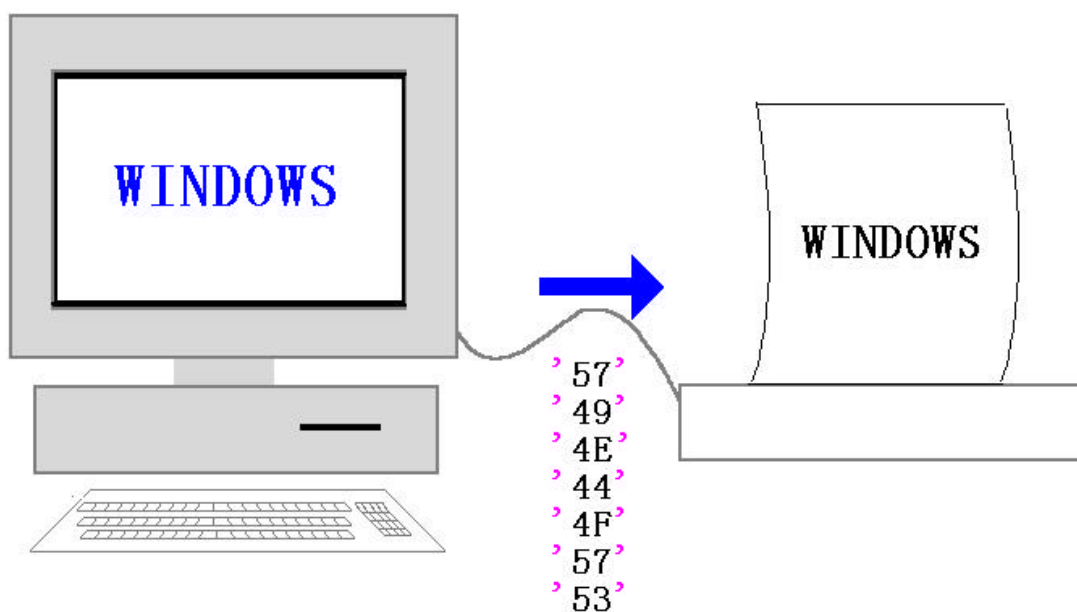
2. 文字コードのしくみと種類

コンピュータ内では文字を符号化

(例) ASCIIコード

		上位4ビット							
		0	1	2	3	4	5	6	7
下位4ビット	0		D _E		0	@	P	'	p
	1	S _H	D _I	!		A	Q	a	q
	2	S _X	D ₂	"	2	B	R	b	r
	3	E _X	D ₃	#	3	C	S	c	s
	4	E _T	D ₄	\$	4	D	T	d	t
	5	E _Q	N _K	%	5	E	U	e	u
	6	A _K	S _N	&	6	F	V	f	v
	7	B _L	E _B	'	7	G	W	g	w
	8	B _S	C _N	(8	H	X	h	x
	9	H _T	E _M)	9	I	Y	i	y
	A	L _F	S _B	*	:	J	Z	j	z
	B	H _M	E _C	+	;	K	[k	
	C	C _L	→	,	<	L	¥	l	l
	D	C _R	←	-	=	M]	m	
	E	S _O	↑	.	>	N	^	n	~
	F	S _I	↓	/	?	O	_	o	

(図) ASCIIコード表



文字コードの種類

主な文字コード

コード体系	説明
ASCII	(American national Standard Code for Information Interchange) 7ビットで128種類の文字が表現可能 英語圏では十分な規格
EUC	(Extended Unix Code) 米 AT&T 社で規定されたマルチバイトコード 各国の文字コードを規定できる
JIS コード	(Japan Industrial Standard) ASCII コードを拡張して、2バイトコードを設定 かな、漢字を割り当てたもの
シフトJIS	(Shifted JIS Code) JIS コードの漢字体系から派生したもの 2バイトで1文字を表す

3. ESC シーケンス

コンピュータが
 41 42 43 44
 を送ると

ASCII コードで解釈・ 41 42 43 44
 A B C D

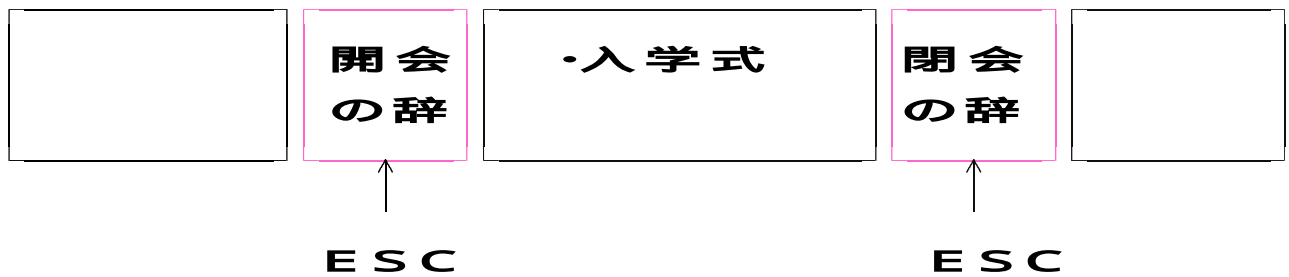
JIS コードで解釈・ 41 42 43 44
 疎 団

ESC を補って文字コードを切り替える
 (ESC は 16 進数で 1B)

(例)

ASCII コード **ESC** JIS コード **ESC** ASCII コード
 41 42 **1B + \$B** 4267 4A2C **1B + (B** 43 44

→ AB **大分** CD



4. ファイル・ビューア

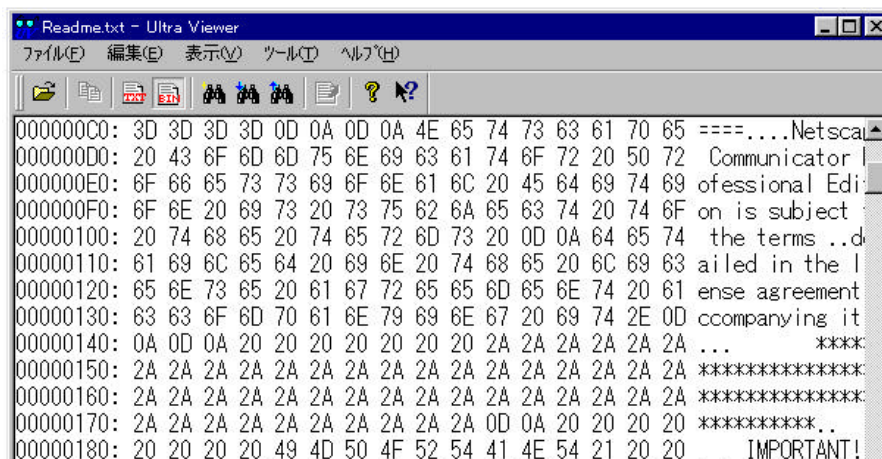
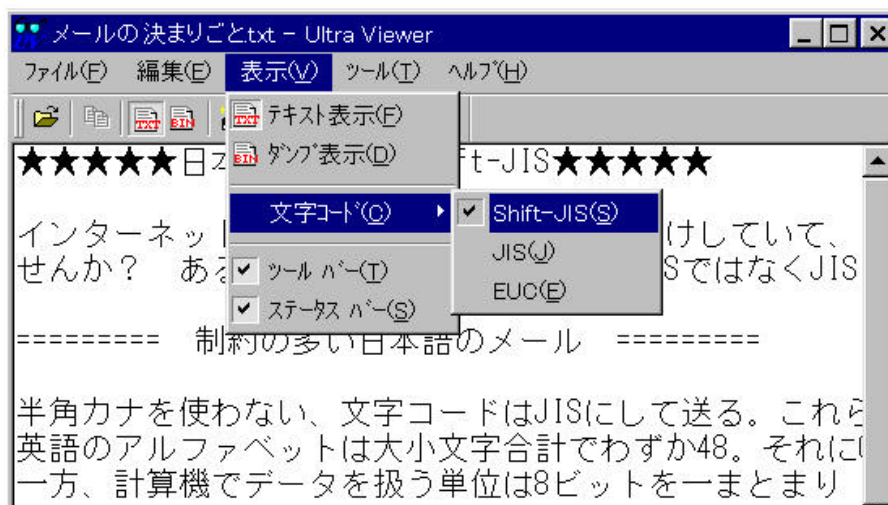
ファイルの内容を閲覧できるソフト

(例) Ultra Viewer

作者 : Kan Tanimura

シェアウェア

URL: <http://www.age-ent.co.jp/UltraViewer/index.htm>



テキスト表示の他に、ダンプ表示が可能
文字コードの設定が可能

文字コード設定可能
バイナリファイルの表示可能
文字化け修復が可能

5. 参考文献

芝野 耕司 編著 『JIS 漢字字典』
1997年 財団法人 日本規格協会

6. 参考サイト

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>