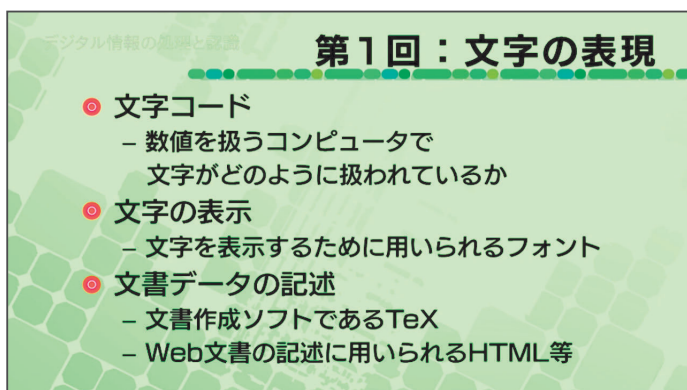


第1回 文字の表現

1



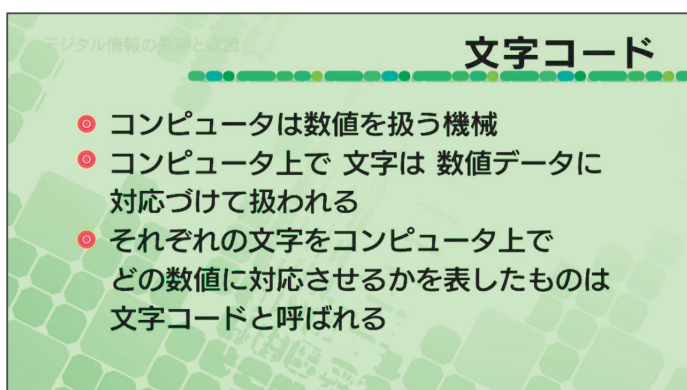
2



3



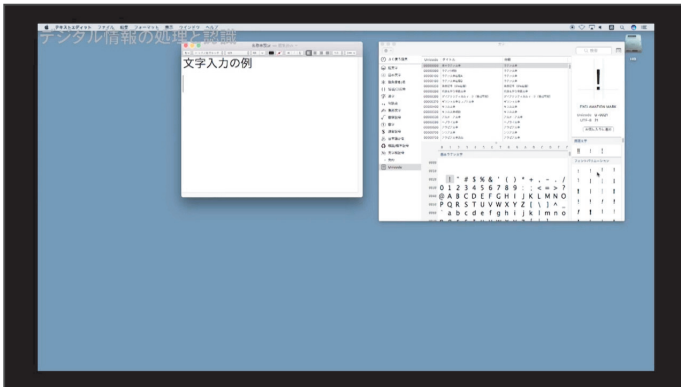
4



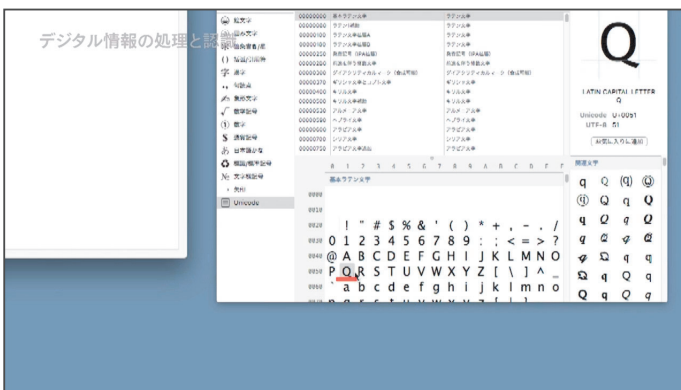
デジタル情報の処理と認識 ('22)

第1回 文字の表現

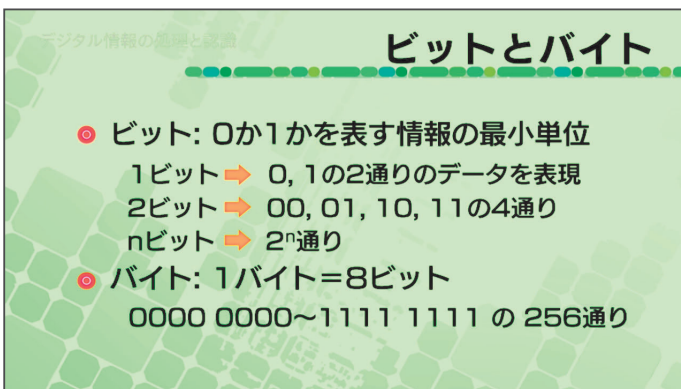
5



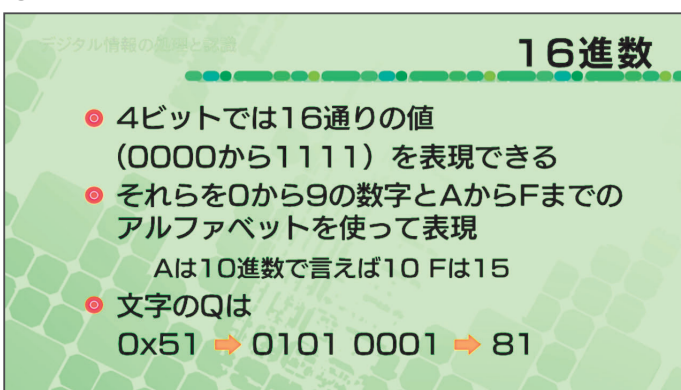
6



7



8



デジタル情報の処理と認識 ('22)

第1回 文字の表現

9

デジタル情報の処理と認識

文字コードの種類

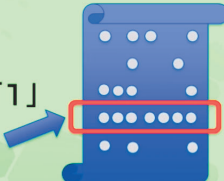
- アスキーコード
 - 英数字を表現
- 漢字コード
 - JISコード シフトJISコード EUC-JP
- Unicode
 - 多言語を表現

10

デジタル情報の処理と認識

アスキーコード

- 英数字を表現するための文字コード
 - 0~31: 改行やタブといった特殊文字
 - 32: スペース
 - 33~126: 英数字
 - ex. 97は「a」 49は「1」
 - 127: デリート (削除)



11

デジタル情報の処理と認識

数値と数字

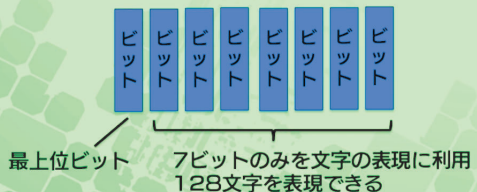
- 数値の5
 - $5 = 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0$
 - 0000 0101
- 数字(文字)の「5」
 - 10進数で表せば53という数値
 - 0011 0101

12

デジタル情報の処理と認識

アスキーコード

- 1バイトコード: 1バイトで文字を表現
- 7ビットコード: 7ビットのみを利用



デジタル情報の処理と認識 ('22)

第1回 文字の表現

13

デジタル情報の処理と認識

半角カタカナ

- アスキーコードで使用しない最上位ビットを利用することでカタカナを表現
- 通常の日本語の半分の幅で表示
- 正しく扱えない場合もある

全角カタカナ ↔ 半角カタ

14

デジタル情報の処理と認識

漢字コード

- 2バイトコード → 約6万5千文字を表現
- 漢字コード
 - JISコード
 - シフトJISコード
 - EUC-JP

15

デジタル情報の処理と認識

JISコード

- 7ビットコード
- エスケープシーケンスを利用

「ESC (B」 「ESC (J」 → 英数字
「ESC \$ @」 「ESC \$ B」 → 日本語

- 「48 33」という2バイトデータ
 - 英数字なら「O!」 日本語なら「亜」

16

デジタル情報の処理と認識

シフトJISコード

- 最上位ビットも使うことでエスケープシーケンスを使用しないように修正
- JISコードの数値を部分的にシフト
- 8ビットコード
- 日本語を表現する場合は 第1バイトの最上位ビットが1



デジタル情報の処理と認識 ('22)

第 1 回 文字の表現

17

デジタル情報の処理と認識

EUC

- 主にUNIXと呼ばれるOSで利用
 - OS：機械部分を直接制御するソフトウェア (Windows macOS UNIX など)
- 日本語を表すEUC-JPは 8ビットコード

18

デジタル情報の処理と認識

Unicode

- 日本語 英語 中国語 韓国語といった多数の言語を1つの文字コードで表現
- UTF-8
 - 可変長バイト
 - 日本語の多くは 3バイトで表現
- Java言語やWebページなどで利用

19

デジタル情報の処理と認識

文字の表示

20

デジタル情報の処理と認識

文字の表示

- 文字コードはそれぞれの文字をコンピュータで扱える数値として表現する
- コンピュータ上で実際に文字を表示するためにはそれぞれの文字に対応する文字の形状やデザインのデータが必要
- このような文字の表示に用いられるデータがフォント

デジタル情報の処理と認識 ('22)

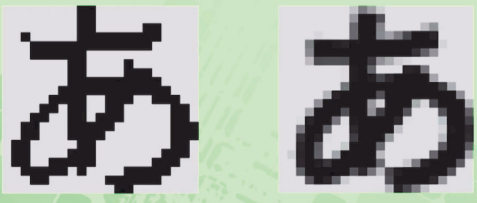
第1回 文字の表現

21

デジタル情報の処理と認識

ビットマップフォント

- 文字を点データの集まりとして表現



ビットマップフォント アンチエイリアス処理

22

デジタル情報の処理と認識

アウトラインフォント

- 文字の輪郭を代表する点とそれらを結んだ線により文字の形状を表現



アウトラインフォント

23

デジタル情報の処理と認識

欧文フォント

- セリフ系 サンセリフ系 筆記体風 etc.



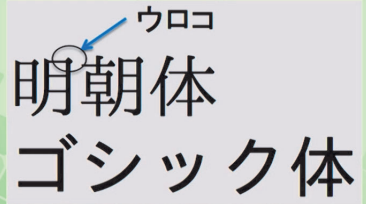
セリフ
Serif
Sans Serif

24

デジタル情報の処理と認識

和文フォント

- 明朝体 ゴシック体 楷書体 etc.



ウロコ
明朝体
ゴシック体

デジタル情報の処理と認識 ('22)

第1回 文字の表現

25

デジタル情報の処理と認識

字 幅

- 和文フォント
 - 等幅フォント
- 欧文フォント
 - プロポーショナルフォント
- カーニング
 - 文字間隔の微調整

AVENUE
AVENUE

26

デジタル情報の処理と認識

文字データの記述

27

デジタル情報の処理と認識

文書の編集

- テキストエディタ
 - テキストデータの編集に用いられる
- ワープロソフト
 - 文字の大きさ 色 レイアウト等の編集も可能
- DTP (Desktop Publishing)
 - コンピュータを使った出版物の編集

28

デジタル情報の処理と認識

文書の編集

- WYSIWYG
(What You See Is What You Get)
 - 表示イメージと印刷イメージが同じ
- 保存形式
 - PostScript
 - 文字や図形をベクタデータとして扱う
 - PDF
 - デザインを変えずに文書を表示する際に利用

デジタル情報の処理と認識 ('22)

第 1 回 文字の表現

29

デジタル情報の処理と認識

TeX

- 特徴
 - 多くのプラットフォームで動作
 - 数式に強い
- 編集方法
 - 文書にコマンドを追加しソースファイルを作成
 - コンパイルして 印刷用のファイルに変換

30

デジタル情報の処理と認識

TeX文書の作成

```

\documentclass{jarticle}
\title{TeX文書の例}
\author{Yoshitomo Yaginuma}
\date{\today}
\begin{document}
\maketitle
\section{はじめに}
TeXは多くのプラットフォームで動作するフリーソフトです。
\section{数式の例}
\begin{equation}
\int_0^1 \sqrt{1-x^2} \, dx = \frac{\pi}{4}
\end{equation}
のように、数式の作成に強いといった特徴があります。
\end{document}

```

31

デジタル情報の処理と認識

TeX文書の作成

① ソースファイル

② コンパイル等

③ 作成された文書

32

デジタル情報の処理と認識

TeX文書の作成

TeX 文書の例

Yoshitomo Yaginuma

平成 29 年 6 月 19 日

1 はじめに

TeX は多くのプラットフォームで動作するフリーソフトです。

2 数式の例

$$\int_0^1 \sqrt{1-x^2} dx = \frac{\pi}{4} \quad (1)$$

のように、数式の作成に強いといった特徴があります。

デジタル情報の処理と認識 ('22)

第1回 文字の表現

33

デジタル情報の処理と認識

World Wide Web

- インターネット上で文書等の情報を共有するための仕組み
- 1989年に CERNで開発
- 閲覧には Webブラウザが必要
– Microsoft Edge Safari など
- 画像や映像の表示も可能

34

デジタル情報の処理と認識

HTML

- Webページの記述には HTML
(HyperText Markup Language)
と呼ばれる言語が利用される
- <>で囲まれたタグによって
構造やレイアウト等が指定される
- 拡張子は、[.html]
(トップページは index.html)

35

デジタル情報の処理と認識

Webページの公開

- 公開にはWebサーバが必要
- URL(Uniform Resource Locator)
– Webページの所在を表わす文字列
<http://www.ouj.ac.jp/index.html>
(放送大学のホームページ)

36

デジタル情報の処理と認識

HTMLの例

```
<html>
<body>
<h1>HTMLのサンプル</h1>
<p><a href="http://www.ouj.ac.jp/">OUJホームページへのリンク</a></p>
<p><a href="movie1.mov">movie1.movをダウンロード</a></p>
<p><a href="movie1.mov"></a></p>
</body>
</html>
```

デジタル情報の処理と認識 ('22)

第1回 文字の表現

37



38

XML

- HTML
 - タグによってその属性を表現
 - 文字だけで記述
 - マークアップ言語
- XML
 - マークアップ言語を作成するための言語 (メタ言語)

39

タグを利用するための規則

- <tag>というタグは </tag>で終了する
- タグは入れ子になって良い
- ```
<tag1><tag2></tag1></tag2>
```

 のようにタグが交差してはいけない
- <tag></tag>は、1つにまとめて <tag/> と書くことができる

40

### XHTML

- HTMLでは 厳密にXMLに準拠していない部分もある
  - 改行を表す<br>タグは 終了タグ無しに 単独で改行位置を指定できる
- XHTML
  - HTMLを厳密にXMLに準拠させたもの
  - <br>タグは <br></br> または <br/>

# XMLの利用

- BML
  - デジタル放送の番組関連情報などの配信
- SVG
  - 2次元ベクトル画像を記述
- SMIL
  - 映像や音声のレイアウトや再生時間を記述
- X3D
  - 3次元データを記述

# まとめ

- 文字コード
  - アスキーコード、漢字コード、Unicode
- 文字の表示
  - フォント
- 文書データの記述
  - TeX、HTML、XML