



データ形式と演習

ガイダンス資料

担当: 宮川 治、小濱 隆司

[ファイル(file)]

- コンピュータのファイルシステムには、ファイル(file)やディレクトリ(directory、別名: フォルダ(folder))というものがある。ファイルとはハードディスク等に保存されたデータを一まとまりにしたものである。

[ファイルなど]

- ファイル(ディレクトリを含む)
 - サイズ
 - 最終更新日時
 - 種類(ファイル／ディレクトリ)
 - ファイル名

[ファイルの内容]

- バイト・データの並び。ただしそのデータを制限して、区別することで色々なものに代わる。
 - バイト: 一般的には、1ビットという単位が8つ並んだもの
 - ビット: 2進数の1桁で0か1である

[1桁の16進数(4ビット)]

10進数	16進数	2進数
0	0x0	0b
1	0x1	1b
2	0x2	10b
3	0x3	11b
4	0x4	100b
5	0x5	101b
6	0x6	110b
7	0x7	111b
8	0x8	1000b
9	0x9	1001b
10	0xA	1010b
11	0xB	1011b
12	0xC	1100b
13	0xD	1101b
14	0xE	1110b
15	0xF	1111b

表記ルール

- ・10進数: 何もつけない
- ・16進数: 先頭に「0x」をつける
- ・2進数: 最後に「b」をつける

[ファイルの分類]

■ 分け方

○ 直接実行できるものと、それ以外

- java.exe(実行できる)
- *.java(ソースコード、実行できない、テキスト)

○ バイナリ、テキスト

- バイナリは普通の人間には意味を解釈できません。2進化されたファイル内部表現の形式である。表記する場合、バイトデータ(8ビット)を16進数の2桁で表現すると見やすい。
- テキストファイルは人間がその内容を見ると文字や数字で構成されている。

[バイナリとテキスト]

- コンピュータが扱うすべてのデータはバイナリ(バイトの並び)であり、テキストもバイナリ的一种ではあるが、通常バイナリとテキストは区別される。

[テキスト・ファイルの分類]

- 文字コードの違いによる分類
 - Javaソースコード
 - Windows-31J (MS932) or Shift-JIS
 - CSVデータ
 - Windows-31J (MS932) or Shift-JIS
 - JSONデータ
 - UTF-8

[文字コードとは]

- コンピュータでの文字の処理や描画には文字コードやフォント(font)が関係している。
- 文字コードの素朴な理解は、文字のそれぞれに番号が振られ、コンピュータの内部ではその番号を処理することです。

【 目的1

- テキストデータをプログラムで処理するには、そのテキストデータの形式(フォーマット)を知る必要がある.
- 使用頻度が高いデータ形式にはCSV (Comma-Separated Values), JSON (JavaScript Object Notation)とXML (Extensible Markup Language)等がある.

[目的2]

- 本授業では、幾つかのデータ形式を取り上げ、その内容や、種々のデータ形式を処理するためのライブラリーの使用方法を学ぶ。演習ではデータ形式の変換など実践的なコーディングを体験する。

[テキストデータの形式 (CSV)]

■ CSV (Comma-Separated Values)

```
id,name,score  
17OB001,小濱,80  
17OB002,高野,95
```

[テキストデータの形式 (JSON)]

■ JSON (JavaScript Object Notation)

```
{  
  "students": [  
    {  
      "id": "17OB001",  
      "name": "小濱",  
      "score": 80  
    },  
    {  
      "id": "17OB002",  
      "name": "高野",  
      "score": 95  
    }  
  ]  
}
```

[用意するもの]

- ノートPCは必要
- JDK8.0 (Javaの開発環境)をインストールしてください
- エディタのインストール

コンピュータプログラミングⅠ・Ⅱ・Ⅲとの関係（パラダイムの違い）

- 構造化（コンピュータプログラミングⅠ・Ⅱ）
 - メインとなる処理では大まかな処理を記述し、振る舞いによって細部を記述していく方法
- オブジェクト指向（コンピュータプログラミングⅢ）
 - 複雑さを隠蔽する技術

コンピュータプログラミングⅠ・Ⅱ・Ⅲとの関係(文法)

- コンピュータプログラミングⅠ・Ⅱ
 - クラスメソッド
 - シグネチャ(メソッドの名前、仮引数の数と型、返却値の型など)
- コンピュータプログラミングⅢ
 - インスタンス(オブジェクト)メソッド
 - 仮引数と返却値の型にインスタンスの型が増える

コンピュータプログラミングⅠ・Ⅱ・Ⅲとの関係

■ パラダイム

- 構造化、オブジェクト指向を両方使用しコーディングを行います。

■ 文法や考え方

- 可能な限り、新しい文法や考え方を取り入れます。
 - 拡張for制御文など

[コーディング(プログラミング)]

- 授業ではコーディングを行います。なので、好きになってください。
- 大量のデータを手作業で処理しない方法を学びます。

[実施方法]

- 毎回、課題があります。コンピュータプログラミングⅢと同様

[評価方法]

- 課題 40%
 - 授業中のプログラムを提出します。
 - 小テストを行います。
- テスト 60%

[URL

■ 授業URL

- <https://usd.sie.dendai.ac.jp/programming/2020/f/>

■ 代替サーバ

- <https://usdaj.github.io/f/>