

インターネット基礎理論 08 マークアップ言語

はじめに

資料置き場

<https://sammyppr.github.io>

に授業資料を置いていきます。復習に使ってください。

欠席した場合などはスライドを確認して追いつくようにしましょう。

スライドあるなら授業受けなくていいや

なんてことは思わないようにお願いします。



今日すること

今日は、ネット上でやりとりされるデータ自身について学びます。
また、その中で最近良く使われているマークアップ言語について学びます。
その前に前回までの内容を軽く復習しましょう。

TCP/IP構造と通信

わりといい資料あがってたので紹介しておきます。
資格を取る人などは、この辺理解しておきましょう。

- [TCP/IP構造と通信](#)

データをやりとりするということ

ファイルの種類について

第四回目においてファイルの種類について

- テキストファイル
- バイナリファイル

の2種類があることを学びました。

いずれにしても、そのデータは0,1の組み合わせで作られています。

データをやりとりするということ

データをやりとりする、というのは単純に言えば、この

データ化されたファイルをやりとりする

と考えて良いかもしれません。

もちろん、ファイルではなく、ただのデータをやりとりすることもあるのですが、ファイルをデータの塊と考えれば同じ事になります。

元をただせば...

テキストファイルもバイナリファイルも2進数の塊のデータであることをもう一度確認しておきましょう。

ということは、

そのデータの塊をTCP/IPを利用して送っているにすぎないということになります。0,1の連続したデータですよ。

どういう形式で送るか

データにはいろいろな用途がある

データにはいろいろな用途があります。

その用途に応じたデータ形式を考える必要があります。

さて、

用途に応じたデータ形式

とはなんですか？

用途に応じたデータ形式

通信を考える以上

- 容量が小さいほうがよい

というポイントが一つあります。

また、いろいろな場所で利用されることを考えると

- 利用しやすいデータ形式

であることが求められます。

用途に応じたデータ形式

この2つは相反するものではありませんが、インターネット上ではデータをいろいろな場面で利用することが多いため、どちらかと言えば

- 利用しやすいデータ形式

の方が利用される傾向にあります。

これは、回線が多く、情報を速く利用することが可能になったことも大きな要因となっています。

バイナリデータ

これに関しては、様々なフォーマットがあり(動画・画像・音声等)基本的に数字の羅列のため、それをそのまま送ることになります。

テキストデータ

例えば、2桁の数字をデータとすることを考えてみます。

45

という数字があったとします。これを、16進数で表せば

2D(16進数)

となり、1bitで表現することができるのですが、これでは**人が認識するには翻訳が必要**です。

普通にテキストデータとして

45 16進数だと34 35

と書いてしまうと、2bit必要になり、情報量としては倍になってしまいます。

読みやすいデータが主流に

しかし、データを受け取ったクライアント（コンピュータ）としても、それを処理する人間にとっても読みやすくなります。

現在インターネットでは、数字だけでなく、文字情報も扱うため、基本的には読みやすい形式が多く利用されています。

データフォーマット

実際のデータ例

もう少し複雑なデータについて考えてみましょう。

例えば、

- 名前: 小林 統
- 年齢: 56
- 生年月日: 1969年10月30日
- 住所: 渋谷区
- 身長: 176

という個人情報のデータがあったとします。(晒してしまった...)

これを、データとして取り扱うにはどのようにすればよいのでしょうか？

CSV

これは

小林 統,56,1969年10月30日,渋谷区,176

というデータで表現することができます。

こういう,(コンマ)で分割された1行のデータを複数持つデータ形式を

- **CSV (Comma-Separated Values)**

とよびます。

CSVの問題は？

仮に100歳以上だったとして、

小林 統,105,1969年10月30日,渋谷区,176

であった場合、どれがどの数字だかわかりづらいですよ？体重かもしれません。

そのため

氏名,年齢,生年月日,住所,身長

小林 統,105,1969年10月30日,渋谷区,176

等とそれが何を意味するか併記した方がよりわかりやすくなります。

実際のCSVデータを見てみよう

気象庁にあるCSVデータを見てみましょう。

- [過去の気象データ・ダウンロード](#)

CSVの問題1

csvでは

- そのデータがどのような内容だか利用する側も知っている
- 1行目等にデータの属性を明示しておく

のどちらかが必要となります。

データが来た時に、どのように記載されているかわからないと困りますよね。
これは、いろいろな人がそのデータを利用するにおいては非常に不便です。

CSVの問題2

テキストファイルには文字コードがあります。
ですので、それを特定する必要があります。

うまくいかないと文字化けしたデータを処理することになってしまいます。

csvの問題3

csvでは改行ごとに一つのデータ、として扱うことになっているため、改行を含むデータの取り扱いに関しては、それぞれのデータの作り手に委ねていました。そのため、結局、使いづらいデータ形式となっていました。

結局、データを実際にみて、このcsvではこのように扱っているんだ、と判断しなくてはいけなかったわけです。

不具合をおこすCSV文書。

Excelで普通に見えるものをCSVにして、テキストエディタで見てください。

(メモ)

- CSV方言.xlsを「CSV UTF-8コンマ区切り」で書き出してテキストエディタで確認
- CSVを右クリックでExcelで開いた時を確認(正常)
- CSVを「ファイル-開く」から開いた時を確認(意味不明)

Excelの開き方でうまくいったりいかなかったり意味わかりませんね。

ExcelのCSVが事実上標準？

CSVの定義では

一括りのデータを、1行にコンマで区切って記載する

ということしか決まっています。

それにExcelが独自の解釈を加えたものが利用されることが多いです。

それでも、1行目等にデータの属性を入れる場合と入れない場合は作り手に委ねられています。

TSV

コンマの代わりにTABという目に見えないコードを利用することもあります。

TABで分割された1行のデータを複数持つデータ形式を

TSV (Tab-Separated Values)

とよびます。

マークアップ言語

マークアップ言語

CSVの不備をなくすために、**マークアップ言語**とよばれるデータ形式が開発されました。代表的なものは

XML Extensible Markup Language

であり、先ほどのデータは

```
<person>  
  <name>小林 統</name>  
  <age>56</age>  
  <birthday year='1969' month='10' date='30' />  
  <address>渋谷区</address>  
  <height>176</height>  
</person>
```

と表記できます。

マークアップ言語とは

文章を構造化する

と言う役目をになっています。

- ここは特に注目してもらいたい情報です
- ここからここまでが1つの意味のかたまり = 段落です

等をタグ・記号等を利用してコンピュータにわかりやすく伝えるための言語となります。

コンピュータに挙動を指示する「プログラミング言語」とは異なり、文章に意味を持たせるのが「マークアップ言語」となります。

XMLの衰退？

1998年に決められましたが、現在では「XML オワコン」と言われるほど利用されなくなってきています。

- [Google Trend XML](#)

理由は、

- わかりやすいが、サイズが大きくなりすぎる

というものです。

ただし、現在のフロントエンドの開発の技術Reactでは似たようなJSX(Javascript XML)が利用されています。

目立たないところではまだXMLは健在

皆さんがよく利用するOfficeのファイル形式は

- docx
- xlsx
- pptx

ですが、この末尾の「x」はXMLから来ています。

Office Open XMLという形式であり、複数のXMLファイルを圧縮してzip化したものが docx, xlsx, pptxの招待となります。

実際に見てみましょう。

RSS これは終わりそう?...

ニュースの内容などをXMLで配信しています。これをRSSと呼びます..が、これは終わるかな??

現在ではRSSリーダーとしては[Feedly](#)が有名です。登録しておく、サイトを横断して記事が読めるので便利です。

Yahoo News RSS一覧

- 衰退したRSS、今でも細々と使われています。
- [【2024年版】RSSリーダーおすすめ一挙紹介！メリットや基礎知識もセット解説](#)

XMLが衰退して、何に変わっているの？

jsonと言う形式が最近では人気です。JSONの方がシンプルであり、開発言語のJavaScriptでの利用が簡単なことから利用されています。

JSON JavaScript Object Notation

- [JSONとXMLの比較](#)
- [Google trend json](#)

AIへの指示を出す際の「構造化データ言語」としても活用されています。

HTML

HTML

ウェブページを作成するために開発されたマークアップ言語が

- **HTML** HyperText Markup Language

となります。見た目は以下のようにXMLと似ています。

```
<!DOCTYPE html>
<html>
  <head>
    <title>ホームページだよ</title>
  </head>
  <body>
    <p>本文はここに書くよ</p>
  </body>
</html>
```

HTMLが開発された経緯

CERN（欧州原子核研究機構）という研究所にいたティム・バーナーズ・リーさんは研究所に存在している膨大な論文を簡単に管理できないか？と考え、HTML(Webサーバ・ブラウザ含む)を生み出しました。

- 世界中の情報整理を後押ししたHTMLの進歩

論文は論理的に記述された文章であるため、HTMLも同じく章立てなど論理的に記述することが得意なマークアップ言語です。

実際のHTMLソース

大学のホームページのソース(HTML)をみてみましょう。

HTMLのバージョン

現在利用されているのは

- HTML Living Standard

と言うバージョンですが、ちょっとわかりづらいので

- HTML5

と呼ばれたりもします。が、現在厳密にはHTML5が存在していないことは覚えておきましょう。

それでは実際に試してみましょう。

それでは、実際にhtmlを書きながらどのように表示されるか試していきます。こんな感じで書くのか...と言うことを体感してもらえれば幸いです。

```
html,head,body
```

```
h1...h6,p,br,hr,img,a,ul,ol,dl,
```

```
table,img,div,
```

等代表的なものを試していきます。

メディア文化コースの「情報デザイン演習II」はHTMLを詳しく学んでホームページを制作する演習となります。

え？あじけないって？

HTMLにて、どのように表記するかをみてもらいましたが、見た目、全然キレイじゃないですね。(阿部寛のホームページみたい...)

次回は、どのように装飾するか、また、どのようにインタラクティブなページを作るかについて説明していきます。

Markdown

マークダウン

HTMLはテキストを囲んだりして定義します。記述する際に少し面倒です。

そのため、もう少し簡単に構造を定義できないか？と考えられたのが「マークダウン」です。

見出し1

見出し2

- 箇条書き項目1
- 箇条書き項目2

などと書くので簡単にメモを取ることができたりします。このスライドも実はマークダウンで記載されています。

マークダウンの普及

マークダウンは元々エンジニアが文章を記述する時によく利用していました。

マークダウンの特徴としては

- **圧倒的な書きやすさ** シンプルな記号を使うだけ
- **高い可読性と互換性** どのテキストエディタでも読め普遍性がある
- **リッチな出力への変換** PDF,HTML,Wordなどに変換可能

があげられます。

ビジネス・一般への利用が現在普及中です。

皆さんが利用しているTeams,Discordなどでも利用ができます。

そして、AIにおいても利用されています。

ChatGPTの結果をコピーして見てみましょう。

AIとHTMLとMarkdown

AIによるアウトプットとして、論理的なフォーマットが必要ですが、

- HTML
- Markdown

どちらを使うべきかちょうど現在揺れています。

- AI時代のドキュメントは「Markdown」から「HTML」へ？ AI活用事例から学ぶ、見やすく伝わる資料の作り方

小レポート

マークアップ言語とは何か、簡潔に説明せよ。(manaba)