

情報社会論

(252S0541)

講義概要

第9回目

WWW (World-Wide Web)

WWWとは、ドキュメントに別のドキュメントの URL への参照（ハイパーリンク）を埋め込むことで、インターネット上に散在するドキュメント同士を相互に参照可能にするシステムである。

9 WWW

WWW(World-Wide Web)は、インターネット上に散在するドキュメント同士を相互に参照可能にするシステムである。ドキュメント（ウェブページ）の記述には、別のドキュメントの URL への参照を埋め込むハイパーリンクという方法が可能であり、主に HTML や XHTML といったハイパーテキスト記述言語が使用される。現在、WWWはインターネットの中心的概念となっているが、その確立は比較的最近である（1986年以降とされる）。

9.1 WWW 以前のプロトコル

- ARPANET 実験ネット開始 (1969年)
- telnet RFC318 (1972年)
遠隔の端末から、ホストにログインして、ホスト上でのテキストベースの作業を行う。
- ftp RFC354 (1972年)
遠隔のコンピュータとのファイルの転送を行う。
- mail 関係 RFC821,822 (1982年)
遠隔のホスト上のユーザーに対して、主にテキスト文書を送る。
- DNS RFC882, 833 (1983年)
遠隔のコンピュータを名前で認識する。

9.2 WWW のルーツ

- ハイパーテキストのアイデア：Vannevar Bush は「As We May Think(Vannevar Bush 1945年)」⁶³で仮想的な機械 memex⁶⁴を提案した。その説明を以下に引用する。

未来での個人が使う道具を考えてみましょう。それは、機械化された個人的なファイルと図書館といったものです。名前が必要なのでそれを仮に“memex”としてみましょう。memexには、個人がその中に、彼の持っている本や、記録、情報すべてが格納されています。そして、非常に高速にかつ柔軟にそれらを調べられるように作られていて、彼の記憶を拡大するための使いやすい補助的道具になります。

今直ちに必要ではない事柄色々については忘れても良くて、そしてもしそれが重要と思ったときは、それを再び思い出せることができるならば、考えを色々と思い巡らせることは、もっと楽しいことになるでしょう。

⁶³URL <http://www.isg.sfu.ca/~duchier/misc/vbush/> で読むことができる。

⁶⁴MEMory EXtention

もちろん、登録されたデータは索引によって調べることができます。もし、ユーザーがある本を探したければ、そのコードをキーボードで入力すると、タイトルページが直ちに表れ、目の前に表示されます。頻繁に使うコードは簡易記憶されているので、コードブックはめったに参照しません、使うときでもキーを一つただけで、参照することができます。また補助的なレバーもあって、それを右に回すと、目の前の本の各ページがちらっと見えながら順次先に進みます。また、10ページずつでも100ページずつでも一度に進むこともできます。左に回すと、逆方向に操作することができます。

ここで提案されている仮想的な機械とは、リンク機能を持つ装置、つまり、1つの文書を通して読むのではなく必要に応じて様々な文書へと移動できる機能を持つものである。

- NLS (oN-Line-System)

1968年、Douglas C. Engelbart らのチーム⁶⁵はNLSを作成し、Demoを行った。それは、現在のコンピュータの持つ主要な機能を備えていた (URL <http://dougengelbart.org/> を参照)。

ウインドウ画面、マウス、マウスクリック等によるリンク機能、グループウェア⁶⁶の始まり。

- Hyper Text

Ted Nelson は、1965年 “Computers, Creativity, and the nature of the written word” と題する Vassar College での講演で、Hyper Text という言葉を使った⁶⁷。

彼は、Xanadu プロジェクト (1960年)⁶⁸を立ち上げ、彼の意味でのハイパーテキストの実現を試みた。

9.3 html と WWW

マークアップ言語：

markup とは、印刷用語ではフォントや組み指定の意味、markup 言語とはそれらのフォントや組み指定を文書の中に文字として書き込む文書 (言語)。

- SGML (Standard Generalized Markup Language)

IBM の GML (Generalized Markup Language 1969年) を発展させ、ISO 規格 (8879) として制定 (1986年)。SGML は異種のコンピュータ間で文書の互換を行うためのもので、文書のもつ「章見出し」「節見出し」「本文」などの論理構造を記述することができる。タグ⁶⁹は文書の論理構造に対応し、その構造や属性は DTD (Document Type Definition) と呼ばれるファイルに記述されるようになっている。DTD に基づいてレイアウトを行えば、異種のコンピュータでも文書の論理構造をそのまま再現できる。

- HTML (Hyper Text Markup Language) :

Hyper Text を記述するための Markup 言語、Tim Berners-Lee は1989年、HTML を用いて、CERN⁷⁰で WWW システムを作った⁷¹。1993年 NCSA⁷²はこれを受け継ぎ、画像・音声を使える Mosaic を開発。

1994年 Tim Berners-Lee らにより **W3C (World-Wide Web Consortium)** が設立され、1995年 HTML2.0 (RFC1866)、1996年 HTML3.2、1997年7月に HTML4.0 の草案 (Working

⁶⁵Augmentation Research Center at Stanford Research Institute, <https://ja.wikipedia.org/wiki/NLS>

⁶⁶ネットワークを利用での共同作業支援システム

⁶⁷彼は、“Literary Machines(1982年)”の中で、ハイパーテキストについて詳説している。

⁶⁸URL <http://www.xanadu.net/>

⁶⁹タグとは、<〇〇>と</〇〇>のような記号列を指し、スラッシュ「/」のないタグとスラッシュのあるタグで囲むことで、囲んだ文字列を修飾する。スラッシュのついていないタグが命令の始まりを示し、スラッシュのついたタグが命令の終わりを示す。タグで囲んだ範囲が、その命令の影響範囲になる。

⁷⁰Conseil Europeen pour la Recherche Nucleaire (欧州原子核共同研究機関)、「せるん」と呼ばれている。

⁷¹彼はこの業績によって、2002年日本国際賞 (計算科学・技術分野) を受賞した

⁷²National Center for Supercomputing Applications : 米国立スーパーコンピュータ応用研究所

Draft) が公開され、同年 12 月には正式な **W3C 勧告**⁷³ となった。その後、修正が加えられた HTML4.01 が 1999 年 12 月に W3C 勧告となった。現在では、HTML5 が標準的なものとなり、2014 年 10 月に最初に勧告され、2016 年 11 月 HTML5.1、2017 年 12 月 HTML5.2 が勧告されている。2019 年には、W3C が WHATWG (Web Hypertext Application Technology Working Group) に協力することになり、2021 年には HTML5 の勧告が廃止され、WHATWG が策定した **HTML Living Standard**⁷⁴ が新たな W3C 勧告となった。

Hypertext Markup Language (HTML) はプラットフォームに依存しない hypertext を作るための単純なマークアップ言語である。SGML は仕様が大きく複雑であるため、HTML は DTD を不要にするなど、仕様を思い切り緩くした。

HTML 文書は広範囲な領域での情報を表現するための適切な生成規則を持った SGML 文書。HTML マークアップによって、ハイパーテキスト・ニュース、メール、文書、ハイパーメディア、そして、オプションメニュー、データベース問い合わせの結果、画像を含む単純な構造化文書、さらに、情報本体の持っているハイパーテキスト構造を表現できる。

- XML (Extensible Markup Language)

1998 年、W3C は XML (XML1.0) を発表した。HTML では決められたタグしか使えず、商取引などに必要なデータの表現が十分にできない欠点があった。そのため、SGML の拡張性と HTML の軽さを取り入れて発展させたのが XML である。

Extensible Markup Language (XML) は、SGML の部分集合で次のような特徴を持つ。

- XML のタグは、データの意味を表し、構造化できる。
- タグは自由に決めることができる。
- XML はレイアウト・デザインの機能を分離している。例えば、スタイルシートを記述する言語で書かれたファイルを用意して XSL 変換^aを行い利用する。

^aURL https://ja.wikipedia.org/wiki/XSL_Transformations を参照。XSL Transformations (XML Stylesheet への変換)。

2000 年、XML を用いた、HTML の再定式版 XHTML1.0 を発表した。

簡単な XML 文書の例

```
<?xml version="1.0" encoding="Shift_JIS"?>
<?xml-stylesheet type="text/xsl" href="testxsl.xsl"?>
<情報社会論>
  <授業>
    <日付>7月12日(火)</日付>
    <資料>あり</資料>
    <内容>www</内容>
    <宿題>なし</宿題>
    <小テスト>あり(前半)</小テスト>
  </授業>
</情報社会論>
```

⁷³ W3C が公開している技術文書 (Technical Report) の内、その技術に関するテストが全て終わって、「国際的に十分な信頼性があると認められた技術文書が完成した」最終の第 5 段階「**勧告 (Recommendation)**」のものを指す。他に、(1)「初期草案 (First Public Working Draft)」, (2)「草案 (Working Draft)」(3)「勧告候補 (Candidate Recommendation)」, (4)「勧告案 (Proposed Recommendation)」のプロセスがある。また、勧告後は、「その後の改訂版 (Edited Recommendation)」も追加されることがある。

⁷⁴ URL <https://html.spec.whatwg.org/>

9.4 ホームページ閲覧の仕組み

用語

- **WWW(World Wide Web)** : HTML で記述された, インターネットでの情報空間。
- **HTTP(HyperText Transfer Protocol)** : HTML を交信するためのプロトコル。⁷⁵
- **URL(Uniform Resource Locator)** : インターネット上の資源の場所を示す標準方法。
例 新潟大学理学部ホームページの URL `http://www.sc.niigata-u.ac.jp/index.html`
例 新潟大学統合型学務情報システムの URL `https://ness.niigata-u.ac.jp/`
例 田中研究室の授業関連のページの URL `http://m.sc.niigata-u.ac.jp/~prtana/`
例 ファイル転送サービスの URL `ftp://ftp.abc.com/pub/abc/` (架空のもの)

ホームページ閲覧の仕組み

- ホームページ閲覧に必要なもの：
 - **Web サーバー (WWW サーバ)** : ホームページを置くコンピュータとそのコンピュータ上で常時稼働している, サーバソフト (Apache, IIS 等)。
 - **Web ブラウザ** : ホームページを見るコンピュータにあるホームページ閲覧ソフト (Internet Explore (IE), Microsoft Edge, Google Chrome, Firefox (Fx), Safari, Opera 等)
 - **Web コンテンツ** : サーバに置く html 文書 (WWW コンテンツともいう)
- 具体例による, ブラウザ (PC) とサーバソフト (Server) のやり取り (第4章や第5章で説明した TCP のコネクション通信により実現される。)

(PC) ブラウザの URL で表示文書を指定 :

`http://m.sc.niigata-u.ac.jp/~prtana/its/its.html`

(PC) DNS サーバー と交信して, 第8章にあるように DNS の仕組みにより, 接続したい Web サーバのドメイン名に対応する IP アドレスを取得する。その IP アドレスを使って, Web サーバとコネクション通信して, パスに従って (ユーザ prtana のホームディレクトリ its にある) Web コンテンツ (its.html) の送信を Web サーバへ依頼する。

(Server) 要求文書 its.html を PC に送信。

(PC) 受け取った its.html を PC のブラウザが解析。

(PC) 文書に含まれるリンクや画像ファイル等があった場合には, Web サーバへそれらを再度送信依頼。

(Server) 要求文書を順番に PC に送信。

(PC) ブラウザが受け取った文書 (リンク先の文書や画像ファイル等) を順にブラウザのウィンドウ (出力装置) に表示。

9.5 HTML 文書の例

授業中に, HTML のスケルトンを取り出す方法を教えます。

⁷⁵参考 : HTTPS(Hyper Text Transfer Protocol over SSL) 暗号化通信を行うもの。