

今回のテーマ:

WWW (World Wide Web)

というか、Internet の話

コンピュータネットワーク

コンピュータ同士を繋いで

情報のやりとり (通信・共有) をする、

という構想は、**1960**年頃からあった

→ 実用的になったのは、**1960**年代末

(ARPANET)

コンピュータネットワークの歴史

ARPANET :

ARPA (米国防総省高等研究計画局)

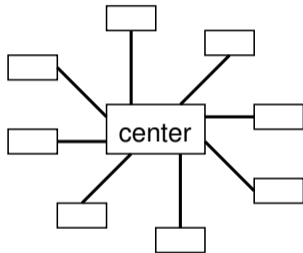
主導のネットワーク

特徴：分散型ネットワーク (特定の中心を持たない)

→ 災害・攻撃に強い

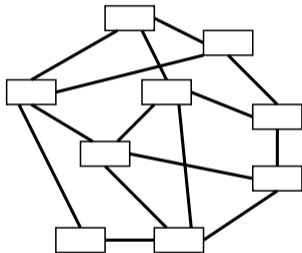
ネットワークトポロジー (繋がり方)

集中型



中心が壊れると
機能しない

分散型



どこか壊れても
機能する

コンピュータネットワークの歴史

ARPANET

- 軍用・学術研究用に分離 (1980年代前半)
- 学術研究用ネットワークが普及
(多くの大学・研究所が参加)

コンピュータネットワークの歴史

日本では、

- **JUNET (1980年代中頃から)**
- **WIDE Project (1980年代後半から)**

大学・研究所が中心の
学術研究目的ネットワーク・プロジェクト

→ **1990年代前半にパソコン通信と相互接続
(異文化同士の接触であった)**

コンピュータネットワークの歴史

当初は学術研究・非営利目的に限定

→ **1990** 年代中頃に商用利用容認

→ 以後、爆発的に広まる

internet は元来普通名詞であるが、
(= **inter-network network**)

現在では、世界に広まったこのネットワークを
“**the Internet**”

と呼んでいる

インターネット利用の情報の流通

- 電子メール
- ファイル転送 (ftp)
- ネットニュース
- **World Wide Web (WWW)**

など

インターネット利用の情報の流通

- 電子メール
- ファイル転送 (ftp)
- ネットニュース
- **World Wide Web (WWW)**

など

World Wide Web (WWW)

1980年代末～90年代初頭に開発

元々は CERN (欧州原子核研究機構) に於いて

各国の大勢の研究者同士で

研究情報を共有するための仕組みとして開発

HTTP (HyperText Transfer Protocol)

によってデータの送受信を行なう

protocol ... 取り決め

World Wide Web (WWW)

1980年代末～90年代初頭に開発

元々は CERN (欧州原子核研究機構) に於いて

各国の大勢の研究者同士で

研究情報を共有するための仕組みとして開発

HTTP (HyperText Transfer Protocol)

によってデータの送受信を行なう

protocol … 取り決め

HyperText

データ (テキスト) 同士が**リンク**で繋がっている
… ハイパーリンク (hyperlink)

HTML (HyperText Markup Language)
などで記述される

ハイパーリンクを手軽に辿れるソフトウェア
(ウェブブラウザ (web browser))
の開発が、WWWの爆発的な普及に繋がった

ウェブブラウザ (web browser)

- HTML で記述された文書を適切に表示
- ハイパーリンクを辿れる
- 必ずしも graphical ではない
(テキストベースブラウザというのものもある)

ウェブブラウザ (web browser)

実際に手元に届いているデータは?

- [表示] → [ページのソース]
- [右クリック] → [ページのソースを表示]

HTML で書かれた内容を
ウェブブラウザが適切に解釈して
表示している

ブラウザによって表示が異なることがあり得る

ウェブブラウザ (web browser)

実際に手元に届いているデータは?

- [表示] → [ページのソース]
- [右クリック] → [ページのソースを表示]

HTML で書かれた内容を
ウェブブラウザが適切に解釈して
表示している

ブラウザによって表示が異なることがあり得る

本授業でも後で (多分年末の 2 回くらいで)

HTML を用いて

web 文書 (ウェブページ) を作ってみます

- **HTML 文書の作成**

- **授業実習用サーバへの掲載**

URL(Uniform Resource Locator)

Internet 上の資源 (の在処) を特定するための文字列

ハイパーリンクでは URL を指定

参考: URI (Uniform Resource Identifier)

URL の例

`http://ccweb.cc.sophia.ac.jp/index.html`

- スキーム名 (`http:`)
- ホスト名
 - ★ ホスト名 (`ccweb`)
 - ★ ドメイン名 (`cc.sophia.ac.jp`)
- パス名
 - ★ ディレクトリ名
 - ★ ファイル名 (`index.html`)

URL の例

http://ccweb.cc.sophia.ac.jp/
News/201010/print_jp.pdf

- スキーム名 (http:)
- ホスト名
 - ★ ホスト名 (ccweb)
 - ★ ドメイン名 (cc.sophia.ac.jp)
- パス名
 - ★ ディレクトリ名 (News/201010/)
 - ★ ファイル名 (print_jp.pdf)

IPaddress と DNS(Domain Name Service)

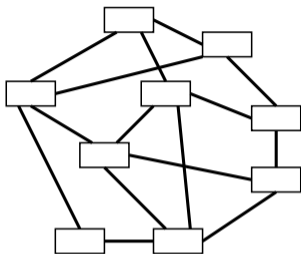
実際のネットワーク通信では、

各ネットワークインターフェイスに対し、

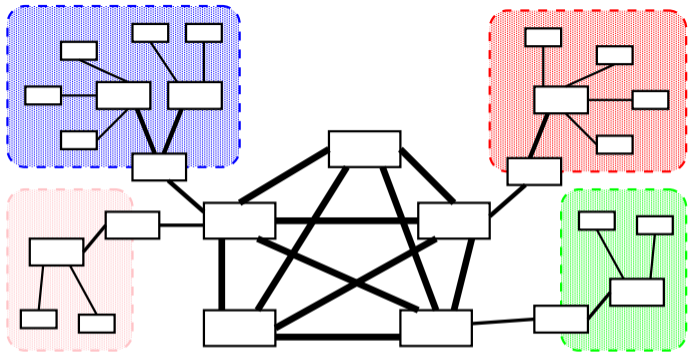
全世界 (Internet 全体) で一意的な番号を付けて、

その番号 (**IPaddress**) を用いて通信する

Internet の特徴は分散型ネットワークであった



- どこか壊れても機能する
- 誰も全貌を把握していない / 出来ない
- 全貌を把握していなくても機能する
→ 刻々と変化するネットワーク構成に対応



ネットワーク上で遠く離れた所に
アクセスするにはどうしたら良いか？

… 経路制御 (**routing**)

IPAddress と DNS(Domain Name Service)

Internet 上のネットワークインタフェースは

IPAddress で一意的に特定されるが、

いつでも番号で呼ぶのは扱い難いので、

通常は

「ホスト名 (ホスト名 + ドメイン名)」

で呼んでいる

<http://ccweb.cc.sophia.ac.jp/index.html>

IPAddress と DNS(Domain Name Service)

「ホスト名 \longleftrightarrow IPAddress」の変換の仕組み

… **DNS (Domain Name Service)**

各ドメインの name server が

自ドメイン内の変換表を持つ
(分散管理)

例えば、上智大学は、

sophia.ac.jp \longleftrightarrow 133.12.xxx.xxx

で、 $(2^8)^2 = 256^2 = 65536$ 個の IPAddress を
管理している

WWW 上の情報検索

WWW での情報共有機能は、

情報検索の的確化・効率化によって

一挙に世界を変えるに至った、

と言っても過言ではない

WWW 上の情報検索

実習:

都道府県庁所在地の一覧を探したい

WWW 上の検索で一覧データを見付けよ

→ **Google** を使う

情報検索のコツ

大量のデータの中から必要な情報を取り出すために

- そこにどんなデータ (文書) があるかを想像せよ
- キーワード検索なら、
自分が見付けたいデータ (文書) が
どんなキーワードを含んでいそうかを
想像せよ
- それが人の書いた文書なら、
どんな文章だろうか、と想像せよ

ネットワーク上の人間と社会

ネットワークの向こうには人間がいる

→ ネットワークは人間社会である

→ 社会には守るべき規範 (慣習) がある

- ネットワーク倫理 (Nethics)
- ネットワークエチケット (Netiquette)

WWW 上の情報検索

昔は次のように良く言われた:

「自分が見付けたいデータ (文書) が
どんなキーワードを含んでいそうかを
想像せよ」

→ 思い付く限りの県名・県庁所在地名を
検索キーワードに並べて指定せよ

しかし、現在ではそこまでしなくても
良い検索結果が得られることが多い

WWW 上の情報検索

昔は次のように良く言われた:

「自分が見付けたいデータ (文書) が
どんなキーワードを含んでいそうかを
想像せよ」

→ 思い付く限りの県名・県庁所在地名を
検索キーワードに並べて指定せよ

しかし、現在ではそこまでしなくても
良い検索結果が得られることが多い

検索結果の順位付け方法

以前は主に、

そのページに含まれるキーワードの多寡・質
で順位付け

- 検索で上位に来ることだけを狙った
ページが増加 (宣伝業者など)
- 検索結果の品質の低下
- WWW のデータベースとしての品質の低下
- **Google** の登場

検索結果の順位付け方法

以前は主に、

そのページに含まれるキーワードの多寡・質
で順位付け

- 検索で上位に来ることだけを狙った
ページが増加 (宣伝業者など)
- 検索結果の品質の低下
- **WWW** のデータベースとしての品質の低下
- **Google** の登場

検索結果の順位付け方法

以前は主に、

そのページに含まれるキーワードの多寡・質
で順位付け

- 検索で上位に来ることだけを狙った
ページが増加 (宣伝業者など)
- 検索結果の品質の低下
- **WWW** のデータベースとしての品質の低下
- **Google** の登場

Google の PagerankTM

WWW 上の全てのページの
重要度 (**Pagerank**) を判定

基本原理:

「多くの良質なページから
リンクされているページは、
やはり良質なページである」

→ キーワードを含むページを重要度の順に提示

Google の PagerankTM

計算法の原理:

リンクをランダムに辿りながら大勢で動き回る

→ 人が集まるページが重要なページ

- 確率過程 (マルコフ連鎖)
→ 線型代数 (行列計算) の固有値問題
- 効率的な数値計算法の利用

これにより高速かつ有効な検索結果が得られた

→ WWW そのものが巨大なデータベースに

Google の PagerankTM

計算法の原理:

リンクをランダムに辿りながら大勢で動き回る

→ 人が集まるページが重要なページ

- 確率過程 (マルコフ連鎖)
→ 線型代数 (行列計算) の固有値問題
- 効率的な数値計算法の利用

これにより高速かつ有効な検索結果が得られた

→ WWW そのものが巨大なデータベースに

Google の PagerankTM

計算法の原理:

リンクをランダムに辿りながら大勢で動き回る
→ 人が集まるページが重要なページ

- 確率過程 (マルコフ連鎖)
→ 線型代数 (行列計算) の固有値問題
- 効率的な数値計算法の利用

これにより高速かつ有効な検索結果が得られた

→ WWW そのものが巨大なデータベースに

メディアとしての WWW の特性

- 誰でも手軽に参加できる
- 強力な検索機能

→ 少数者・新規参入者も

多数者・既得権益者と同等

→ (相対的には) 少数者に力を与えるメディア

メディアとしての WWW の特性

- 誰でも手軽に参加できる
- 強力な検索機能

→ 少数者・新規参入者も

多数者・既得権益者と同等

→ (相対的には) 少数者に力を与えるメディア

メディアとしての WWW の特性

- 誰でも手軽に参加できる
- 強力な検索機能

→ 少数者・新規参入者も

多数者・既得権益者と同等

→ (相対的には) 少数者に力を与えるメディア

期末レポート課題の例

次のような事柄について調べよ:

- **ARPANET** 以来のインターネットの歴史
- インターネット上での情報の流通
- **Google** の **Pagerank** について 等々

今後も授業内容に関連した課題例を追加予定

詳細は追ってお知らせします

今日の提出課題 (授業時)

都道府県庁所在地の一覧が掲載されている
web page を WWW 上の検索で探し、

そのうちの 1 つの URL をメールの本文に記して、

課題提出用のメールアドレスに送る

次回 (10/25) 予告:

図書館利用案内

図書館データベース (OPAC) 検索実習

その次の週 (11/1) は創立記念日のため全学休講