

インターネット数理科学第4回

～ネットワークのあちら側を支える数理科学その1～

2006年10月26日

株式会社インターネット総合研究所代表取締役所長
東京大学大学院数理科学研究科客員教授

藤原 洋

目 次

1. ネットワークのあちら側を支える数理科学とは？
2. 「ネットワークのあちら側」の要素技術
～ Webサーバサイド・コンピューティング～
3. 動的Webページを創るCGIとSSI
4. WebサーバプログラミングへのJavaのインパクト

1. ネットワークのあちら側を支える数理科学とは？

③(ネットワークの)あちら側

⇒「グラフ理論」「金融工学理論」に基づくデータベース、検索エンジン最適化、検索連動データベース、ネット金融サービス

①ネットワークそのもの

⇒「グラフ理論」による動的ルーティング、帯域制御、放送型ルーティング
「デジタル信号処理理論」に基づく変復調技術

②(ネットワークの)こちら側

⇒「デジタル信号処理理論」に基づくコンテンツ符号化技術

③(ネットワークの)あちら側

Web1.0(ポータル)⇒Web1.5(SNS)⇒Web2.0(ロングテール)

①ネットワークそのもの

ダイヤルアップ/2Gモバイル⇒ブロードバンド/3Gモバイル⇒IP放送/NGN/WiMAX

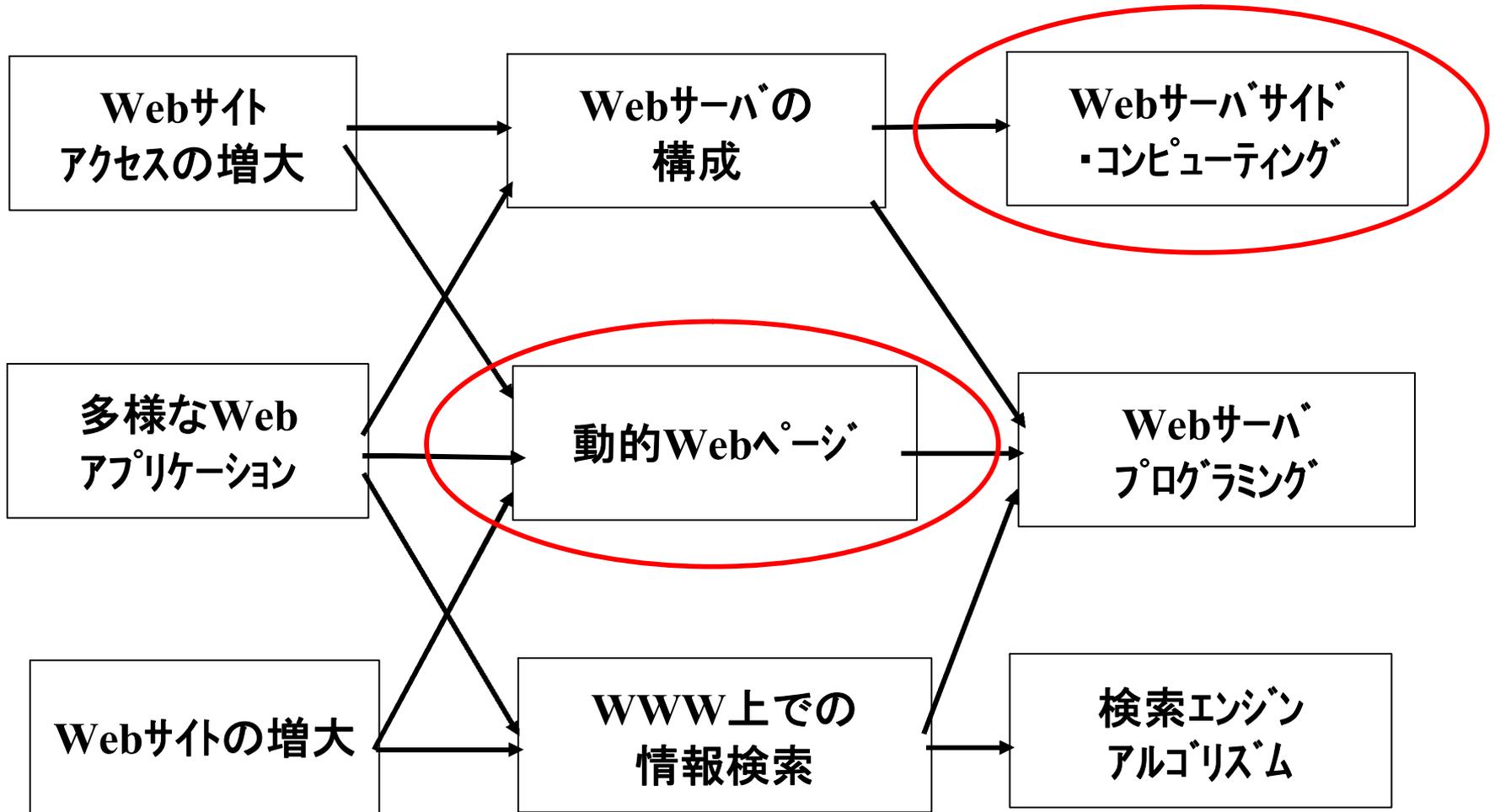
②(ネットワークの)こちら側

文字情報(Eメール)⇒HTML(ブラウザ)⇒動画(デジタル符号変換)

課題

着眼点

具体策



Web上で、キーワードから所望のURLを探し出す様々な「検索エンジン」が出現、勝ち組みが、スタンフォード大学院生だったジェリー・ヤンとデビット・ファイロが95年に設立したYahoo!社でした。このYahoo!に代表される検索エンジンを用いたディレクトリ(検索)・サービスというビジネス形態が生まれました。同社は、ポータル(玄関口)サービスへと発展しましたが、検索エンジンのアルゴリズムを徹底的に追及したのがスタンフォード大学の大学院生だったラリー・ページとセルゲイ・ブリンで、リンク数の統計から優先順位をつけて自動的に探し出すクローラー(ロボット)型検索エンジンの仕組みを作りGoogle社を98年に設立しました。

私自身は、96年にIRIを設立し、同社に注目はしていましたが、所詮Googleは、優れた検索エンジンでしかなく、Yahoo!と比較して、ビジネスとしては、難しいだろうとみていました。一方、別の技術をもったOverture社(設立時GoTo.com、現在はYahoo!の完全子会社)が、97年9月にIdealab社のビル・グロスによって設立、広告主が関連性のある特定キーワードに入札し、それが検索結果に表示を可能にするPay-For-Performance検索サービス(スポンサードサーチ)を開始しました。Googleは、2002年から高額の特許料を払い、同技術を取り入れ、「検索連動型広告」を中心にビジネスを組み立て直し一気に高収益企業となりました。Yahoo!とGoogleは、昨年地上波4大ネットワークの広告収入を上回るまでになりました。今後さらにポータルと検索エンジンが更に進化し、巨大なメディアになると思われます。

「広告産業は変わるか？」マスメディアは、これまで、マス4メディアとも呼ばれるように、テレビ、ラジオ、新聞、雑誌を指してきか、例えば、ヤフージャパンの閲覧ページ数は、2002年10月に100億ページビュー、2005年10月には300億ページビューを突破したように、ポータル・サイトは、マスメディアの地位を確立した。また、アマゾン・ドットコムに代表されるロングテール(売れ筋商品だけでなく、少量販売数品目を扱うことで開拓される新市場)市場の開拓によって、インターネットは、ロングテールメディアとしての新たなメディア属性が認識された。インターネットの爆発的普及によって、消費者のライフスタイルや価値観に大きな変化が起こり、既存メディアによる広告手法には限界が見えてきた。インターネットは、既存メディアではできなかった消費者との直接接触を可能にした、初めてのメディアだが、そしてこのメディア属性こそが、インターネットが広告産業を根本的に変えてしまう可能性を秘めている。

インターネット広告は、1994年10月27日のホットワイアード創刊に14社分のバナー広告が掲載されたのが最初だが、アメリカでは、1996年4月にマイクロソフトやインフォシークが立ち上げたIAC(Inretnet Advertising Council)を前身とするIAB(Interactive Advertising Bureau)が設立され、インターネット広告の概念や広告取引の標準化が進行。

「ネットワークのあちら側」が変化した結果？

日本では1999年4月、日本広告主協会がインターネット広告の課題や効果などを研究する専門組織Web広告研究会が発足、広告会社や媒体社はインターネット広告推進協議会をスタート。ジュピターリサーチの予測、アメリカのインターネット広告市場は、2004年に93億ドル、2005年に119億ドル、2010年には189億ドルにまで成長。2005年の総広告市場の伸びが対前年比約3%であるのに対してインターネット広告市場の伸びは13.9%に達する。

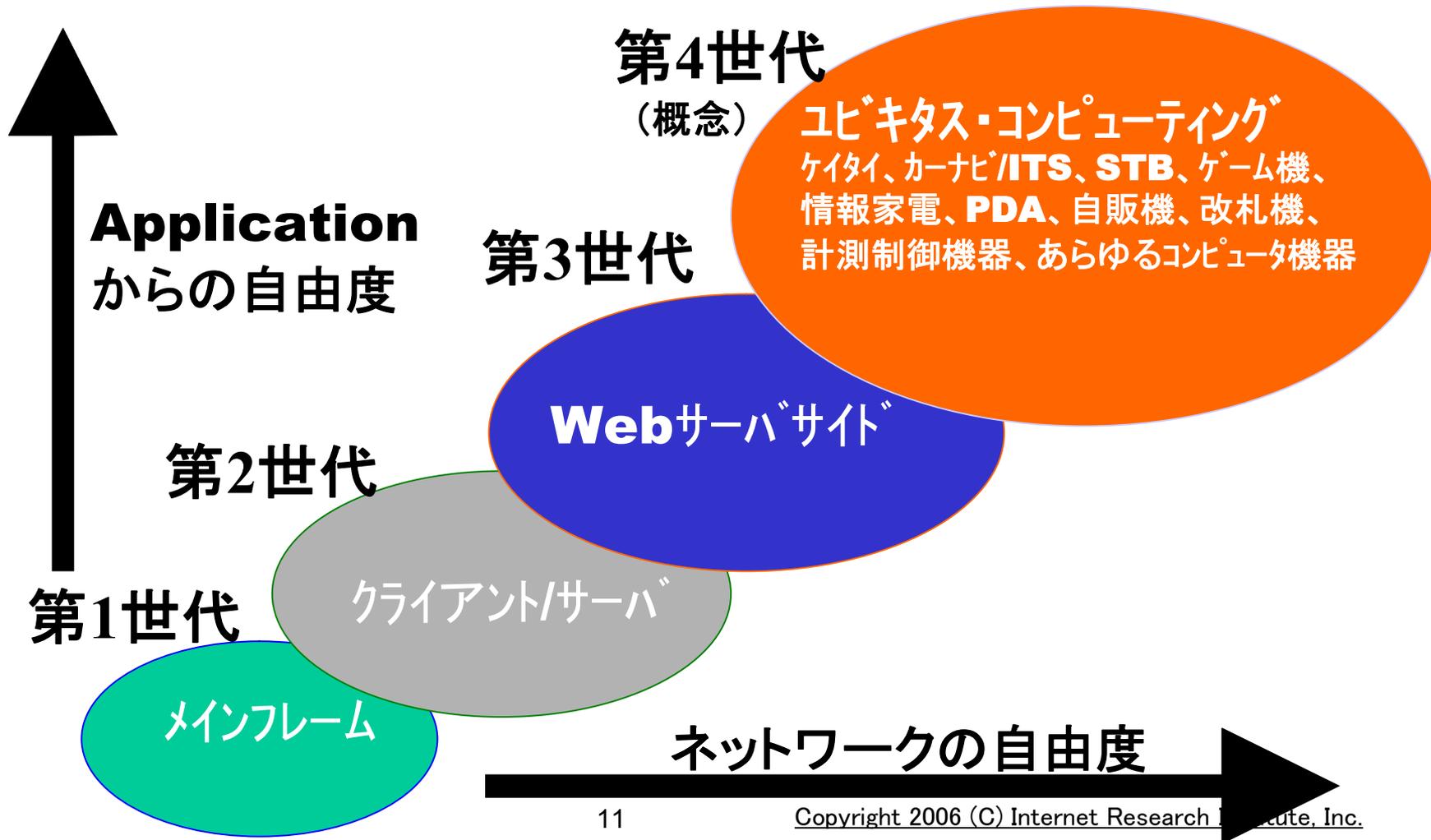
一方、日本の同市場は、電通によると、2808億円(そのうちモバイル広告288億円、検索連動型広告590億円)と推計されており、総広告費5兆9,625億円のうちの4.7%に相当。主要メディア広告費の前年比は、テレビ99.9%、新聞98.3%、雑誌99.4%、ラジオ99.1%、インターネット154.8%。このアメリカ以上の日本での同市場の高成長要因は、ブロードバンドの普及とモバイルや検索連動型広告など広告の多様化にあり、インターネットの広告メディアとしての価値を高く評価する広告主が増加。

2006年の総広告費は前年比102.1%の見通だが、インターネット広告は50%以上の高成長を維持することが確実視されている。2004年にラジオ広告費を抜き、2007年に雑誌広告費を抜くと予測されるが、このままいくと、2006年に早まる可能性が高まってきた。国家のGDPの約1%を占めると言われる広告市場の中でインターネット広告はまだ5%程度になったばかりだが、国民生活の構造変化と共にその比率は、さらに高まり続けていくことだろう。

2. 「ネットワークのあちら側」の要素技術 ～ Webサーバサイド・コンピューティング～

“The Network is The Computer.” Bill Joy.

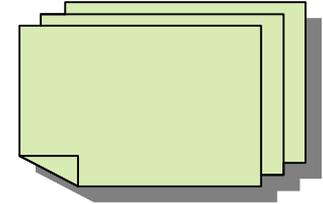
“C&C : Computer & Communication” 小林宏治



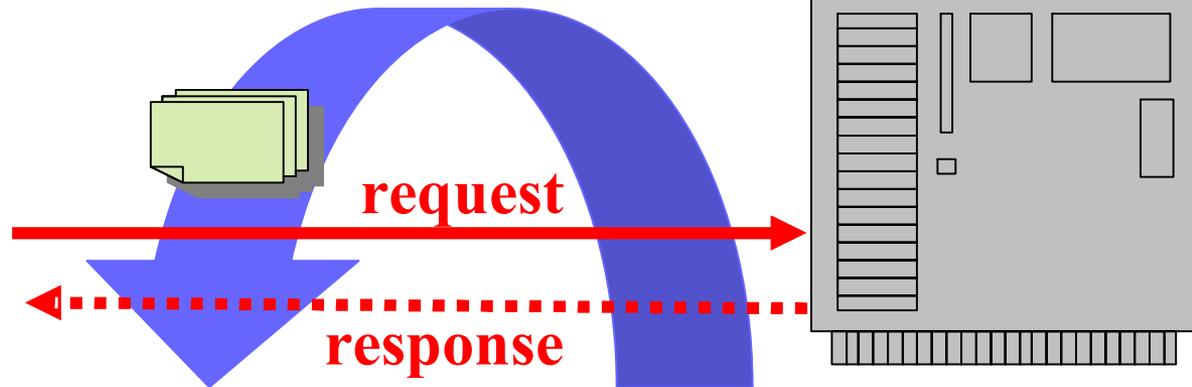
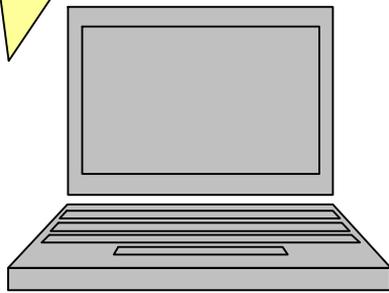
情報検索
予約サービス
電子商取引
Web掲示板

これが基本！

①記述言語:HTML



④送受信プロトコル:HTTP



③情報受信:Webブラウザ
(Internet Explorer, Navigator,etc.)

②情報発信:Webサーバ
(IIS, Apache,etc.)

WebブラウザにURLを入力するか、
Webページのリンクをたどればよい！

最近のWebサーバは、ユーザー入力
で起動するWebアプリケーションを実装する！

CGI,SSI,Java

- Apache
 - OS環境 : Linux、Windows、Mac-OS
 - 開発言語: CGI、PHP(Hypertext Preprocessor)
 - フリーウェア、Tomcatと連携可
- IIS (Windowsに付属)
 - OS環境 : Windows
 - 開発言語: CGI、ASP、ASP.NET
 - フリーウェア、Tomcatと連携可
- Tomcat
 - OS環境 : Linux、Windows、Mac-OS
 - 開発言語: Javaサーブレット、JSP (Java Server Pages)
 - フリーウェア、Tomcatと連携可

- Webサーバサイド
 - データベース処理等の機密性の高い処理
- クライアントサイド
 - マウスクリックに即座に反応すべき処理
 - アニメーション、ポップアップメニュー等
- Webサーバサイドコンピューティングへ収束
 - クライアント側 : 汎用Webブラウザだけを実装
 - Webサーバ側: 専用ハードウェアやソフトウェアも実装
 - 前提条件 : HTTP以外の通信プロトコルは使用せず

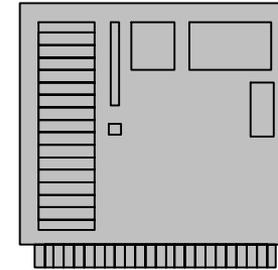


【クライアント】



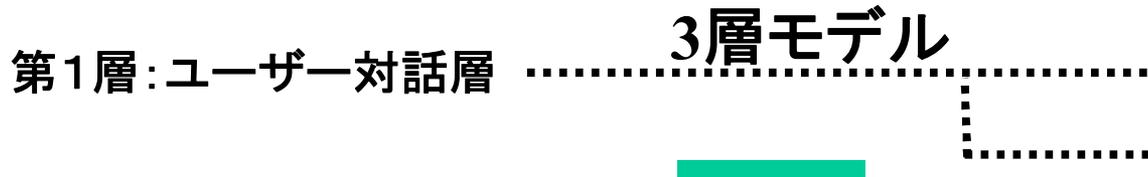
2層モデル

第1層: クライアント層



【サーバ】

第2層: サーバ層



3層モデル

第1層: ユーザー対話層

第2層: ビジネスロジック層

第3層: データベース層

進化

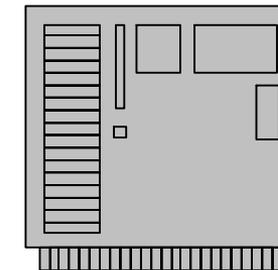


【Webブラウザ】



3層モデル

第1層: 汎用ブラウザ



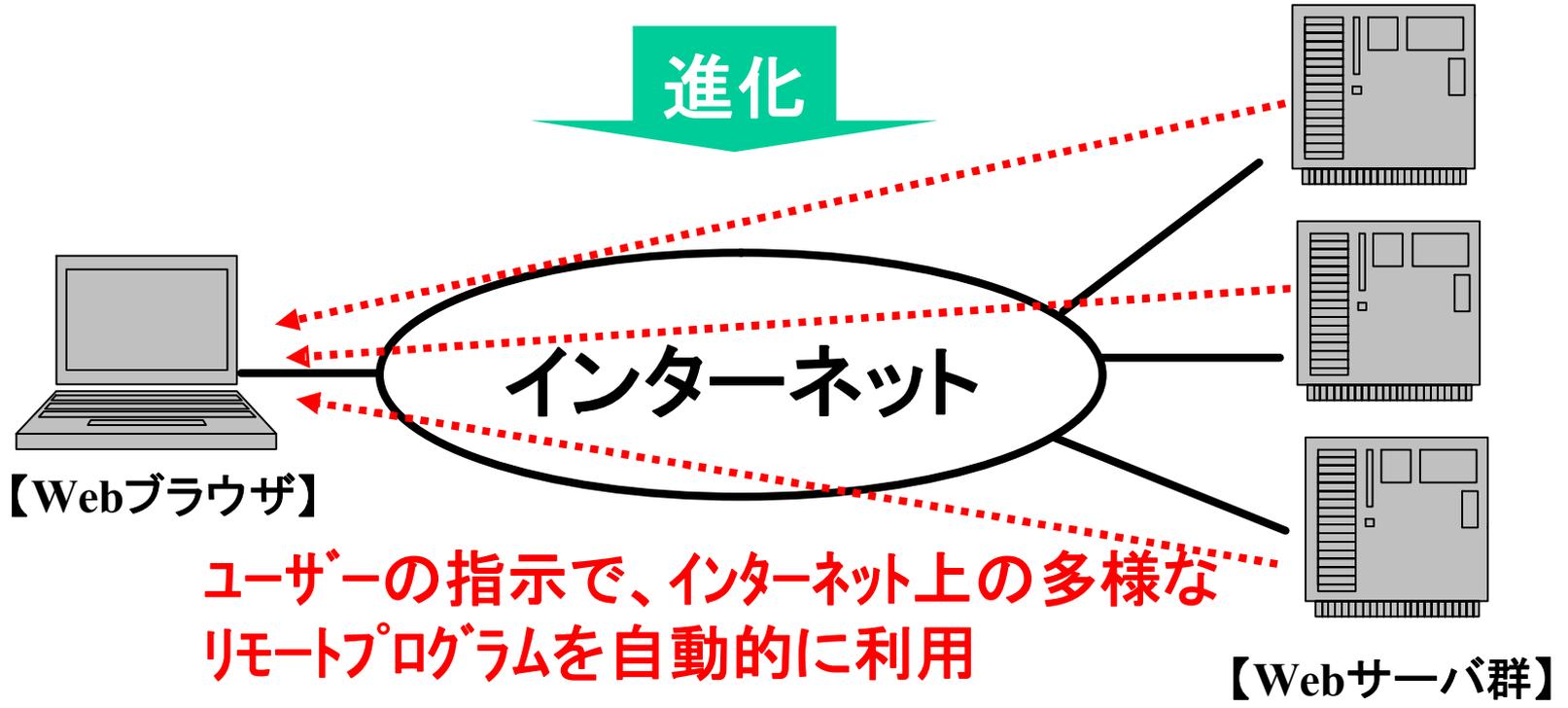
【Webサーバ】

第2層: ビジネスロジック層

第3層: データベース層



進化



- WWW関連の技術を使い、ソフトウェアの機能をネットワークを通じて利用できるようにしたもの。
- 企業間の大規模なものから、単一の機能を持つものまで、様々な種類のものがある。
- 機能記述や呼び出し手順などの標準化が進行中。
- 複数のWebサービス同士をつなぎ合わせてアプリケーションを構築する方式が次世代のソフトウェア環境の主流になるといわれている。
- 従来のOSやミドルウェア(OS上で動作し、アプリケーションソフトに対してOSよりも高度で具体的な機能を提供するソフトウェア)は、Webサービスを開発・実行する環境としての役割を担うようになり、サービス(およびそれらを組み合わせたアプリケーション)を利用するエンドユーザは、現在のWebブラウザを拡張したクライアントソフトを通して、すべてのソフトウェアを利用するようになるといわれている。
- 具体的なWebサービスとしては、様々なソフトウェアやサービス、開発環境、仕様などが各社から提案されており、Microsoft社のMicrosoft .NETや、IBM社などが中心となって進めているUDDIプロジェクトなどが知られている。
- 現在のところ、XMLをベースにした標準や標準案が多く、中でも、ソフトウェアの機能の遠隔呼び出しの手順を定めるプロトコルは、XMLベースのSOAP(Simple Object Access Protocol、XMLとHTTPなどを基本に、別のコンピュータにあるデータやサービスを呼び出すためのプロトコル。Microsoft社やUserLand Software社、Developmentor社が中心となって開発。)がデファクトスタンダードとなっている。
- Webサービスは広い概念で、まだ発展途上であり、各分野に応じた数々の標準仕様が登場していくものと思われる。

- 性能評価指標: **Webサーバ群の性能を定量的に見積もる**
 - ① スループット
 - ② アクセス数
 - ③ 接続数

- 負荷分散方式: **サービス実現のWebサーバ群の構成方式**
 - ① ラウンドロビン
 - ② ロードバランサー

① スループット:

- ・Webサーバ群として単位時間あたりの処理可能トラフィック量
- ・bpsで表現
- ・ネットワーク構成のボトルネックがポイント

② アクセス数:

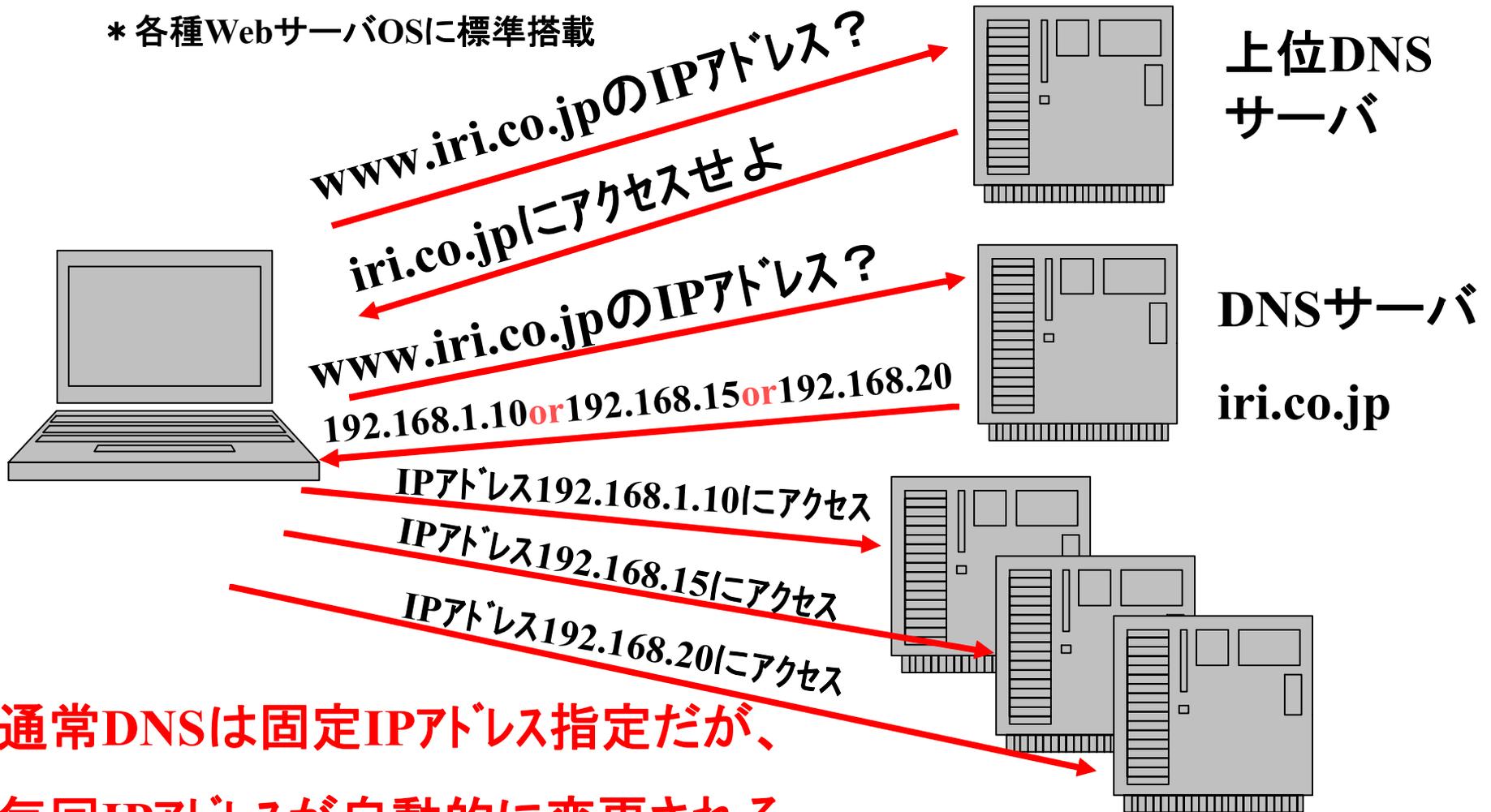
- ・HTTPトラフィックの特徴＝トラフィック量は少ないが接続が多発
- ・HTTP1.0の問題点＝HTMLページ内オブジェクト別にTCPコネクション
(オブジェクト数＋HTMLデータ転送に1本の本数)
- ・HTTP1.1の改良点＝HTMLページ全体で1つのコネクション

③ コネクション数: HTTP1.1改良によって発生した新たな性能指標

- ・Webサーバ/Webブラウザ間TCPコネクションは2回目以降も保持
(Keep-alive機能)
- ・通常は同時に張れるTCPコネクション数をサーバOSで制限
- ・WebサーバではWebブラウザからの要求制限として最大スレッド数を規定
(TomcatではMax50)

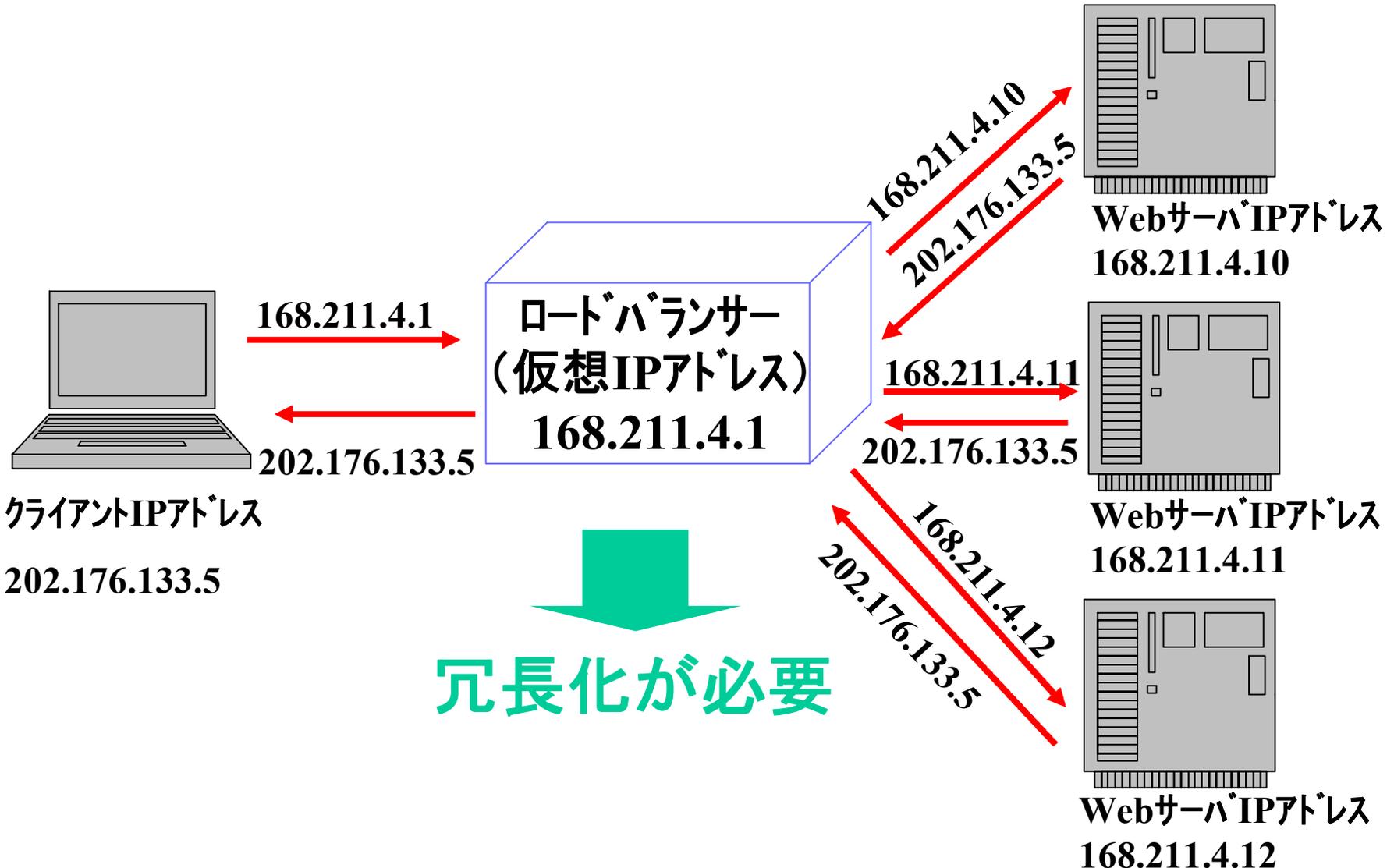
「ラウンドロビン」方式

- * 各サーバに均等にトラフィックを振り分ける方法で、サーバの処理能力に違いがない場合
- * 各種WebサーバOSに標準搭載

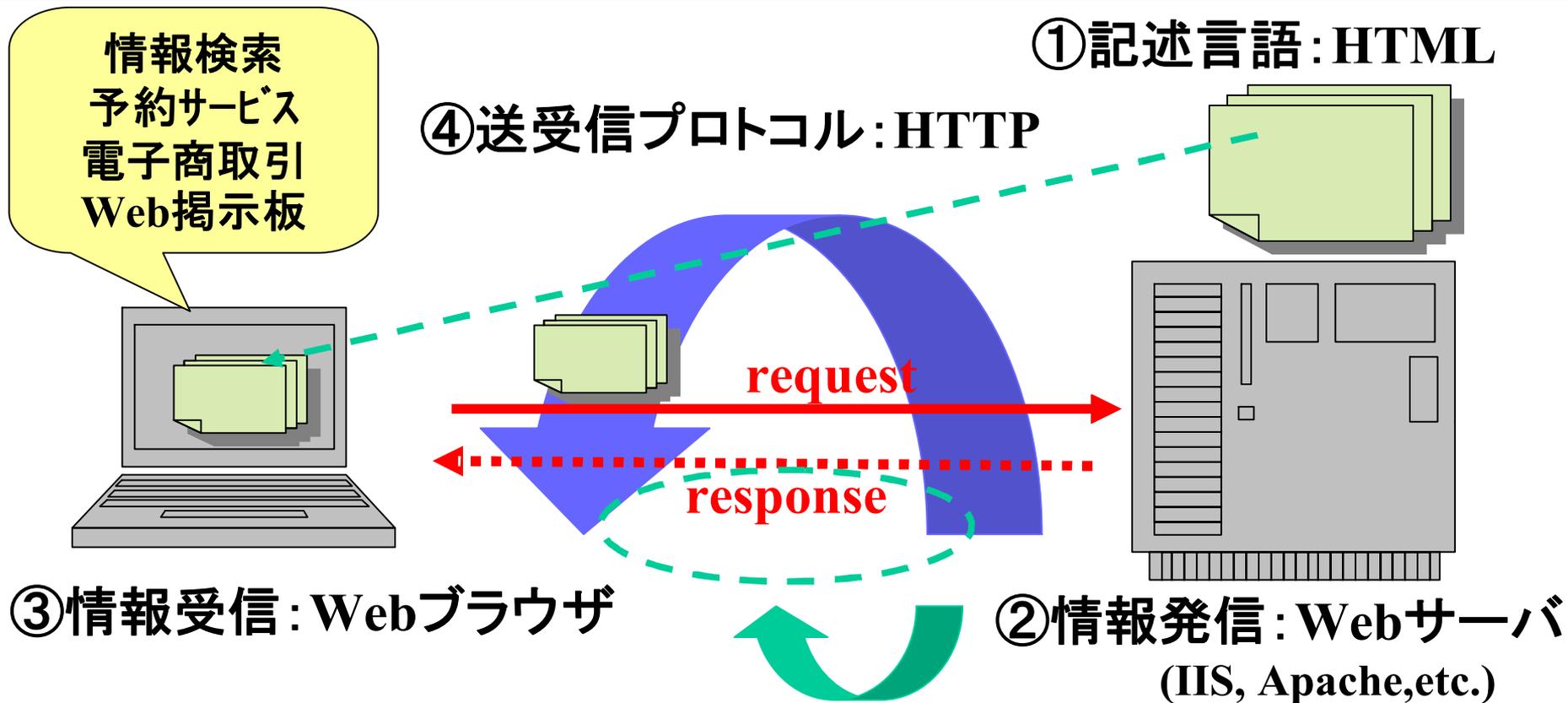


通常DNSは固定IPアドレス指定だが、
毎回IPアドレスが自動的に変更される

「ロードバランサー」方式



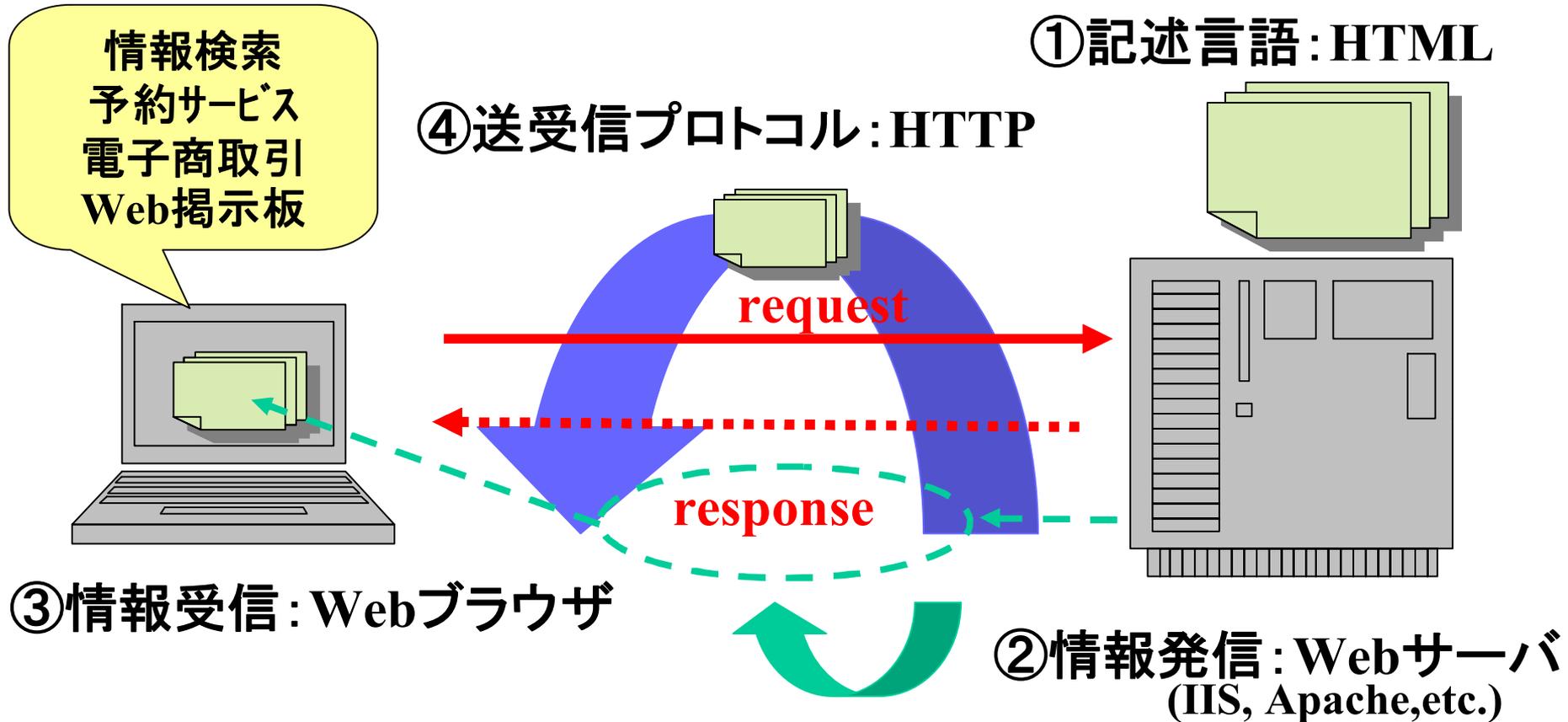
3. 動的Webページを創るCGIとSSI



ブラウザの要求でWebサーバ内の指定されたプログラム
処理結果をWebサーバを通じてブラウザへ返す仕組み

CGI: Common Gateway Interface

Webの4大要素とSSIの役割



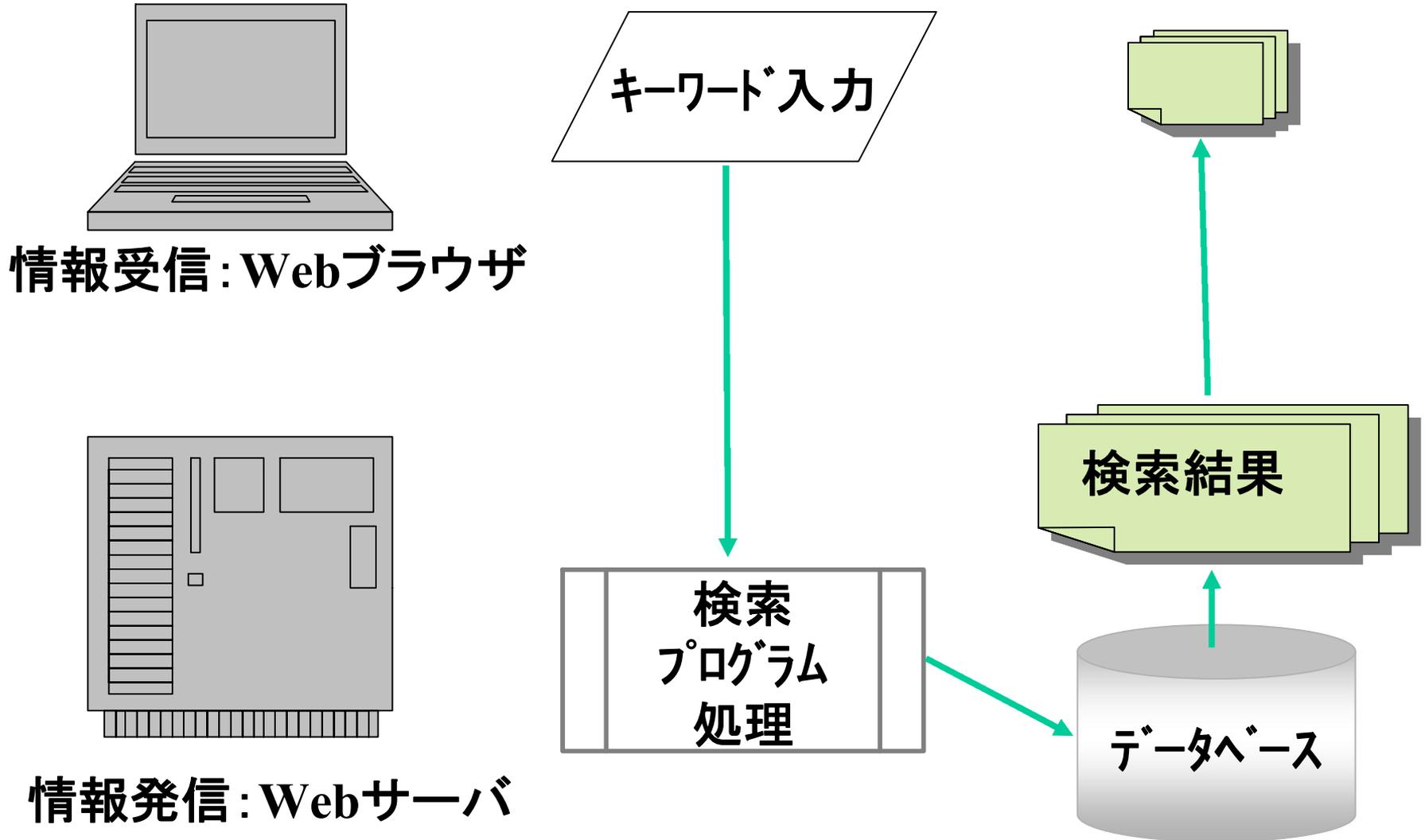
(Internet Explorer, Navigator,etc.)

WebブラウザにURLを入力するか、
Webページのリンクをたどればよい!

ブラウザの要求でWebサーバ内の指定されたプログラム
処理結果をブラウザ側HTML文書内へ返す仕組み

SSI: Server Side Include

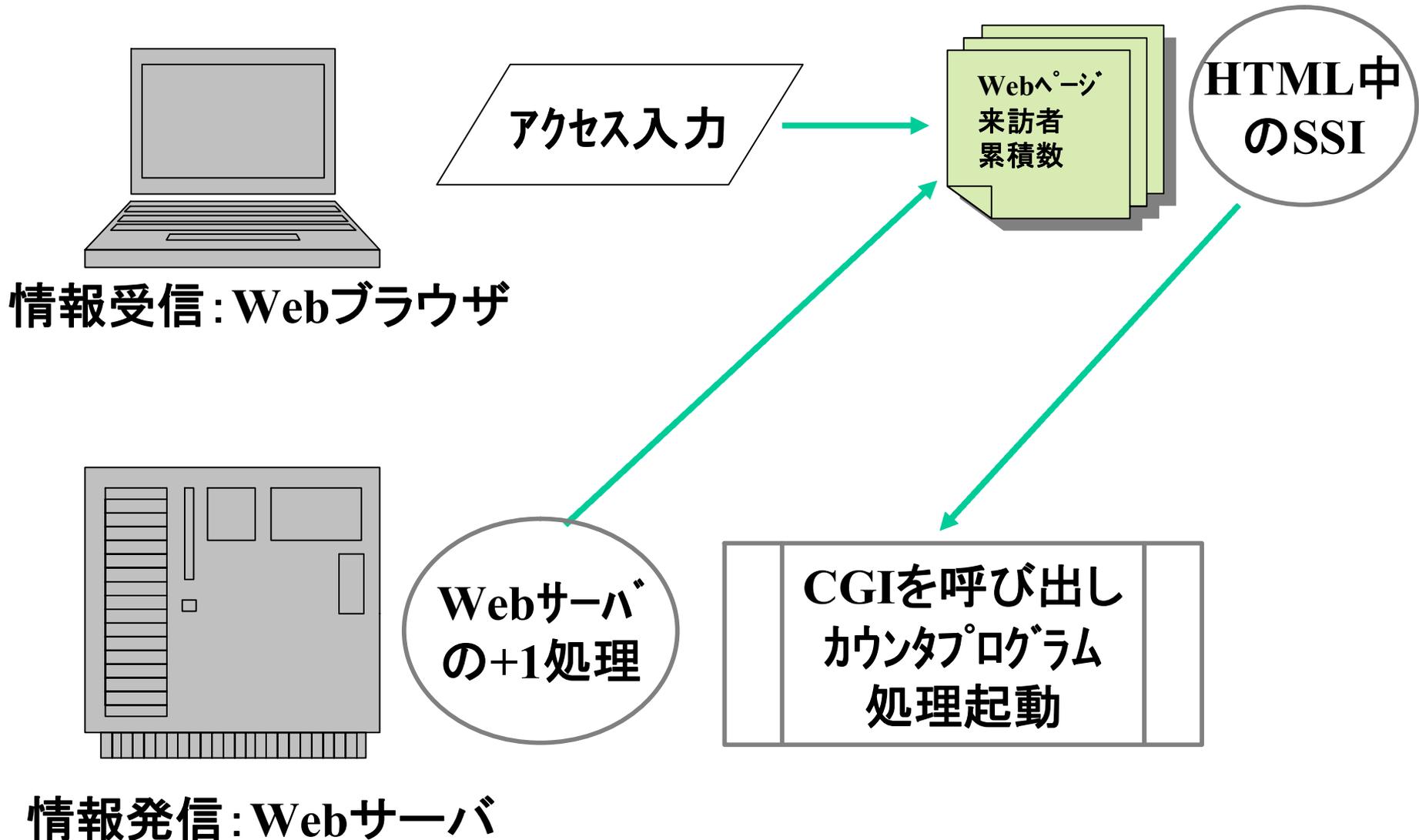
- プログラミング言語: 自由 Perl, C, Ruby, Visual BASIC
 - Webサーバサイドプログラミングで最多利用
 - ユーザーから直接入力が可能
-
- 検索エンジン、Web掲示板などのサービスに利用
(例) ユーザー入力された検索のキーワードは、
HTTP経由でWebサーバ上の検索エンジンプログラムがWebサーバ内のデータベースを検索し、
HTTP経由で検索結果をブラウザへ渡す。
- ⇒ ユーザーは、自分側のコンピュータにデータベースを持つことは不要



- CGIは、プログラミング言語ではなく、Webサーバ上のプログラムとWebブラウザを結ぶインタフェース
- Webサーバ上のプログラムは、言語に依存しない。
- CGIプログラミングに用いられる言語
 - Perl(Practical Extraction Report Language) :
インタプリタ言語
Larry Wallが開発
 - C : コンパイラ言語
Dennis M.Richie,とBrian W.Kernighanが開発
- 例: `<form action=「Webサーバ側プログラムのアドレス」
 methods=「データ形式」
 enctype=「MIMEタイプ」
 ~</ form>`

- プログラミング言語: 制約あり C, C++, Visual C++
OS環境でコンパイル結果が相違
- Webサーバサイドプログラミングで広く利用
- ユーザーからの直接入力是不可能
- 現在時刻表示、アクセスカウンタなどのサービスに利用

SSIによるアクセスカウンタの動作例



- ブラウザ側での記述例

HTMLによるアクセスカウンタ

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML4.01//EN"
```

```
<html>
```

```
<head>
```

```
  <title>アクセスカウンタ</title>
```

```
</head>
```

```
<body>
```

```
  貴殿は<!-- #exec cmd="./ counter.cgi"-->番目の来訪者です。
```

```
</body >
```

```
</ html>
```

- Webサーバ側は、Perl、C言語などでこれに対応するプログラムを行う。

- アクセス権限

Permission: Webサーバのファイルアクセスの許可の仕組み

type: dはディレクトリ型

owner: 所有者に対する許可

group: 所有者が属するグループへの許可

other: その他全員への許可

⇒各々の権限を下記のrwxの合計値として数値化する。

* r=4 : 読込み権限 w=2 : 書込み権限 x=1 : 実行権限

- セキュリティ向上策

- ・ファイルのPermissionに注意を払う

- ・ブラウザ～Webサーバ間の情報を暗号化: http://⇒https://

4. WebサーバプログラミングへのJavaのインパクト

- Sun Microsystems社が開発したプログラミング言語

単なるプログラミング言語ではない！

Java Virtual Machine(仮想マシン)に特徴！

- Write Once, Run Anywhereが設計思想
- 家電製品、PDAなど組込型用途だったが、インターネットの普及でWeb用言語として発展
- NetscapeのNavigatorでJavaアプレットが動作したことでインターネット上での言語として定着
- Sun Microsystems社からの10周年メッセージ：
誕生から今年で10年目となるJavaテクノロジーは、様々なタイプの消費者とビジネスユーザに多様な価値を提供する完全なソフトウェア・エコシステムに発展しました。Javaテクノロジーは、次の3つのニーズに対応したプラットフォームを提供しており、開発者は目的に応じた最適なテクノロジーを選択することができます。
 - ・小型デバイス & モバイル・デバイス向けのJavaテクノロジー
 - ・デスクトップ向けのJavaテクノロジー
 - ・中～大規模企業向けのJavaテクノロジー

Javaのバイトコードを実行するためのスタック型仮想マシン。この環境を移植することで、さまざまな環境でJavaプログラムを実行可能。APIやいくつかのツールとセットでJava 2 Platformとして利用。

* **バイトコード**: 機械語より抽象度の高い中間コード。プログラミングにおいて、特定のハードウェアへの依存度を下げ、またインタプリタの実装を容易にし、コンパイラの中間コードとしても用いられる

* **Application Programming Interface**)とは、アプリケーションから利用できる、OSやプログラミング言語で用意されたライブラリなどの機能の入り口。主に、ファイル制御、ウインドウ制御、画像処理、文字制御などのための関数として提供されることが多い。

* **グリーンスレッド**: OSのスレッドを使わずJavaで仮想的なスレッドを作って実行。

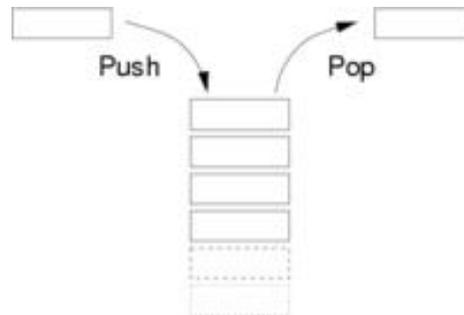
* **ネイティブスレッド**: OSのスレッドを利用。

スタックとは？

抽象データ型の一つであり、コンピュータで用いられる基本的なデータ構造の一つで、データを後入れ先出し (LIFO: Last In First Out; FILO: First In Last Out) の構造で保持する。

コンピュータシステムで、広範囲に使われており、ハードウェアとソフトウェアの機能の組合せによって実装される。

「スタックベース・コンピュータシステム」は、一時情報をまずスタックに格納するシステムを意味し、CPUのレジスタに格納するシステムを「レジスタベース・コンピュータシステム」を意味する。



ネットワークを通してWebブラウザに読み込まれ実行されるJavaのアプリケーション。Netscape Navigatorに搭載されて普及した。単にアプレットとも言う。Web ページの一部として自動的に読み込まれて動作するため、セキュリティ上の観点から一般のアプリケーションプログラムと比べさまざまな制限 (サンドボックス) が課せられる。ユーザーの許諾によりこの制限を外せる。

Webの普及初期に、インタラクティブ性を高められる技術の一つとして注目を浴び、当時のWeb普及に寄与した。

しかし、当時は、ブロードバンド環境のユーザが少なかったことと、Javaがまだ生まれて間もない技術だったためWebブラウザ上のJava VM (Virtual Machine)の動作が遅く、ベンダごとのVMの実装時に互換性がなく、あまり利用されない時期が続いた。

最近になってモバイルとブロードバンドのインターネット環境の整備でJavaアプレットの利用が急速に進み始めた！

Webサーバ上で実行されるモジュール化されたJavaプログラム。サーブレットを追加することにより、Webサーバの機能を拡張することができる。

サーブレットはJava言語で記述されているため、特定のOSやハードウェアに依存することがなく、サーブレットAPIを実装したあらゆるWebサーバで稼働させることが可能。

CGIなどの他のサーバサイドプログラムと異なり、一度呼び出されるとそのままメモリに常駐するため、高速処理が可能。

*** CGIでは、WebブラウザからCGI実行要求が来ると別プロセスとして当CGIを実行。Javaサーブレットでは、同一プロセス内に新規スレッドを作成し、Javaサーブレットを実行。**

また、データを永続的に扱うことができるため、複数のユーザ間で情報を共有することが可能。当初はSun Microsystems社のJava Web Serverの機能の一つとして提供されていたが、仕様がJava Servlet APIとして独立し、現在では様々なWebサーバ向けにサーブレットの実行環境が用意されている。

スレッドとは、CPU利用の単位。プログラムを実行するときのコンテキスト情報が最小で済むので切り替えが早くなり、ライトウェイトプロセスとも呼ばれる。スレッド (thread) は、*thread of execution* (実行の脈絡) という言葉を省略。

プログラミングの観点からみると、アプリケーションの処理の「実行の脈絡」はひとつでないことが多い。これをシングルスレッドで実現しようとするするとシグナルやタイマーを駆使することになる。

また、複数のプロセスに分割してプロセス間通信で協調動作させるという方法もある。しかし、いずれの場合もそれらの機能を使うための余分な、本来のアルゴリズムと関係ないプログラミングが必要となる。

スレッドを使用したプログラミングでは、本来のアルゴリズムに集中しやすくなり、プログラムの構造が改善される。

コンピュータで、複数の処理を同時実行する(マルチタスク)場合、処理を分割し、同時に実行する部分を指定可能な、処理の分割単位として、スレッドとプロセスがある。

新たなプロセスを動作させるためには、CPUやメモリ空間などを割り当てる必要がある。それぞれのプロセスは、割り当てられた資源内で独立して動作。しかし、この方法では、メモリ空間が独立してしまうため、独立したメモリ空間が不必要な場合、メモリの利用効率が悪い。

プログラムによっては、共有メモリを利用しながら複数の処理を行った方が効率的あるため、プロセスの機能から必要なところだけを利用したものがスレッドである。

集合包含関係： スレッド ⊂ プロセス ⊂ タスク

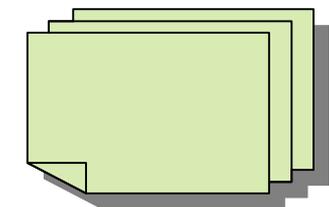
スレッドを使うことで、複数処理を同一メモリ空間上で実行でき、メモリ消費量などが軽減できる。しかし、その一方で共有ライブラリ使用時には、その共有ライブラリがスレッドセーフ(リエントラント)になっているかどうか気をつけてプログラミングしなければならない。

複数のスレッドを同時に動作させることをマルチスレッドと呼ぶ。

1つのタスクは、1つ以上のプロセスから構成され、1つのプロセスは、1つ以上のスレッドから構成。1つのCPUが、ある瞬間に同時に実行しているのは割り込み処理なども含めて1つのスレッド、1つのプロセス、1つのタスクである。

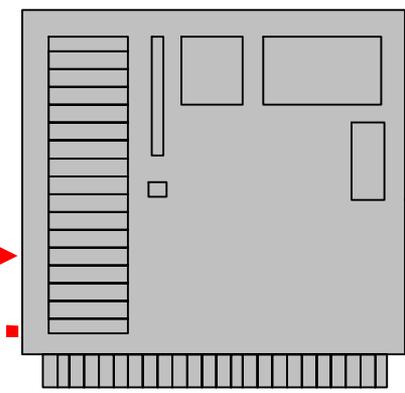
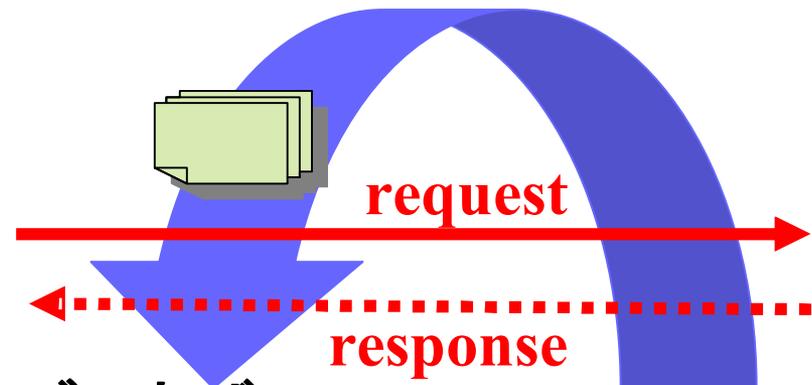
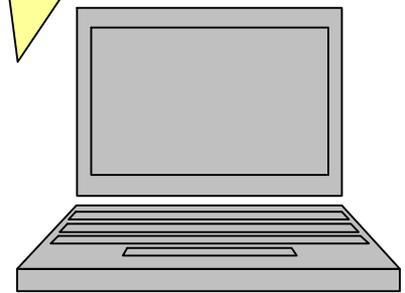
これが基本！

①記述言語:HTML



④送受信プロトコル:HTTP

情報検索
予約サービス
電子商取引
Web掲示板



③情報受信:Webブラウザ

②情報発信:Webサーバ

Javaアプレット:Webブラウザ
上で動作するJavaプログラム
(サンドボックスを適用)

Javaサーブレット:Webサーバ
上で動作するJavaプログラム
(サンドボックスを適用せず)

ご清聴ありがとうございました