永年利用可能なURLの実現手法

東京工業大学 理学部 情報科学科
上田 仁
(学籍番号 9306114)

平成8年度学士論文

指導教官 大野 浩之 講師

平成9年2月6日

目次

第 1 章

はじめに

World Wide Web(WWW) とよばれるインターネット上の情報の提供・閲覧サービスを利用する人が、近年爆発的に増加している。WWW の利用者は「ホームページ」上でさまざまな情報をインターネットに向け公開できる。ホームページという言葉に厳密な定義はないが、本論文ではホームページを「個人または組織が公開する情報の中で最初に参照されることを想定して作られた情報」とする。ホームページには、他の公開しているページにリンクが張られていることが多い。

WWW の利用者が増大するにつれ、ホームページは人々の身近なものとなり多様な目的で利用されるようになった。例えば電子メールに自分のホームページの URL を記載することは広く行われ、人々の間で日常的に交換される名刺にさえ URL が記載されることもある。また、学術論文等がホームページで公開され、他の学術論文等に参考文献としてその文献名と URL のみが記される場合がでてきた。参考文献に URL を示すことはまだ一般には認めらていないが、徐々に増えている。学術論文等は関連研究として長い間過去に遡って調べられるので、参考文献として挙げられている文献のホームページは、発表されてから何年も後で参照される可能性がある。しかし、長い年月の間にはホームページを移動させる場合がある。例えば自分のホームページを登録しているプロバイダを変更して別のプロバイダと契約した場合や、学校を卒業したことに伴いホームページをその学校に置けなくなった場合である。

現在の WWW のシステムでは、ホームページを移動させるとホームページに張られていたリンクはすべて無効になり、そのリンクを張った人はホームページを参照できない。このように World Wide Web の利用が多様化すると、長期に及んでホームページが参照される可能性が生まれた。ホームページを長期に及んで参照可能にするには、ホームページの移動に伴いリンクが外れる問題を解決しなければならない。

ホームページの移動に伴いリンクが外れる問題を解決するため、次の四つの方法を検討した。

- 移動先へのリンクを張ったページを残す方法
- ページの移動情報を一元的に管理する方法
- WWW サーバがページの変更情報を記憶する方法

- DNS を利用をする方法

これらの方法を、従来のシステムと共存できるか、ホームページが存在した元の場所に情報を残す必要がないか等の条件から評価した。その結果、最適な手法として「DNS を利用した手法」を選び実装した。

DNS を利用した手法では、各ホームページにはインターネット上で一意に区別できる「ホームページ名」を与え、ホームページのインターネット上の位置をあらわす URL と対応させた。利用者がホームページを参照するときは WWW クライアントに URL を指定する代わりに、ホスト名をホームページ名に代えた URL を指定する。この様な URL を本論文では、永年利用可能な URL と呼ぶ。ホームページが移動しても永年利用可能な URL は変化しないので、 WWW の利用者は移動前と同じ永年利用可能な URL でホームページを参照できる。まず、永年利用可能な URL を実現するシステムの問題点を明らかにし、解決方法を示した。そして実際に実装を行ない、永年利用可能な URL が有効なことを確認した。 WWW クライアントに手を加えていないため、 WWW の利用者は従来と同じ操作方法で情報の参照ができた。最後に、考察を行ない今後の課題と展望を述べた。

第 2 章

World Wide Web の利用の多様化

2.1 World Wide Web

現在、WWW(World Wide Web の略) と呼ばれるインターネット上の情報の提供・閲覧サービスを多くの人が利用している。WWW では、情報はインターネット上に分散して存在し、各々の情報は網の目のように相互に関連づけられている。この網の目のような関連づけはリンクと呼ばれる。また、網の目 (Web) のような形状が World Wide Web の名称の語源となった。このリンクによって相互に関連づけられた情報を本論文では資源と呼ぶ。資源は、文書や画像, 音楽, 映像などさまざまな形式の情報である。WWW は資源とそれを関連づけるリンクとによって複雑に絡み合った情報空間 (図 2.1) を形成している。リンクは WWW における情報の構造を示しており、リンク情報をもとに WWW の情報空間を定量的に調査する研究 [7] も行われている。

WWW においてページと呼ばれる HTML[1] [2] 文書も資源である。インターネット上の各々の資源は URL[2]によって識別される。URL はインターネット上の資源の位置をあらわす識別子である。WWW の利用者はブラウザなどの WWW クライアントに URL を指定 (図 2.2) することで目的の資源を参照 (図 2.3) できる。URL を指定された WWW クライアントは、WWW で情報をやりとりするのに使われる通信プロトコルの HTTP[3]を用いて WWW サーバと通信をし、目的の資源を得る。

URL

URL[3] は次のような形式である。

$$URL =< scheme >:< scheme - specific - part >$$

scheme は ftp や http,gopher,mailto,news,nntp,telnet,wais,file,prospero など情報をやりとりする各種の通信プロトコルである。scheme-specific-part は文字列であり shceme

---

[1] HyperText Markup Language

[2] Uniform Resource Locator

[3] HyperText Transfer Protocol

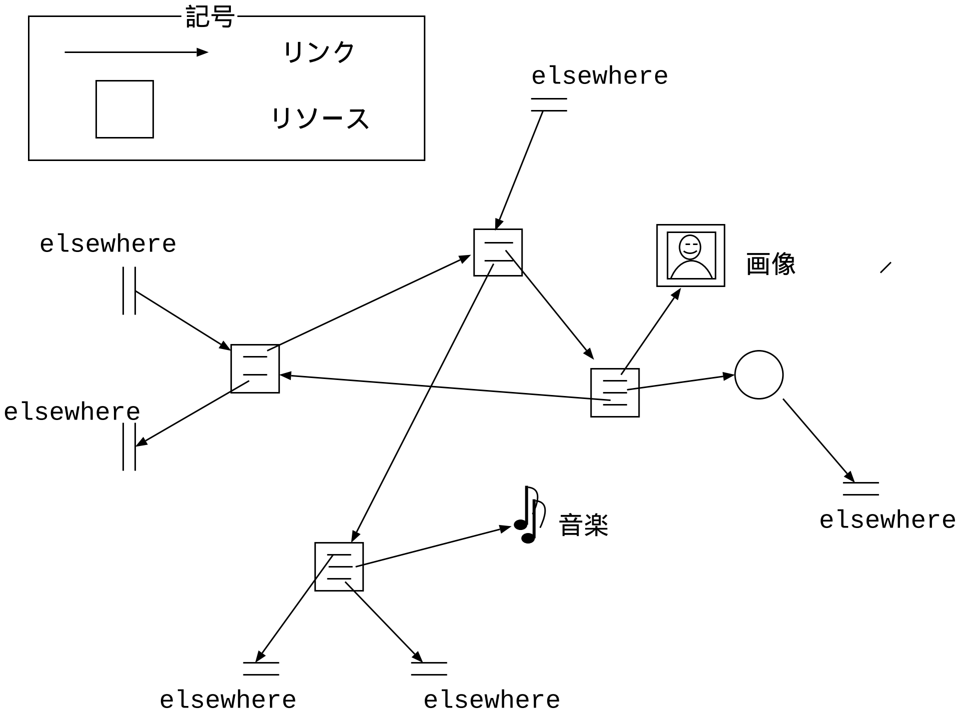


図 2.1 WWW の情報空間の例

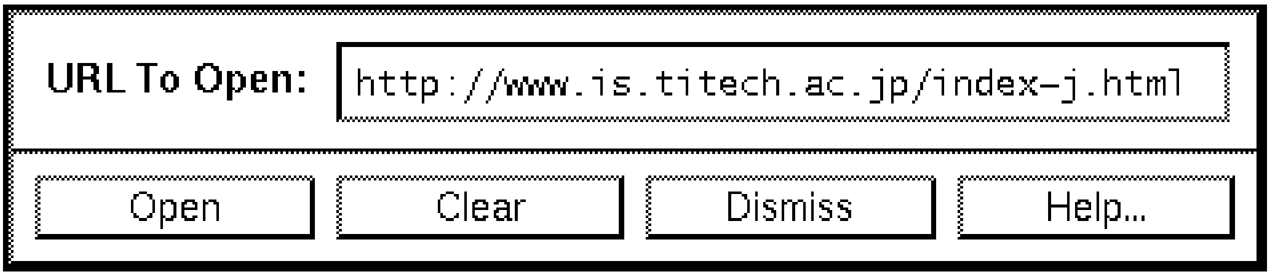
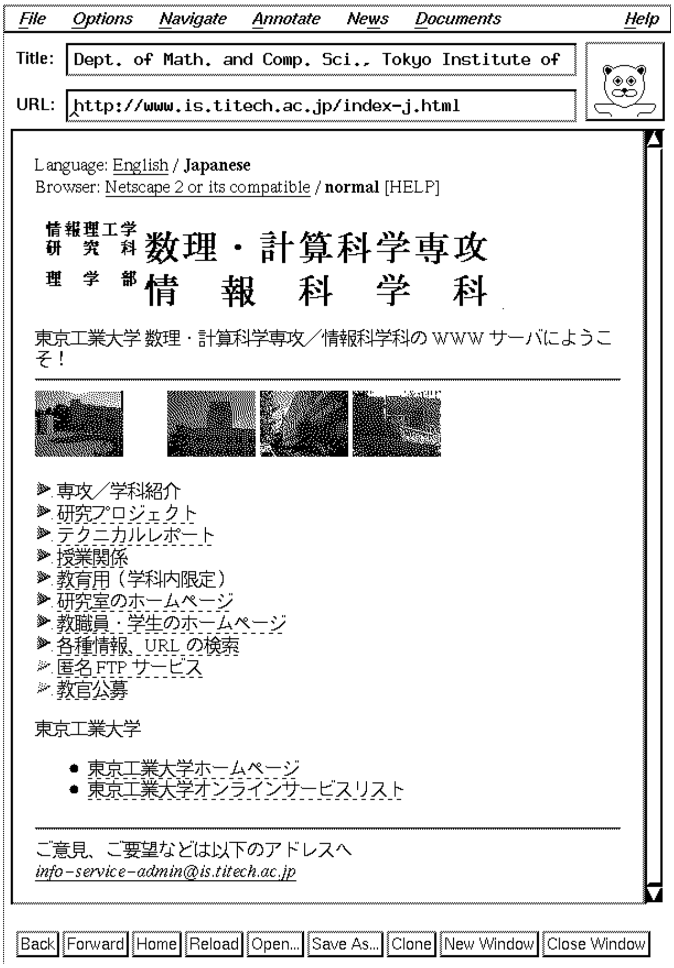

URL To Open: http://www.is.titech.ac.jp/index-j.html

図 2.2 URL の入力



Title: Dept. of Math. and Comp. Sci., Tokyo Institute of

URL: http://www.is.titech.ac.jp/index-j.html

Language: English / **Japanese**
Browser: Netscape 2 or its compatible / **normal** [HELP]

情報理工学研究科 数理・計算科学専攻
理学部 情報科学科

東京工業大学 数理・計算科学専攻／情報科学科の WWW サーバにようこそ！

- 専攻／学科紹介
- 研究プロジェクト
- テクニカルレポート
- 授業関係
- 教育用（学科内限定）
- 研究室のホームページ
- 教職員・学生のホームページ
- 各種情報、URL の検索
- 匿名 FTP サービス
- 教官公募

東京工業大学

- 東京工業大学ホームページ
- 東京工業大学オンラインサービスリスト

ご意見、ご要望などは以下のアドレスへ
info-service-admin@is.titech.ac.jp

図 2.3 情報の表示

に依存する。WWW で情報をやりとりするのに使われる通信プロトコルは HTTP である。HTTP で使われる URL は、HTTP URL2.4ともよばれ URL と同じ RFC の文献番号 1738 に定義されている。本論文で以後 URL と述べた場合は HTTP URL を指すものとする。< port >と< searchpart >は、省略可能である。

```
HTTP URL  =  http://<host>:<port>/<path>?<searchpart>
    host         インターネットドメイン表記されたホスト名
                 または IP アドレス
    port         ポート番号
    path         HTTP のセレクタ
    searchpart   質問文字列
```

図 2.4 HTTP URL の定義

$$HTTP \quad URL \; = \; http://<host>:<port>/<path>?<searchpart>$$

host インターネットドメイン表記されたホスト名または IP アドレス

port ポート番号

path HTTP のセレクタ

searchpart 質問文字列

2.2 WWW の利用の仕組み

資源は WWW サーバを通して閲覧者に公開される。WWW サーバは HTTP の要求が行われると URL のなかの< path >に従って資源を要求元に返す。WWW サーバと WWW クライアントの関係は図 2.5である。

本論文では WWW を利用する人を利用者と呼ぶ。利用者の中でも特にページを参照している人を閲覧者、ページを公開している人を公開者と呼ぶ。ある人が、WWW クライアントを利用してリンクをたどり情報を参照しているときは閲覧者であり、WWW を利用してインターネットに情報を公開しようするときは公開者である。利用者は、両者の概念を含む。

個人や組織が WWW で情報を公開するとき、HTML で記述されたページをつくる。そのページは、閲覧者がまず最初に参照し、そこから閲覧者が他のページへリンクをたどりやすいように記述されている。本論文ではこのようなページを特にホームページ (図 2.6) と呼ぶ。ホームページも資源である。

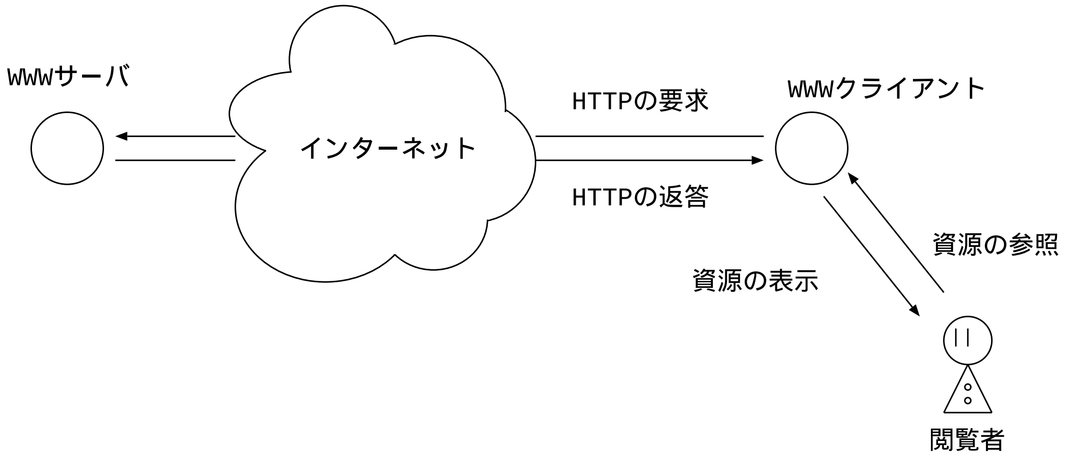


図 2.5 WWW サーバと WWW クライアントの関係

2.3 WWW の利用の多様化

2.3.1 利用者の拡大

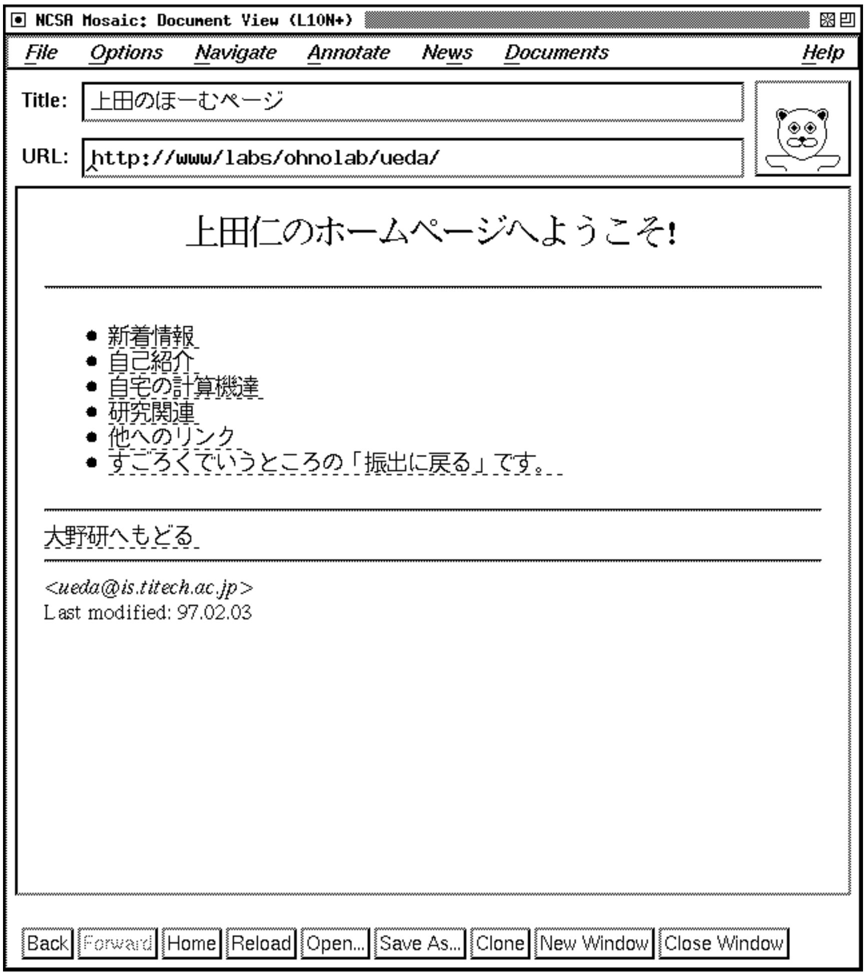
近年、 WWW のサービスを提供する WWW サーバが爆発的に増加している。

可能なかぎり多くの WWW サーバ名を収集している NETCRAFT 社の Web Server Survey [5] によれば、 WWW サーバの数の増加は図 2.7のようになる。

このように WWW サーバが爆発的に増加するにともない、 WWW の利用者も急増している。

2.3.2 利用の多様化

WWW の利用者が増加するにともない、ホームページは人々のより身近なものになった。いまや多くの人が自分のホームページを持つ。また電子メールに自分のホームページの URL を記載することは広く行われている。人々の間で日常的に交換される名刺にさえ URL が記載されることもある。 URL が名刺に記載されていると、名刺を受け取った人は好きなときに、名刺に書かれた人が公開しているホームページを参照できる。学会では事前に発表内容がアナウンスされる。このアナウンスには発表内容の資源への URL が示されることも多くなった。発表を聞く聴講者は発表までのあいだ随時更新される最新の発表内容を参照できる。また論文では参照文献として文献名と URL か書かれることもある。論文などは関連研究として長い間過去に遡って調べられるので、何年もしくは何十年も後で論文を読んだ人が資源を参照しようとする可能性がでてきた。



Title: 上田のほーむページ

URL: http://www/labs/ohnolab/ueda/

上田仁のホームページへようこそ!

- 新着情報
- 自己紹介
- 自宅の計算機達
- 研究関連
- 他へのリンク
- すごろくでいうところの「振出に戻る」です。

大野研へもどる

<ueda@is.titech.ac.jp>
Last modified: 97.02.03

図 2.6 ホームページの例

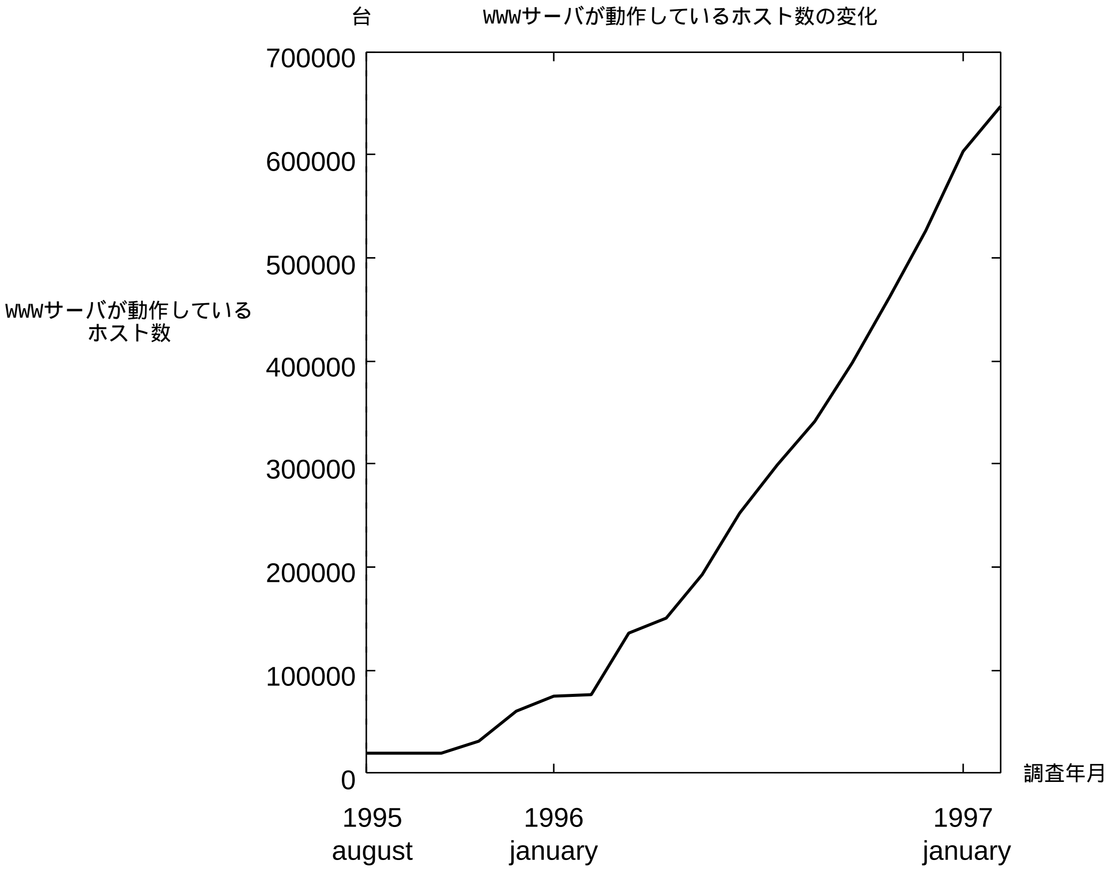


図 2.7 WWW サーバの数の推移

2.4 新たな需要

WWW の利用法が多様化するにつれ新たな需要が生まれてきた。WWW サーバはインターネットに接続された計算機で動作し、管理は教育機関や企業, 非営利組織等によって行われている。WWW を利用して資源を公開していると、何らかの都合により資源の置き場所を変更したいことがある。例えば学校等に在籍していた学生は卒業とともに資源をその学校以外の他の場所、就職した先の企業やインターネットとの接続を提供する商用プロバイダなどに移さなければならない。なぜならほとんどの学校には卒業生のために資源を置いておく余裕ながい。また、もはや学校に在籍していない者に、資源を置いておくハードディスクなどを提供することは法律的にも問題となる可能性がある。

このことは大学内に資源を置く学生だけに限らない。インターネットを利用する環境の変化が著しい現在、接続料金が安い・接続速度が早い、などの理由でいままで資源を置いていた商用プロバイダとは違う別の商用プロバイダと契約することがある。このように資源を移動させたい要求が生まれてきた。

2.5 新たな需要に対応できない問題

資源を移動させたいという新たに生まれた要求は、従来の WWW のシステムでは対応できない。WWW では資源を移動させると資源のインターネット上の位置をあらわす URL が変わるのでさまざまな問題が生じるからである。一つには、URL を指定することで張られていたリンクが無効になる。利用者は無効になったリンクをたどって資源を参照できない。もう一つの問題は古い URL では移動した後の資源を参照できないことである。先に上げた名刺の例では、名刺を渡したあと名刺に記された資源が移動したとする。名刺を受け取った側が資源を参照しようとしても、もはやできない (図 2.8)。なぜなら資源が移動したため名刺に書かれた古い URL と移動後の資源を示す URL が異なるからである。論文の例でも同じことがいえる。論文の例ではより長い年月経過したあとで参照されることもあるため、その可能性は増える。

このように従来の WWW では資源が移動してしまうと、資源に張られてリンクが無効になる問題と、古い URL を使って移動後の資源を参照できない問題とがある。これは WWW の利用方法が多様化してきた現在、無視できない問題である。次章ではこれらの問題を解決する手法を提案する。

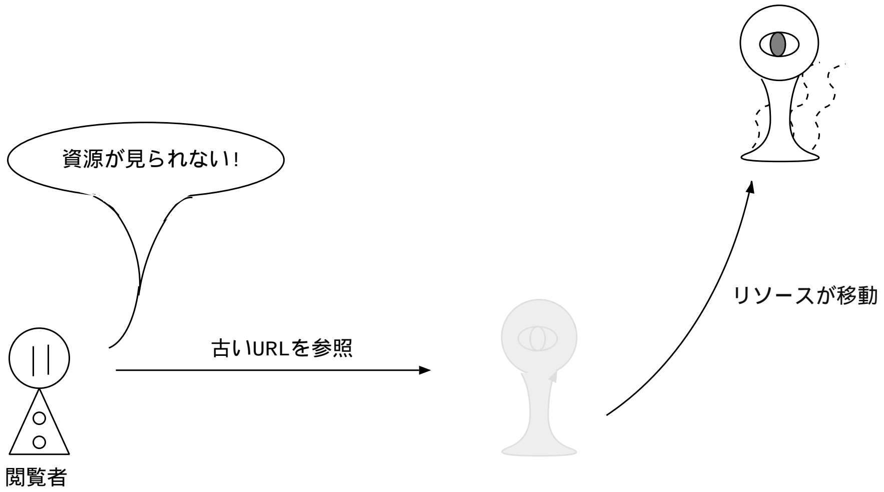


図 2.8 古い URL での参照

第 3 章

資源の移動手法

この章では張られているリンクを維持しつつ資源を移動する方法を考案する。まず資源を移動させるときに満たすべき必要な条件を明らかにする。つぎに考えられる方法を列挙する。最後に各方法の優劣を表にまとめその中から最適な方法を選ぶ。

3.1 必要な条件

前章で述べた問題を解決するためには、以下に挙げる条件を満たしつつ資源を移動できる方法が必要である。

- 移動前の URL を指定して移動後の資源が参照できる。
- 従来と同じ方法で資源を参照できる。
- 従来と同じ方法で資源を公開できる。
- 既存の資源に適応できる。
- 従来のシステムと共存できる。
- 資源が存在した元の場所に情報を残さない。
- トラフィックを増やさない。

移動前の URL を指定して移動後の資源が参照できる

資源のインターネット上での位置は URL によって示されている。資源を移動させると、移動後の資源のインターネット上の位置は新しい URL によって示される。このように、移動前の URL を古い URL とよび、移動後の URL を新しい URL とよぶ。“移動前の URL を指定して移動後の資源が参照できる” とは古い URL を指定して移動後の資源が参照可能なことをいう。

従来と同じ方法で資源を参照できる

現在の WWW と同じ操作で資源を参照できることをいう。リンクをたどり自由に資源を参照でき、直接 URL を WWW クライアントに指定することでもリースの参照が可能である。

従来と同じ方法で資源を公開できる

現在の WWW と同じ操作で資源を公開できることをいう。個人や組織が WWW に情報を公開するとき、ホームページからさまざまなメディアの情報に向けてリンクを張ることで行う。

既存の資源に適応できる

すでに WWW 上に存在する資源にたいして" リンクを維持した資源の移動を可能にする手法" を適応することでリンクを維持した自由な資源の移動が可能なことをいう。

従来のシステムと共存できる

“リンクを維持した資源の移動を可能にする手法” を適応した資源と、現在の資源も同様なシステムで共存して扱える。

ホームページが存在した元の場所に情報を残さない

WWW でリンクをたどりページを参照していると図 3.1のように “このページは移動しました” と書かれていて、その移動先へのリンクが張ってある資源がある。これは資源が移動したために新しい資源に移動してもらうための資源である。

この様に本論文では移動先の資源へ移動させるためだけに置く資源を" 足跡" と呼ぶ。“ホームページがあった元の場所に情報を残さない” とはこの様な足跡を残さないことを言う。

トラフィックを増やさない

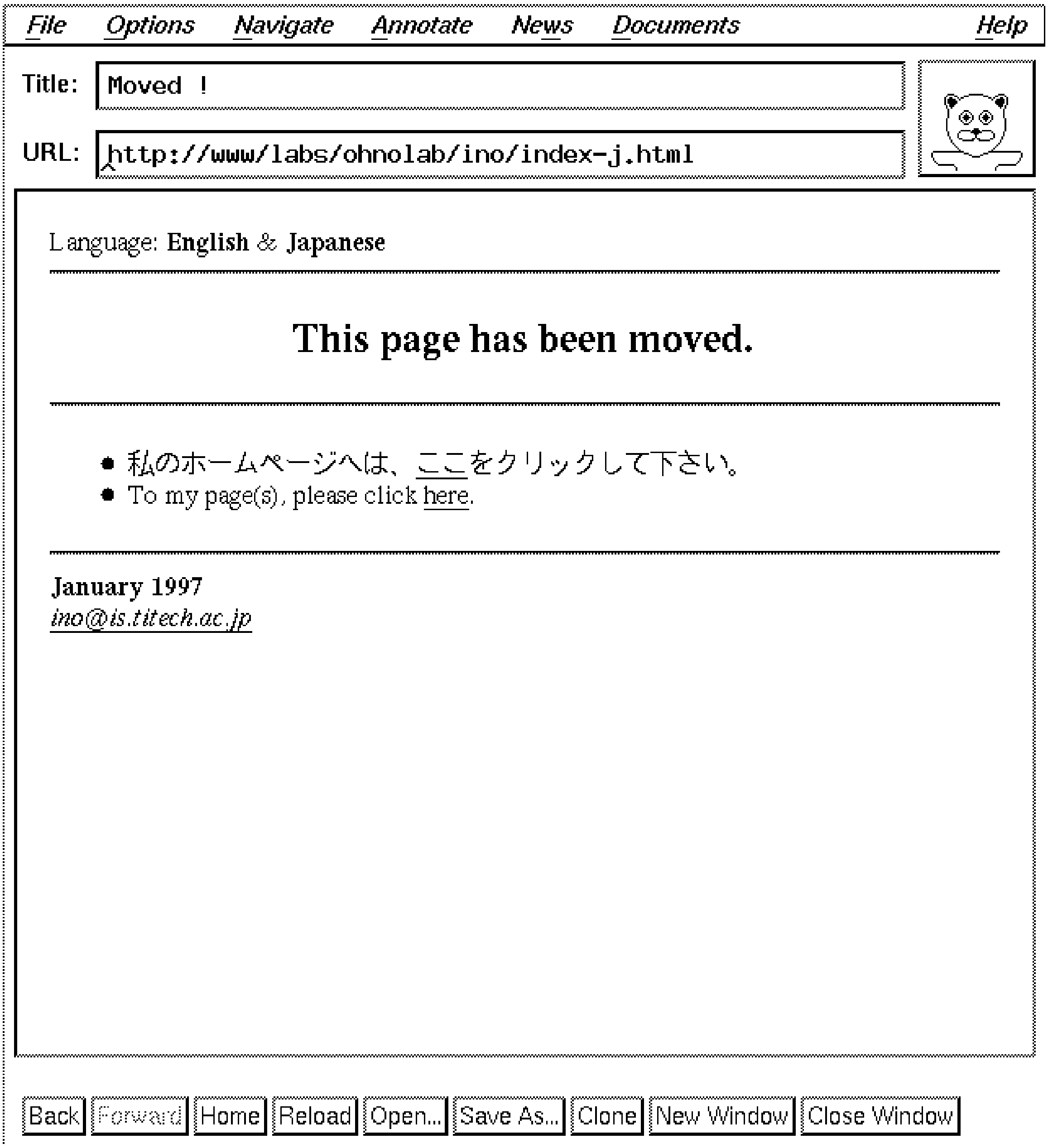
WWW の利用者は爆発的に増加し、 WWW が使用する通信プロトコルの HTTP はネットワークトラフィックの第一位を占めるまでになった。ゆえにこれ以上のトラフィック増加は避ける。

3.2 さまざまな実現手法

現在行われている方法を含め、今回検討した手法を 4 つ挙げる。

3.2.1 移動先へのリンクを張った資源を残す方法

インターネット上で広く行われている方法である。資源をを移動させるとき移動先へのリンクを張った資源を新たにつくり元の URL があらわす所に置く (図 3.2)。移動した資



Title: Moved !

URL: http://www/labs/ohnolab/ino/index-j.html

Language: **English** & **Japanese**

This page has been moved.

- 私のホームページへは、ここをクリックして下さい。
- To my page(s), please click here.

January 1997
ino@is.titech.ac.jp

図 3.1 足跡のページ

源を参照する閲覧者は足跡の資源をまず参照する。その後足跡から張られたリンクをたどり目的の資源を参照する。

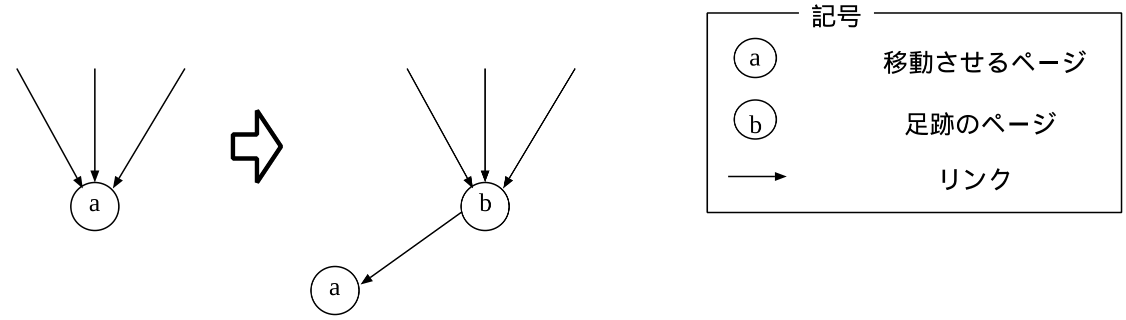


図3.2 移動先へのリンクを張った資源を残す方法

WWWクライアントのnetscapeなどは、指定された時間が経過した後に特定のURLが示す資源に自動的に飛ぶ機能[1]があるので、足跡の資源を表示したのち移動先のページに行くこができる。さらに別の機能[2]では、閲覧者が見ようとする資源が移動していることを知ると、移動先の資源を収得し足跡を画面に表示することなしに移動先のページを直に表示する。このとき利用者は要求した資源が移動していたことを意識しない。

いずれにせよ、移動が何度も繰り返されるたびに足跡が残る。どこか途中の足跡がなくなるとそれ以前につけられた足跡をたどって移動先の資源を参照することはできない。資源の移動が繰り返されるとその分だけ途中の足跡が消失する危険性が高くなり最終的な移動先の資源にたどり着ける可能性が減る。ゆえに移動した資源が存在する限りその資源が残した全ての足跡を保存しておかなければならない欠点がある(図3.3)。

3.2.2 資源の移動情報を一元的に管理する方法

全ての資源の移動情報を管理するシステムをつくり、そのシステムに資源の移動前の位置と移動後の位置を保持させる。WWWクライアントは資源を参照するとき常にその一元管理するシステムに問い合わせ、目的の資源が移動しているかどうかを調べる。目的の資源が移動していたら移動先を参照する(図3.4)。目的の資源が移動していないならそのまま資源を参照する。資源の移動が行われるたびに、一元管理するシステムでは移動情報を書き換える。

この方法ではWWW上で移動した資源全ての移動情報が一元管理するシステムに蓄積される。基本的に移動情報が削除されることはないので蓄積しなければならない移動情報は増え続け、膨大な量になることが予想される。また全てのWWWクライアントは資源を参照するたびに移動情報を一元管理するシステムに問い合わせるので一元管理するシス

---

[1] METAタグにより起動

[2] location機能

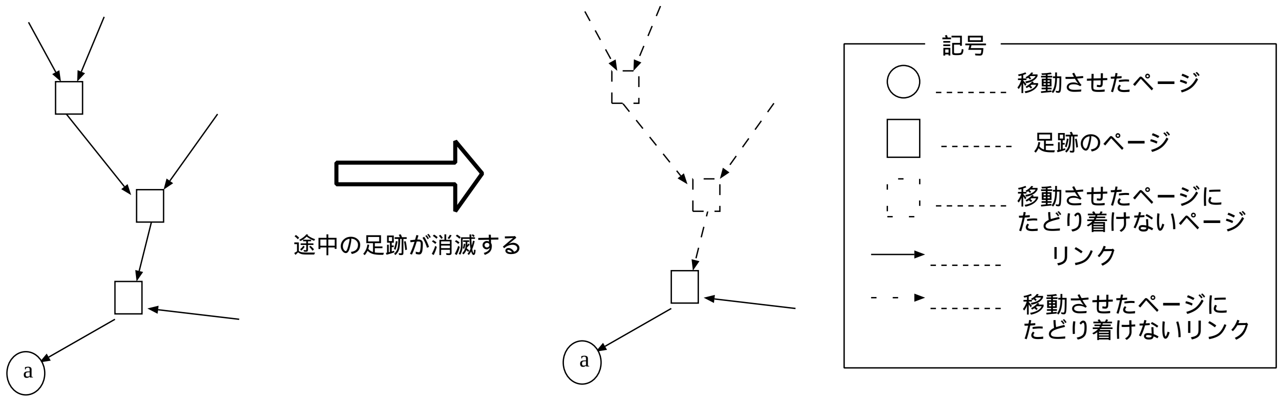


図 3.3 足跡の消失

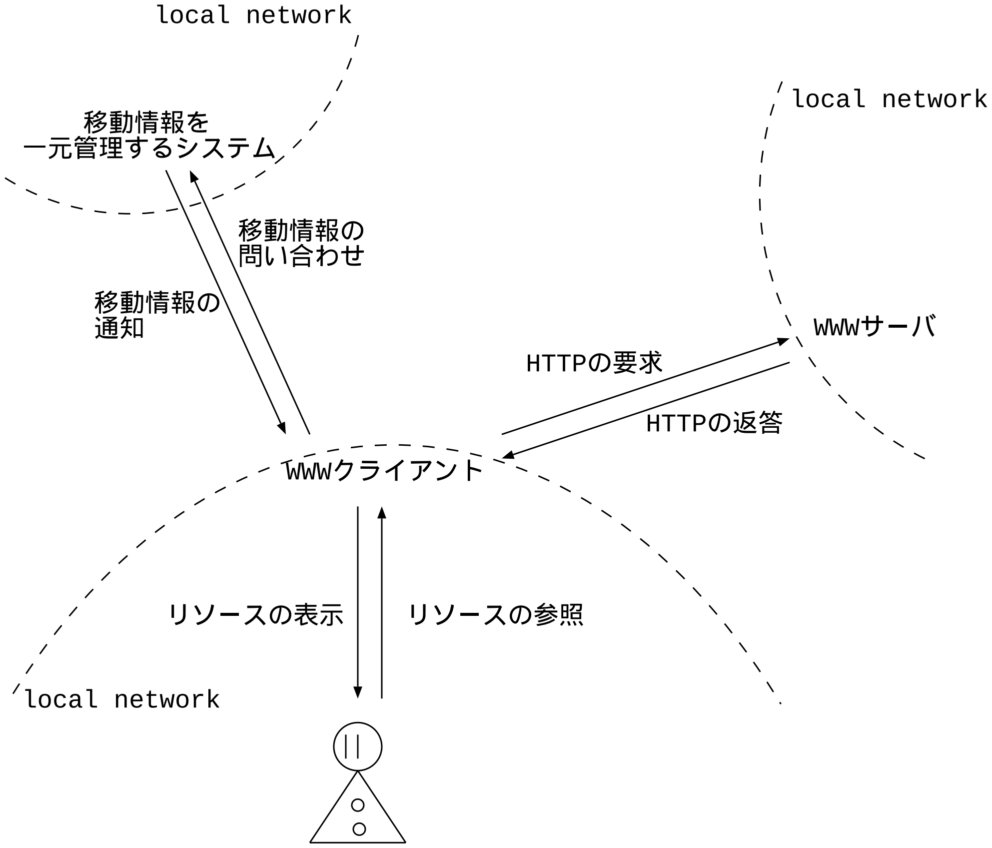


図 3.4 一元的に管理したときの参照

テムが動作するネットワークとその周辺のネットワークではトラフィックが非常に増大する。そしてもう1つの問題は、移動情報を管理するシステムに必ず問い合わせをしなければならないことであり、その機能のない既存のWWWクライアントは対応できない。

3.2.3 WWWサーバが資源の変更情報を記憶する方法

WWWクライアントから資源の参照要求があると、WWWサーバは自分が管理する資源を返答する。このとき返答する資源から張られているリンクをあらかじめ全て先読みをして調べ、そのリンク先の資源が移動しているか調査し結果を移動情報として記憶する。リンク先が移動しているとき、WWWクライアントへ返す資源中のリンク先情報を移動先の新しいリンク先へと内容を書き換えて返答する。資源のの移動情報を得たWWWサーバはそれを記憶し、次からはその情報を優先する。

問題点はまず先読みによるトラフィックの増加とページの移動情報をいつまでも保持していなければならない点があげられる。そして最大の問題点は利用者が直にURLを指定して資源を参照するときリンクの移動に対応できないことである。

3.2.4 DNSを利用をする方法

DNS[1]はホスト名と、ネットワーク上の計算機の位置を示すIPアドレスとの関連付けを管理する分散データベースである。DNSではインターネットドメイン表記されたホスト名からIPアドレスを得ることができる。

現在のWWWのシステムでは、WWWクライアントは目的のWWWサーバが動作している計算機のインターネットドメイン表記された名前をURLから知り、DNSを利用してのネットワーク上の位置を特定する。DNSでは計算機にホスト名の別名を割り当てることができるので「DNSを利用する方法」ではこの仕組みを応用する。まず移動する予定の資源に固有の資源名をつける。この名前と同じ名前のホストをDNSの機能で架空に作る。この「架空」のホストと、実際に資源が管理されている「実在」のホストとを関連づける。

ホームページを移動させるときは、まず実際に資源を移動させる。そして架空ホストと「実在」のホストとの関連づけを「架空」のホストと、移動後の資源を管理しているWWWサーバが動作している「実在」のホストとの関連づけに変更する。リンクは資源固有の資源名に張られているので、リンクを変更することなしに移動後の資源を参照することができる。

この方法の欠点はあらかじめ資源を固有の資源名に割り当てていなければならない点である。

3.3 最適な実現手法の選択

前に述べた4つの方法から、張られていたリンクを維持しつつ資源を移動する最適な方法を選択する。評価の基準はこの章の冒頭で述べた満たすべき要素である。4つの方

法をそれぞれ評価し利点と欠点を表にまとめた (図 3.5)。図中で方法 A ~ D は次の方法に対応する。

方法 A 移動先へのリンクを張った資源を残す方法

方法 B 資源の移動情報を一元的に管理する方法

方法 C WWW サーバが資源の変更情報を記憶する方法

方法 D DNS を利用をする方法

	方法 A	方法 B	方法 C	方法 D
移動前の URL を指定して移動後の資源を参照できるか	△	○	×	○
従来と同じ方法で資源を参照できるか	○	○	○	○
従来と同じ方法で資源を公開できるか	○	○	○	○
既存のページに適応できるか	○	○	○	×
従来のシステムと共存できるか	○	×	×	○
ホームページがあった元の場所に情報を残さないか	×	○	○	○
トラフィックを増やさないか	△	×	△	○
総合評価	△	×	×	○

図 3.5 各方法の利点と欠点

この表を見ても分かるとおりどの方法も WWW 上でリンクを移動させると張られていたリンクが無効になる問題の決定的な解決方法ではない。その主な原因は、ネットワーク上の資源の位置は URL によってあらわされるのに URL を変えずにネットワーク上の位置を変えようとするからである。理想的には資源を一意に識別できる資源の内容だけに依存した名前を資源につけ、その資源名から資源のインターネット上の位置を知ることができれば良い。現実には世界中のさまざまな資源に一意に識別できる名前をつけることは大変困難である。またその名前からインターネット上の位置を知るのは難しい。

以下では資源をホームページのみに限定して議論する。ホームページは資源の一部なのでホームページでの議論は簡単に資源にも適応できる。

ホームページの名前とそのホームページが置かれている場所を示す URL を関連づける情報を管理する必要がある。この情報は膨大な量になると予想されるので、分散型のデータベースで管理することが望ましい。その意味で DNS を用いる方法は適切である。 DNS を用いた方法は既存のページに対して適応できない欠点がある。しかし既存の WWW のシステムと共存できるため、新たにホームページを公開する場合に従来のホームページと共存ができるので、移行が容易である。ホームページの公開者は従来のシステムを使うか本論文のシステムを使うかの選択を自由に選ぶことができる。この意味で DNS を用いた方法の欠点は問題がないと考える。本論文ではリンクを維持した資源の移動を実現するために DNS を利用した手法を実装し評価する。

第4章

永年利用可能な URL の実現

前章では、リンクを維持しつつ資源を移動するために DNS を利用することを述べた。この章では実際にシステムを実装する。

4.1　ホームページ名

ホームページを「永年利用可能」にするには、ホームページを一意に識別する「ホームページ名」が必要であり、それは適切な名前空間を形成しなければならないことを述べる。「永年利用可能」とは、ホームページを移動させても移動前と同じ識別子で参照できるため長年に及んで利用できることをいう。既存の WWW のシステムでホームページ名を利用するには、URL と対応づけなければならない。そのとき生じる問題を述べ、解決方を示す。

4.1.1　ホームページ名の必要性

ホームページを永年利用可能にするには、ホームページをあらわす識別子がホームページを移動しても変化しない必要がある。このため本システムではホームページ名とよぶ、ホームページを移動しても変化しない識別子を各ホームページに与え、WWW の利用者は常にそのホームページ名を利用する。ホームページ名は存在する全てのホームページを一意に区別する。

4.1.2　ホームページ名がつくる名前空間

ホームページ名は永年利用可能で、かつ存在する全てのホームページを一意に区別する。ホームページ名に、ホームページを管理している人が所属する組織の名前や、ホームページが置いてある場所の名前を含めることはできない。なぜなら場所も所属も年月の経過で変わる可能性があるからである。各々のホームページは年月の経過により変化しない名前で区別される名前空間を形成しなければならない。

4.1.3 ホームページ名と URL の差異による問題点とその解決方法

既存の WWWW のシステム上でホームページ名を利用するために、URL の中でホスト名の代わりにホームページ名を指定する。その場合生じる問題点と、解決法を述べる。

ホームページ名を利用したホームページの参照

既存の WWWW のシステム上でホームページ名を利用するために、URL 内でホスト名の代わりにホームページ名を指定する。この様な URL を本論文では、永年利用可能な URL と呼ぶ。

- `URL = http://<host>:<port>/<path>?<searchpart>`
- 永年利用可能な `URL = http://<homepagename>:<port>/<path>?<searchpart>`

永年利用可能な URL が WWW クライアントに入力されたら、WWW クライアントはそのホームページ名から目的のホームページを管理する WWW サーバが動作しているホスト名を調べ、そのホストに要求を出す。ホスト名の代わりにホームページ名が指定されるため、ホスト名の名前空間内にホームページ名の部分空間を作る。ホームページ名の名前空間内にあるホームページ名はすべて、ホスト名の名前空間内にある実際にホームページがあるホスト名に対応づける。この対応づけにより WWW クライアントはホームページ名からホスト名を収得することができる。

HTTP の要求にホームページ名が含まれない問題

WWW サーバが受けとった HTTP の要求にはもはやホームページ名が含まれていない。なぜなら DNS によりホームページ名は実際のホスト名に置き換えられてしまうからである。このため要求を受け取った WWW サーバは、ホームページ名からホームページのホスト内の位置を知ることができない。ホームページはホスト内に幾つもあるのでどのホームページを要求されたか WWW サーバは知ることができないためである。

問題の解決法

しかし現在[1]広く使われている HTTP/1.0 の次のバージョンである HTTP/1.1 では、要求には後述するホストヘッダフィールドが必ず付加されるため WWW サーバはホームページ名を収得できる。すでに Apache/1.2[6] が HTTP/1.1 に対応している。また、WWW クライアントの netscape-3.01 がすでに、ホストヘッダフィールドを要求に付加することを筆者は確認している。他の WWW サーバや WWW クライアントも、早い内に HTTP/1.1 に対応すると予想される。よって本システムでは通信プロトコルとして HTTP/1.1 を前提とする。

---

[1] 1997 年 1 月

HTTP/1.1

HTTP/1.1 は 1997 年 1 月に出された RFC の 2068 番 [4] で勧告された。 HTTP/1.1 と HTTP/1.0 の主な変更点の一つは、 WWW のクライアントからサーバに送られるリクエストにホストヘッダフィールドをつけなければならない点である。 RFC2068 番によればホストヘッダフィールの定義は以下である。

$$Host = "Host" \;\; ":" \;\; host \;\; [\;\; ":" \;\; port \;\;]$$

ホストヘッダフィールドは利用者が入力した URL のホスト名とポート番号をふくむ。たとえば

`<http://hostname/directory1/directory2/>`

で示される要求は、以下のようにホストヘッダフィールドをもつ。

```
GET /directory1/directory2/ HTTP/1.1
Host: hostname
```

である。要求を受け取った WWW サーバは、この要求がホスト名 hostname にたいして行われたものであることが分かる。ポート番号は省略可能であり省略された場合はデファルト値が使われる。

計算機上の位置の収得

ホスト名を含んだ要求を受け取った WWW サーバはホームページ名に対応するホームページを応答する。 1 つのホスト中には多数のホームページが置かれるので、 WWW サーバはホームページ名に対応するホームページを選ばなければならない。既存の WWW サーバにはそのような機能はない。よって WWW クライアントと WWW サーバの間に「振り分けサーバ」を介在させる。振り分けサーバの位置は、 WWW サーバと同一セグメント上かその近辺にに置く。

振り分けサーバは HTTP の要求内にあるホームページ名の情報から、 HTTP の要求内に存在するパス情報にそのホームページが存在する位置情報を追加する。パス情報は request-URI とよばれるもので、 WWW サーバが動作するホスト中にどの位置に要求する資源があるかの情報をしめす。振り分けサーバについては後で詳しく述べる。

情報の応答

WWW サーバは要求された情報を振り分けサーバを介在して WWW クライアントに返す。

4.1.4　ホームページの移動

ホームページを移動させると、ホームページが置いてあるホストとホスト内の位置が変わる。ホームページ名を変えずに移動したホームページを参照するには、ホームページ名とホスト名の対応づけと、ホームページ名とホスト内の位置との関連づけを変えなければならない。

ホームページ名とホスト名との対応づけの変更

ホームページ名とホスト名はDNSの機能により対応づけられているので、この対応づけを変えるにはDNSの設定を変えればよい。DNSの設定を変えるサービスは「CNAME変更デーモン」が行う。CNAME変更デーモンは後で詳しく述べる。

ホームページ名とホスト内の位置の対応づけの変更

ホームページ名とホスト内の位置は振り分けサーバにより対応づけられているので、この対応づけを変えるには振り分けサーバの設定を変えればよい。

4.2　永年利用可能な URL を実現するシステム

永年利用可能なURLを実現する。まずシステム全体図を考え、つぎにシステムの構成要素を述べる。

4.2.1　永年利用可能な URL を実現するシステムの全体図

永年利用可能なURLを実現するシステムの全体図は図4.1である。図中の○で囲われた数字は次の手順を示す。

1. ホームページ名からホスト名の問い合わせ
2. ホスト名の応答
3. frisへホームページの要求
4. request-URIを変更した後、WWWサーバへホームページの要求を転送
5. frisへホームページの応答
6. WWWクライアントへホームページの応答を転送

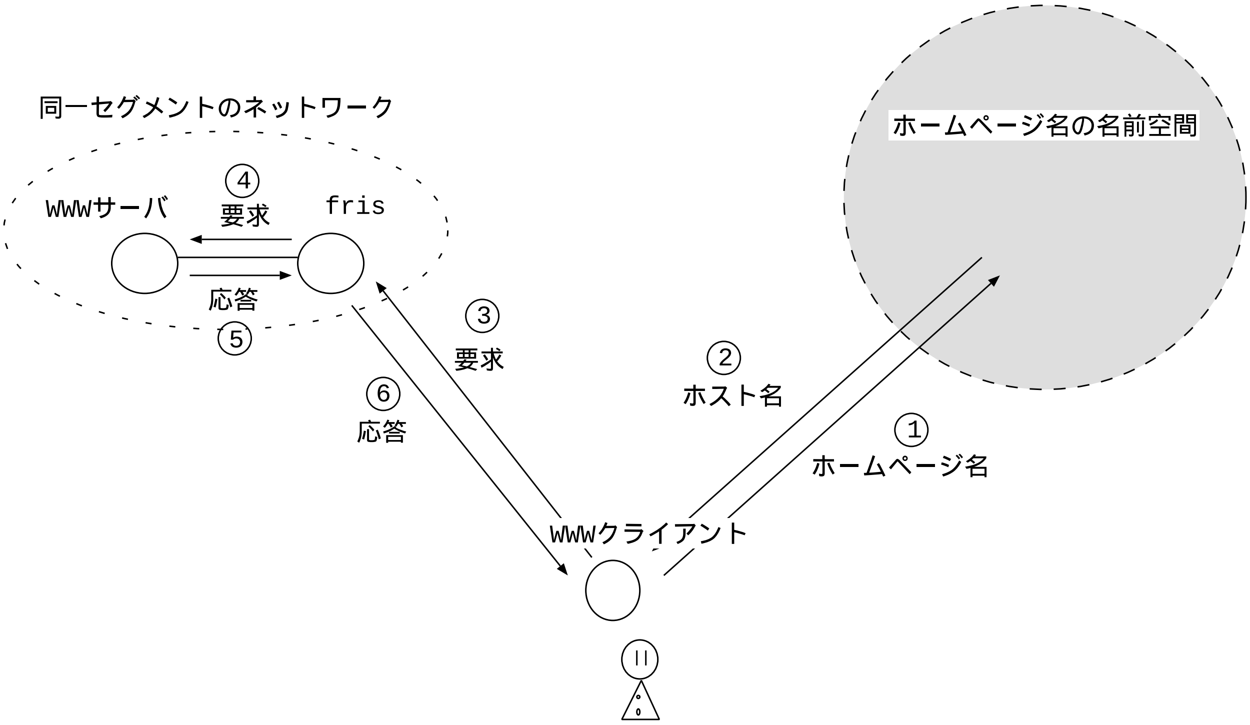


図 4.1 システムの全体図

4.2.2 システムの構成要素

振り分けサーバ (fris)

WWW サーバとネットワーク的に近い位置で動作するサーバであり、機能は次の通りである。 WWW クライアントと WWW サーバとの間に介在しデータのやりとりを仲介する。

- HTTP のプロトコル中の request-URI を書き換えてデータを中継する。
- ホームページが移動したらホームページ名と、ホームページのホスト上の位置との対応づけを変更する。

HTTP プロトコルの中継

1. WWW クライアントからの HTTP/1.1 の要求を受ける
2. 要求中の request-URI を振り分け設定ファイルに従って書き換える
3. 書き換えた要求を WWW サーバへ送る
4. WWW サーバから応答を受ける
5. WWW クライアントにその応答を返す

対応づけの変更

ホームページの公開者からの要求に従ってホームページ名と、ホームページのホスト中の位置との関連づけを変更する。

CNAME 変更デーモン (ccd)

ホームページ名の名前空間内で動作する次の機能をもつデーモンである。

- DNS の設定の変更

DNS の設定の変更

ホームページが移動したとき、ホームページの公開者はホームページ名の名前空間内の DNS の設定を変更しなければならない。ホームページ名が移動前のホスト名に対応づけられていたのを移動後のホスト名に新たに対応づける。

第 5 章

実験および評価

この章では実際にホームページ名の名前空間を作り、振り分けサーバ (fris), CNAME 変更デーモン (ccd) の各システムを実装し、本システムの評価を行った。

5.1 実験環境の設定

ここでは本システムの実験環境を述べる。

5.1.1 ホームページ名の名前空間

ホームページ名の名前空間として wwwhps.is.titech.ac.jp ドメインを設定した。
wwwhps.is.titech.ac.jp ドメインは is.titech.ac.jp ドメインのサブドメインである。
また wwwhps.is.titech.ac.jp ドメインではネームサーバを動作させた。ホームページ名とホスト名との対応づけは DNS の機能の 1 つ、 CNAME による別名機能を利用した。 DNS の設定ファイル中には、このドメインにあるホームページ名の数だけ次のような行が書かれている。

```
ホームページネーム    IN    CNAME    nsk.is.titech.ac.jp.
```

ホームページ名の名前空間に登録したホームページは、大野研究室に所属する人が公開する個人のホームページである。これらのホームページは、 www.is.titech.ac.jp と名付けられている計算機上で動作する WWW サーバによって公開されている。

fris

大野研究室に所属する人が公開しているホームページを管理している WWW サーバと同一セグメントにある計算機上で fris を動作させた。 firs は次の動作をする (図 5.1)。

1. WWW クライアントから要求を受ける
2. 振り分け設定ファイルに従って要求の中の request-URI を書き換え、 www.is.titech.ac.jp 上で動く WWW サーバに送る。

3. WWW サーバからの応答を受ける

4. 応答を WWW クライアントに送り返す。

設定ファイル (図 5.2) の各行はホームページ名と、そのホームページのホスト中の位置をあらわす文字列とからなる。先頭が # で始まる行はコメント行であり、何ら fris に影響を与えない。

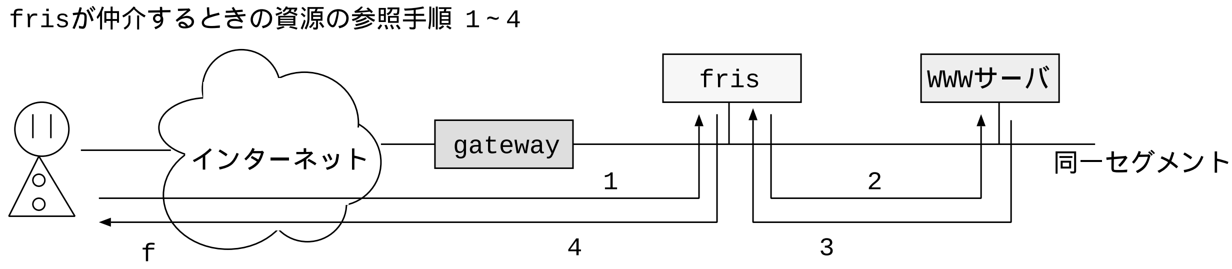


図 5.1 WWW サーバと fris

ccd

ccd への利用者のホームページ変更要求はメールを通して行う。メールを処理する関係から ccd はメールサーバ上で動作させる。メールでホームページの変更要求がきたら、 DNS が動作している計算機上の DNS の設定ファイルを変更し、変更の新旧を区別するシーケンス番号を更新する。そののち新しく変更された設定を有効にするために、ネームサーバを再起動する。

5.1.2　ホームページの参照

上記のように設定した本システムを使ってホームページ名指定することでホームページが参照できることを確認した。現在では HTTP/1.1 に従った要求を出す WWW クライアントが存在しないので、 telnet アプリケーションを使用し、手動で HTTP/1.1 の要求を出した。要求と応答の部分をログとして図 5.3に示す。この応答のログを WWW クライアントが解釈すると、第 2 章図 2.3の画面が表示される。

5.1.3　ホームページの移動

上記のように設定した本システムを使って ueda のホームページを移動した。 WWW サーバ内でホームページの位置を変えるだけなら、振り分け設定ファイルを書き換えるだけであった。ホームページが移動後、他の WWW サーバの管理下に入るときは wwwhps.is.titech.ac.jp ドメインの DNS の設定を変える必要があった。

```
# 振り分けサーバ用設定ファイル
# シャープが行頭にあったらコメント行です。
#
# ホームページ名 資源の位置
#
hohno.wwwhps.is.titech.ac.jp       /labs/ohnolab/hohno/
onogi.wwwhps.is.titech.ac.jp       /labs/ohnolab/onogi/
kimoto.wwwhps.is.titech.ac.jp      /labs/ohnolab/kimoto/
yoshiaki.wwwhps.is.titech.ac.jp    /labs/ohnolab/yoshiaki/
narita

.wwwhps.is.titech.ac.jp      /labs/ohnolab/narita/
masaki.wwwhps.is.titech.ac.jp      /labs/ohnolab/masaki/
honjo.wwwhps.is.titech.ac.jp       /labs/ohnolab/honjo/
monma.wwwhps.is.titech.ac.jp       /labs/ohnolab/monma/
noda.wwwhps.is.titech.ac.jp        /labs/ohnolab/noda/
ino.wwwhps.is.titech.ac.jp         /labs/ohnolab/ino/
ueda.wwwhps.is.titech.ac.jp        /labs/ohnolab/ueda/
koreeda.wwwhps.is.titech.ac.jp     /labs/ohnolab/koreeda/
sakai.wwwhps.is.titech.ac.jp       /labs/ohnolab/sakai/
tada.wwwhps.is.titech.ac.jp        /labs/ohnolab/tada/
tatuya.wwwhps.is.titech.ac.jp      /labs/ohnolab/tatuya/
```

図 5.2 振り分け設定ファイル

```
>telnet ueda.wwwhps.is.titech.ac.jp 6789
GTrying 131.112.40.11...
Connected to nsk.is.titech.ac.jp.
Escape character is '^]'.
GET / HTTP/1.1
Host: ueda.wwwhps.is.titech.ac.jp:6789

HTTP/1.0 200 Document follows
Date: Tue, 11 Feb 1997 14:09:26 GMT
Server: NCSA/1.4.2
Content-type: text/html

<head> <title>Ueda's home page</title> </head>

<BODY Bgcolor="ffffff">


<center><h1>>eED$N%[!<%‘%Z!<%8$X$h$&$3$=!</h1></center>

<hr>
<ul>
  <li> <a href="new.html"> ?7Ce>pJs </a>
  <li> <a href="introduction.html"> <+8J>R2p </a>
  <li> <a href="research.html"> 8&5f4XO" </a>
  <li> <a href="link.html"> B>$X$N%j%s%/ </a>
</ul>

<hr>
<a href="http://www/labs/ohnolab/"> BgLn8&$X$b$I$k </a>

<hr>
<address> &lt;ueda@is.titech.ac.jp&gt; </address>

Last modified: 97.02.11

Connection closed by foreign host.
```

図 5.3 要求と応答のログ

5.2 考察

第3章ではリンクを維持しつつ資源を移動する手法が満たすべき条件を述べた。この章では本システムがその条件を満たしているか考察する。

5.2.1 リンクを維持しつつ資源を移動する手法が満たすべき条件

移動前の URL を指定して移動後の資源が参照できる

永年利用可能な URL でホームページにリンクを張れば確かに参照できる。

従来と同じ方式で資源を参照できる

閲覧者は従来と同様に、リンクをたどることで資源を参照可能であり、かつ URL を指定する方法でホームページを参照できる。

従来と同じ方式で資源を公開できる

ホームページの公開者は従来と同様に HTML 文章を利用してホームページを公開できる。

既存の資源に適応できる

3章でも述べた通り、すでにリンクが張られている資源には本システムを適応できない。

従来のシステムと共存できる

この章では大野研の人にシステムを適応したが、従来の WWW システムと共存可能であった。このシステムは従来のシステムと共存できる。

資源が存在した元の場所に情報を残さない

ホームページの移動にともなう作業は、振り分けサーバ設定ファイルの変更と CNAME の関連づけの変更だけであり、資源が存在した元の場所に情報を残す必要はない。

トラフィックを増やさない

振り分けサーバを介在する分だけトラフィックは増える。トラフィックの増加が問題となる程度かどうか、トラフィックの測定を行なわなければならない。

5.2.2 WWW サーバの管理者から見た本システムの利点

本システムを導入するにあたり WWW サーバの管理者がすることは、振り分けサーバを導入することだけである。 WWW サーバには、何ら変更を加える必要がない。しかも振り分けサーバを導入するのは、2つの点で抵抗が少ないと思われる。ひとつは振り分けサーバは request-URI を変更する他は HTTP プロトコルを単に中継するだけだからで

ある。ふたつめは fris はインタプリタの perl で実装したため管理者は比較的容易にプログラムを解析でき fris の細かい動作まで把握できるためである。

5.2.3 DNS の拡張

DNS はインターネットドメイン表記されたホスト名のデータベースである。 DNS を拡張して、インターネット上に存在する資源のデータベースをつくる方法もある。このとき WWW クライアントが、拡張した DNS を引くと、ホームページが存在するホスト名とホスト上の位置が分かるため、 fris がもはや必要なくなる。しかし、資源のデータベースはホスト名のデータベースより巨大になることから、本当に現在の DNS を単純に拡張するだけで対応できるか不明である。

第 6 章

今後の課題と展望

6.1 課題

本システムが持つ課題を列挙する。

- 認証機能の実装

  ホームページを移動させる公開者は ccd にホームページ移動要求を出す。しかしホームページを移動できるのは確かにそのホームページの公開者のみでなければならない。このため ccd は移動要求が本人から出されたものであることを認証する必要がある。移動要求はメールで行なわれるため、認証にはメールに電子署名を付加する予定である。

- システムの信頼性の継続的な向上

  本システムはネットワークサービスを行うので、一般のシステムに比べてより高い信頼性を求められる。常に継続的な信頼性の向上が望まれる。

- ホームページの適切な名前付けによる名前空間の構築

  ホームページ名の名前空間の中での名づけを適切に行わなければならない。このホームページ名は永年利用可能にするために、所属や地域など年月とともに変わる要素を名前に付けられないので適切な名前空間を作ることが難しい。

- ホームページ振り分けサーバが介在する分の遅延時間の計測

  WWW サーバと WWW クライアントの間に振り分けサーバを介在させるので、振り分けサーバがないときに比べいくらかの遅延が発生する。この遅延があまりに大きいと閲覧者はそのホームページを見ようとしないだろう。よって実状を知るために、この振り分けサーバが介在したときの遅延時間を計る必要がある。

6.2 展望

本システムでは現在のDNSのシステムを何ら変えずに、ホスト名の空間の中にホームページ名の空間を構築している。このことはホームページ名だけではなく他の名前空間もDNSの名前空間の中に構築できることを示している。応用例としては、メールアドレスの名前をつくり永年利用可能なメールアドレスを実現することが考えられる。

第 7 章

まとめ

本論文では、WWW が広く利用されるにつれ WWW の利用に多様性が生まれ、従来の WWW では対応できない問題が生じてきた。その問題はホームページを移動させたときに張られていたリンクが無効になる問題である。本論文では DNS を用いた方法でその問題を解決した。DNS を用いた方法では、各ホームページにインターネット上で一意に識別できるホームページ名を与え、ホームページ名と URL を対応させた。ホームページ名を既存の WWW のシステムで利用するため、利用者は URL を指定するとき URL の中のホスト名の代わりにホームページ名を指定する。この様な URL を本論文では永年利用可能な URL と名付けた。永年利用可能な URL を用いれば資源が移動しても同じ永年利用可能な URL で資源を参照することができることを示し、実験環境下で確かめた。最後に、考察を行ない今後の課題を述べた。

謝辞

本研究を進めるにあたり、数々の指導をいただいた大野浩之講師に感謝します。特に、研究の方向性を決める段階では時間を掛けていただき、ありがとうございました。大野研究室の皆様には研究生活をさまざまな面から支えていただきました。感謝します。そして辛い時期、相談にのってくださった国立がんセンター研究所がん情報研究部がん発生情報研究室の中川晋一先生に御礼を申し上げます。

参考文献

[1] Paul Albitz and Cricket Liu. DNS & BIND. 株式会社アスキー, 8月 1995年. 監訳 浅羽 登志也 / 上水流 由香.

[2] T. Berners-Lee and D. Connolly. Hypertext markup language - 2.0. *RFC1866*, November 1995.

[3] T. Berners-Lee, L. Masinter, and M. McCahill. Uniform resource locators (url). *RFC1738*, December 1994.

[4] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, and T. Berners-Lee. Hypertext transfer protocol – http/1.1. *RFC2068*, January 1997.

[5] NETCRAFT. Web server survey. *http://www.netcraft.com/Survey/Reports/*.

[6] Apache project. Apache http server. *http://www.apache.org/*.

[7] 中川格. World wide web 情報空間の特徴の分析と把握. 図書館情報大学 修士論文, 3月 1997年.