

Internet における情報資源の利用と提供の実際

阪口 哲男

図書館情報大学

〒305 茨城県つくば市春日1-2

Tel: 0298-52-0511, Fax: 0298-52-4326, E-mail: saka@ulis.ac.jp

概要

Internet では様々な方法で情報の提供が行われている。本稿では WAIS と WWW を中心に、現時点で実際に用いられている Internet 上の情報提供のシステムについて述べる。

Practice of Using and Providing Information Resources on the Internet

Tetsuo Sakaguchi

University of Library and Information Science

1-2, Kasuga, Tsukuba, Ibaraki, 305, Japan

Phone: + 81-298-52-0511, Fax: + 81-298-52-4326, E-mail: saka@ulis.ac.jp

Abstract

There are various methods for providing information on the Internet. This paper describes WAIS, WWW, and such information providing systems which are currently used on the Internet.

Keywords

the Internet, distributed system, WAIS, full-text search, information retrieval, ANSI Z39.50, WWW, NCSA Mosaic, hypertext, SGML, HTML, HTTP, URI, URL

1. はじめに

Internet では誰でも自由に情報を提供することができる。情報提供のために用いられるプロトコルは TCP/IP に基づいており、現在までに多くの標準的なものが定義、利用されている。情報共有のためには、プロトコルだけではなく提供されるデータのフォーマットも標準化されていなければならない。旧来、標準的に用いられていたものは書式などの情報を一切持たないプレーンテキストなどが中心であったが、近年では画像、音声、動画などを含むハイパテキストなどのフォーマットも標準化されている。

本稿では、Internet 上で用いられている情報を提供するシステムの中で、主に WAIS と WWW を取り上げる。WAIS は分散型の全文検索システムであり、WWW は分散型のハイパテキストシステムである。これらの概要と実際にツールを用いた情報の提供について述べる。また、WAIS や WWW 以外のシステムについても簡単に触れ、最後に、Internet を通じた情報提供の今後について論じる。

2. WAIS (Wide Area Information Servers)

2.1. WAIS の概要

WAIS は Brewster Kahle が考案し、Dow Jones、Thinking Machines、Apple Computer と KPMC Peat Marwick の共同プロジェクトとして改良が行われ、現在では商用化も検討されている。WAIS は Internet に接続された複数の計算機上に、様々な全文データベースを格納している。WAIS の利用者は、Internet に接続された計算機であればどこからでも検索を行い、その結果得られる論文や記事などを読むことができる。

WAIS は Internet に接続された複数のサーバとクライアントから構成される。サーバは索引付けされたテキストのデータベースを備えており、クライアントからの検索要求に応じてデータベースを検索し、結果を返す。また、データベースに格納されているテキストを要求に応じて転送する。クライアントは利用者から指定された検索条件を複数のサーバに分配し、各サーバから返される検索結果を統合して利用者に提示する。サーバとクライアントの間の通信は ANSI Z39.50 に準拠したプロトコルに基づいて行われる。また、検索結果から利用者が選択したテキストの内容を表示することも可能である。データベースには、付与キーワードなどで索引付けを行ってあれば、テキスト以外にも画像を格納することができる。数多くのサーバから利用者の目的に合ったデータベースを備えたサーバを探すために、特別なデータベース `directory-of-servers` が準備されている。利用者は最初に `directory-of-servers` を検索し、引き続き目的に応じたサーバで検索を行って目的のテキストや画像を得る。

ここでは、`freeWAIS` という最も一般的に用いられている UNIX ワークステーション上の WAIS 処理系を例にしてシステムの概要について述べる。なお、`freeWAIS` は日本語に対応していないが、最近では日本語に対応したパッケージソフトなども存在している。

2.2. データベースとサーバシステム

WAIS のデータベースに格納するデータは、索引付けされていればテキストでも画像でも良い。索引付けには `waisindex` というコマンドを用いる。`waisindex` は標準的に対応している形式で作成された (複数の) ファイルを読み込んで索引に用いる語を抽出し、検索に用いる索引ファイルを作成する。現在のシステムで対応している形式は約 50種類ある。こうして索引付けされたファイルを WAIS ではデータベースと呼ぶ。例えば、もっとも単純に全文データベースを構築する際は、書式情報を含まないプレーンテキストを格納したファイルのあるディレクトリに集め、そのディ

レクトリを指定して `waisindex` を実行するのみである。

上記のように構築されたデータベースを検索するために、**WAIS** では二通りの方法が準備されている。一方は、利用者が使用するアプリケーションによって直接検索する方法であり、もう一方はネットワークを通じて送られる要求に応じて検索を行うサーバシステムを用いる方法である。**Internet** 上で利用する場合は、後者のサーバシステムを用いる方法を用いる。**freeWAIS** ではサーバシステムを起動する `waisserver` というコマンドが準備されている。`waisserver` を実行する際、データベースを登録しているディレクトリを指定することにより、そのデータベースを検索するサーバシステムが起動される。サーバシステムはネットワークを通じて送られる要求に応じて、データベースの検索とテキスト (または画像) データの転送を行う。

サーバは **WAIS** のプロトコルを採用していれば、その実現はどんなものでも良い。このため、現在では **freeWAIS** で提供される `waisserver` を用いずに、`waisserver` では対応していない機能を提供するサーバも稼働している。

2.3. クライアントシステム

クライアントシステムは利用者と対話して、検索要求をサーバシステムに送り、返された結果を利用者に提示する。**freeWAIS** における標準クライアントシステムは **X-Window** 上で用いられる `xwais` というものである。`xwais` でシェークスピアの全文データベースを検索した例を図1に示す。クライアントシステムはデータベースを検索する際、データベースが設置されているサーバシステムのネットワーク的な所在情報が必要となる。**WAIS** ではその所在情報をデータベース毎にデータベースの概略説明等とともに記述したものをソースと呼び、検索を行うときにクライアントシステムに与えるようになっている。

ソースにはデータベースの名称、版、サーバのホスト名またはホストアドレス、ポート番号、サーバの使用料金、データベースの保守者のメールアドレス、概略説明などが記述される。ポート番号はある計算機 (ホスト) が複数のサービスをネットワーク上で提供する際に、そのサービスを識別するための番号である。クライアントシステムは、このソースに記述された情報を基に、サーバのホストと接続し、通信を行う。従って、**WAIS** を用いて検索を行う際は、このソースを何らかの手段で入手しておく必要がある。

2.4. `directory-of-servers`

世界各地に数多く点在しているすべてのサーバについてソースを入手しておくのは不可能であり、また仮に入手できたとしても利用者がその中から自分の目的に合ったサーバを探すのは困難である。そこで、利用者の目的に合ったサーバを探すために、`directory-of-servers` と呼ばれるデータベースを設け、利用者はまずこの `directory-of-servers` を検索することにより、所望のデータが得られると考えられるサーバのソースを入手することができる。一度入手したソースは利用者の使用した計算機上のファイルに格納しておくことができるので、次に検索するときは `directory-of-servers` を検索するステップを略して直接目的のサーバにアクセスすることができる。

新たに **WAIS** のサーバを設置する際、`directory-of-servers` に登録を依頼する必要がある。本稿の執筆時点では **Thinking Machines** の `quake.think.com` というホスト上に `directory-of-servers` が設置されており、新たなサーバ上のデータベースのソースを電子メールで送って登録を依頼する。実際には、`waisindex` コマンドに "`-register`" という引数を与えると、索引付けと同時に生成したソースを用いて登録依頼を自動的に行うようになっている。

なお、`directory-of-servers` は必ずしも `quake.think.com` 上のものを使う必要はない。ソースを集めることができれば独自の `directory-of-servers` を設置することができる。実際に、**Internet** に

は数多くの `directory-of-servers` が存在するが、混乱を避けるために `quake.think.com` 上のもののみが名称として “`directory-of-servers`” を使用し、他のものは若干異なる名称を使用している。

3. WWW (World Wide Web)

3.1. WWW の概要

WWW は CERN (European Particle Physics Laboratory) で開発された、Internet をまたいだリンクを可能とした分散型のハイパテキストシステムである。WWW では Internet に接続された複数の計算機上に、相互にリンクされたテキストや画像、音声、動画などが格納される。

WWW では、これらのテキストなどを格納されている計算機を意識せずにリンクをたどって読むことができる。

WWW も WAIS と同様に Internet に接続された複数のサーバとクライアントから構成される。サーバにはリンク付けされたテキストや画像、音声などが格納される。クライアントは起動時に指定されたサーバより最初のテキストを取り寄せ、利用者に提示する。提示する際、そのテキスト中に埋め込まれたタグに従ってリンク情報を利用者に示す。利用者はテキスト中の興味を持ったリンクを指定してリンク先であるテキスト等を取り寄せることができる。

テキスト中にはリンクの他に文字の大きさや字体の指定と、画像や音声の埋め込みができる。これらの指定を行うタグの形式は、SGML に準拠して定義された HTML (Hypertext Markup Language) に従っている。HTML で記述されたテキストをサーバとクライアントの間で転送するために、HTTP (Hypertext Transfer Protocol) を用いている。リンクを表すタグには、この HTTP で接続するサーバのホスト名と、そのホスト上でユニークなテキスト (または画像や音声) の名前が URL (Uniform Resource Locators) と呼ばれる標準形式で埋め込まれている。リンクをたどる際は、このホスト名で示されるサーバと接続し、指定されたテキストを転送する。なお、タグには転送に使用するプロトコルも同時に指定できるので、HTTP だけではなく FTP (File Transfer Protocol) などの他のプロトコルを用いたサーバも利用できる。

この WWW の処理系には様々なものがあるが、ここではその機能の豊富さからよく用いられる NCSA Mosaic を例にとる。

3.2. HTML テキストと WWW サーバ

WWW ではサーバ上に格納するテキストを HTML に従って記述しておく必要がある。HTML は SGML に準拠しているが、HTML で記述されたどのテキストも同じ DTD (Document Type Definition) となるので、WWW で実際に用いられるテキストでは DTD が略されている。HTML で書かれたテキストの例を以下に示す。

```
<HTML>
  <TITLE>A sample HTML instance</TITLE>
  <H1>An Example of Structure</H1>
  Here's a typical paragraph.<P>
  <UL>
    <LI>
      Item one has an
      <A HREF="http://ulispsn.ulis.ac.jp:8001/">
        anchor
      </A>
```

```

<LI>
  Here's item two.
</UL>
<H2>An Example of Embedded Image</H2>
Here's a bitmap image.
<IMG SRC="escherknot.xbm">
</HTML>

```

このテキストを Mosaic で見ると図2 のようになる。HTML では “<” と “>” で囲まれたものがマークとなっている。HTML の文書タイプ名は “HTML” であるので、テキストは <HTML> で始まり、 </HTML> で終る。例ではタイトル (<TITLE>, </TITLE>), ヘッダ (<H1>, </H1>, <H2>, </H2>), リスト (, ,), パラグラフ区切り (<P>) などが文書構造を表すために使用されている。また、リンクをはるためにアンカ (<A>,) が使われており、この HREF= で指定された URL へとつながっている。 はテキスト中に画像を挿入するもので、SRC= で画像ファイルを指定する。

前節で述べたように、WWW のサーバは HTTP プロトコルを使用するものが基本であり、その他のプロトコルを使用したサーバも WWW サーバとして利用することができる。WWW サーバのプログラムそのものは Mosaic には付属しておらず、CERN や NCSA から提供される WWW のサーバプログラムが必要となる。WWW サーバは HTML テキストをクライアントからの要求に従って転送することが主な機能であり、その HTML テキストは WWW サーバ上にファイルとして格納しておく必要がある。また、実際に提供される WWW サーバのプログラムには検索機能や他のシステムとの連携機能なども含まれている。

HTTP 以外のプロトコルを用いるサーバの場合は、プロトコルによってサーバ上に格納するデータの形式などが異なる。例えば FTP サーバの場合は、テキストや画像などが単なるファイルとして格納してある。また、各種のゲートウェイを行うサーバも準備されている。例えば WAIS とのゲートウェイを行うサーバでは、WWW から WAIS の検索を行うことを可能としている。

3.3. WWW クライアント: NCSA Mosaic

WWW クライアントには CERN が提供する文字端末向けのシステムもあるが、扱い易さと機能の豊富さの観点から NCSA Mosaic がよく用いられる。Mosaic は NCSA (National Center for Supercomputing Applications) で開発されたツールで、WWW だけではなく、様々な Internet のサービスにもアクセスできるように設計されている。Mosaic はマルチウィンドウ環境を前提としており、UNIX の X-Window システム上で動作するもの他に Apple Macintosh や MS-Windows で動作するものも提供されている。

Mosaic では、HTML テキストに埋め込まれたタグに従って文字の大きさや字体を変えて表示したり、リンク付けされている部分はアンダーラインや縁取りが施され、一目でわかるようになっている。そのアンダーラインのついた語や縁取りされた部分をマウスでクリックすることにより、リンクをたどることができる。また過去の履歴を保持しており、一度たどったことがあるリンクについては、そのアンダーラインが異なる色の破線となる。Mosaic を起動してからのたどったリンクについてはたどった順序を逆に戻ることができるほか、履歴ウィンドウを表示して直接それまでに読んだテキストの一つを指定して戻ることができる。そのほか、テキスト表示ウィンドウの複数化や利用者個人のコメントをつけ加える機能も有している。

Mosaic を起動した際には、ホームページと呼ばれるテキストが表示される。ホームページには WWW やその他のサービスへのリンク情報が含まれており、このリンクをたどることで各種のサ

サービスの利用を開始する。標準のホームページは NCSA で準備されているが、利用者が独自のホームページを持つことも可能である。サービスの利用は基本的にマウスによるクリックで利用者が各種の指示を行うが、WAIS へのアクセス時など文字列を指定する必要がある際は、テキスト中に文字列指定用のフィールドが現れ、利用者が文字列を入力することができるようになる。

なお、Mosaic は日本語に対応したものもある。

4. その他のシステム

WAIS や WWW 以外にも表1 のように数多くの情報を提供するためのシステムがある。

表1 その他の情報提供システム

名称	概要
anonymous ftp	フリーソフトなどのファイルの転送サービス
archie	anonymous ftp で提供されるファイル名の検索サービス
finger	特定のホストにおける利用者情報の提供
gopher	Internet 上の情報資源をたどることができるメニューシステム
telnet	計算機の遠隔利用システム (多くの目録検索システムなどが telnet によって提供されている)
whois	Internet 上の個人/組織情報サービス

anonymous ftp は単独ではファイルのありかを探すのが困難なので、archie によるファイル名の検索機能の提供が行われるようになった。finger はある計算機上の利用者の情報を得ることができるが、悪用を避けるため管理者によって使用できないようにしている場合も多い。whois は一般にネットワークやシステムの管理者などの連絡先を提供し、専らネットワーク運用に役立てる目的で利用されている。telnet によって提供されるサービスには、従来より運用されている情報検索サービスなどをユーザ登録なしで利用できるようにしたものが多い。gopher は Internet 上に分散した階層型のメニューシステムであり、メニューの各項目が他のホストのメニューやテキスト、画像、telnet を利用したサービス、anonymous ftp などに対応している。

5. 今後の展望

以前は様々なサービスが独立して提供されていたため、利用者はその形態に合わせた使い方を覚える必要があった上、どのような情報資源があるか探すことも困難であった。WWW や gopher によってそのような問題は解決され、様々なサービスをメニューやリンクをたどって容易に情報資源を利用することができるようになった。特に WWW では、HTML というハイパテキストの記述法の一標準を定義した上、URL や URI (Universal Resource Identifiers) という情報資源の識別法を考案している。

URI は既に RFC 文書となっており、URL は現在 HTML や HTTP と共に Internet の RFC 文書として標準化が進められつつある。URI は Internet 上の文書やサービスなどのオブジェクトを識別するための記法を定義している。URL は URI に基づいて既存のプロトコルを用いて、情報資源をアクセスするためのアドレスを表現するものである。URL は Internet 上の情報資源が設置されているホストとプロトコルを特定するため、様々な形態のサービスを統一的に表現することができる。現時点で URL は既に利用されているサービスを対象にしているが、新たな形

態のサービスが提供されるようになっても拡張可能なようになっている。

今後も Internet 上で新たなサービスが開発されると考えられる。その際、既存のサービスと互換性はなくても URI に基づいて情報資源を特定できるようにすれば、従来からのシステムとの統合が可能である。しかしながら、URI による参照や WAIS の `directory-of-servers` のような単純な検索のみでは、急速に増加しつつある情報資源の中から目的に合うものを捜し出すのは困難になりつつある。今後は情報資源の利用の際に、利用者の知的なサポートを可能とするシステムが必要になると考えられる。

6. おわりに

本稿では WAIS、WWW を中心に述べてきた。現在は WWW が Mosaic というユーザフレンドリなツールと組合わさって、急速に普及しつつある。

WWW サーバを設置するには、NCSA や CERN の提供するプログラムを導入する必要がある。また、NCSA、CERN 共にかなり高度な機能を備えているため、導入の際には UNIX ワークステーションや C 言語プログラミングについての知識などがある程度要求される。そのため、現時点では (特に日本では) WWW サーバを設置しているのは、まだまだ計算機寄りの分野に偏っていると思われる。そのような観点から著者は現在 WWW サーバの基本的な機能を含めた新たな Internet 上の情報提供システムについての検討を行っている。

参考文献

(Internet 上の情報提供システム全般について)

- [1] Krol, Ed. *The Whole Internet Catalog & User's Guide*. O'Reilly & Associates. ISBN 1-56592-025-2. 396p. 1992. (邦訳: 村井純監訳. インターネットユーザズガイド. オーム社. ISBN 4-900718-00-9. 568p. 1994)

(WAIS について)

- [2] Pierre, M. St. ほか. *WAIS over Z39.50-1998*. RFC 1625. 7p. 1994.

(WWW について)

- [3] Berners-Lee, T. *Hypertext Transfer Protocol*. Internet Draft. 25p. 1993.
<ftp://info.cern.ch/pub/www/doc/http-spec.txt>
- [4] Berners-Lee, T.; Connolly, Daniel. *Hypertext Markup Language*. Internet Draft. 36p. 1993. <ftp://info.cern.ch/pub/www/doc/html-spec.txt>
- [5] Berners-Lee, T. *Universal Resource Identifiers in WWW*. RFC 1630. 28p 1994.
- [6] Berners-Lee, T. *Uniform Resource Locators*. Internet Draft. 22p. 1993.
<ftp://info.cern.ch/pub/www/doc/url-spec.txt.Z>

Appendix: フリーソフトウェアリスト

(WAIS, WWW に関するフリーソフトウェアを探す際のファイル名を示す。多くの anonymous ftp サーバによって提供されているので、archie によって最寄りのサーバを探すことができる。)

WAIS

freeWAIS-0.3.tar.gz (サーバとクライアントの両者が提供される。なお、WAIS に関する詳しい文書も含まれている。)

WWW

Mosaic-2.4.tar.gz (WWW クライアント、UNIX ワークステーション用の NCSA Mosaic。欧米語のみに対応)

Mosaic-2.4-110n-0.patch.gz (Mosaic-2.4.tar.gz を日本語などに対応させるパッチ)
cern_httpd_3.0pre6.tar.gz (CERN が提供する WWW サーバプログラム)

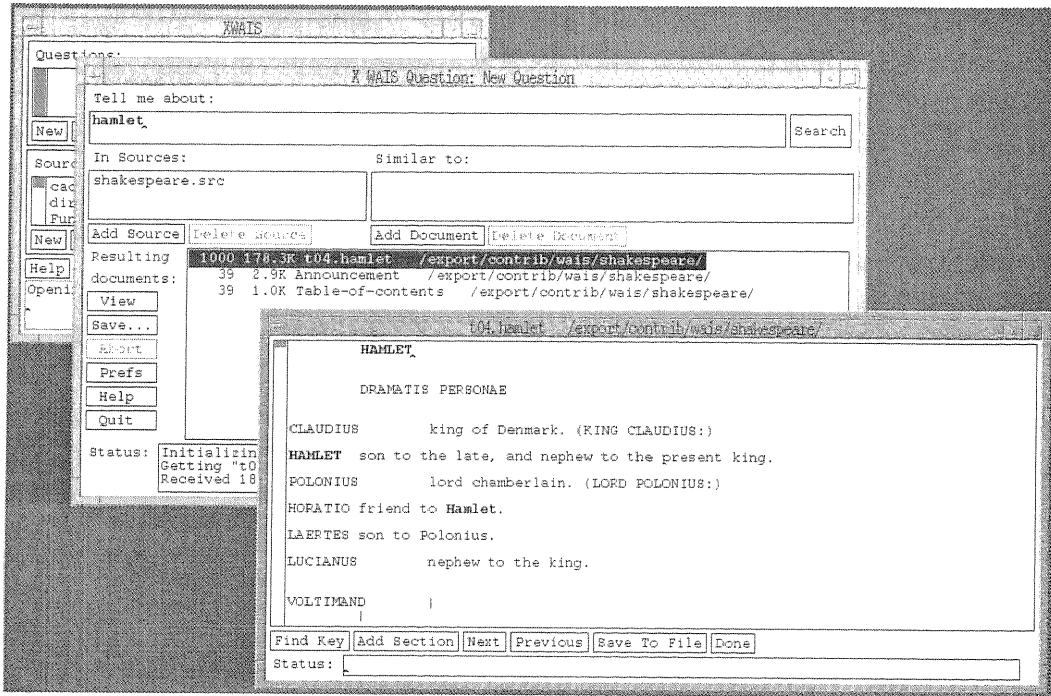


図1 xwais による検索画面例

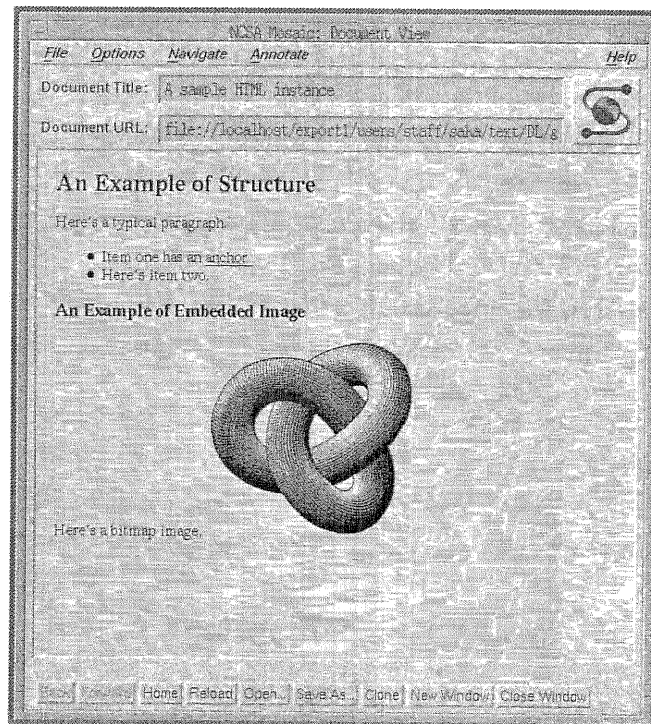


図2 Mosaic による HTML 表示例