

Frequently Asked Question



いまさら聞けない



いまだから聞きたい

このコーナーでは読者の皆さんのインターネットに関する疑問や質問にお答えします。「?」と感じたことはどのようなことでも構いませんので、下記のメールアドレスまでご質問ください。なお、ご質問へのメールでの回答はできませんのでご了承ください。

ご質問はこちらまで
im-faq@impress.co.jp

1 URLの仕組みとURIやURNとの関係

2 ウェブサーバーのログから得られる情報

今月のポイント



ウェブページをhttp://internet.impress.co.jp/のようにURLで表しますが、この仕組みはどうなっているのでしょうか。また、最近URIやURNという言葉も聞きます。何が違うのでしょうか。(富山県 M.Tさん)



「URL」は、「Uniform Resource Locator」の略で、リソースの「場所」とそこへのアクセス方法を表します。

URLの上位概念として、「URI」があります。URIは、「Uniform Resource Identifier」の略で、リソースを一意に特定する識別子です。URIの書式は、「(スキーム名)(スキームごとに決められた部分)」のように「:」で区切られた文字列になっていて、スキーム名はIANAという組織が管理しています。

上位概念であるURIに含まれるものの1つが、URLというわけです。URLの先頭要素であるhttpやftpなどは、すべてスキーム名になります。図1は、httpやftpで共通に利用できるURLの書式の一部です。

URIの下位概念には、URLのほかに「URN」があります。URNは「Uniform Resource Name」の略で、リソースの存在する場所に依存しない、普遍的な名前を意味します。URIの書式によるURNの

記述は、「urn(名前空間ID)(空間内での識別)」となります。たとえば、書籍コードを表すISBNはURNであり、「urn:isbn:4-8443-1671-0」と表せます。URIと、URL/URNの関係を図で表すと図2のようになります。

(本誌アドバイザー井上尚司)

図2 URIとURL/URNの関係

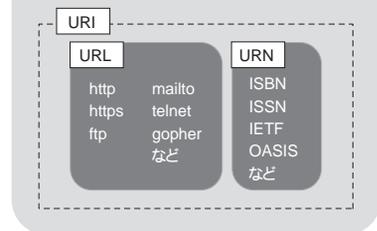
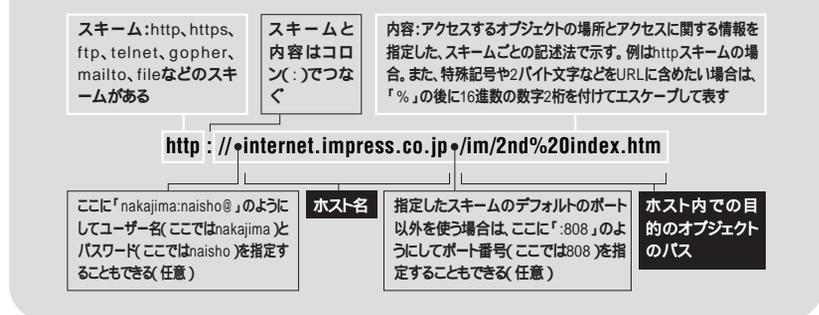


図1 URLの構造(httpの場合の例)



URLもURNもURIのサブセット
URLは場所を、URNは名前を示す



Q

上司から、ウェブサイトのアクセスログはちゃんと確認して情報を得ておくように言われたのですが、何をどうすればいいのでしょうか。(東京都 N.Kさん)

A

ウェブサーバーは、ブラウザなどのクライアントからの要求を処理するごとに、その情報をアクセスログに記録しています。たとえば図3は、Apacheウェブサーバーが取っているログのうちの1行を示したものです。ここには、要求したクライアントのIPアドレス、日時、要求の内容であるリクエストヘッダー、終了状態、送信バイト数などが記録されています。

ログの書式は、ウェブサーバーの種類ごとに異なりますし、サーバーの設定によっても変更することができます。Apacheウェブサーバーの場合、環境設定ファイルであるhttpd.confでログの項目を指定します。この設定を変えることで、いろいろなログデータを取ることが可能となります。たとえば使われたブラウザの種類や、どのリンクをたどってきたのかなどの項目も取れます。

このログを見れば、ウェブサーバーへのアクセスの状況を知ることができるわけですが、もちろんこのログだけでは利用者の個人情報を知ることができません。ただ、このログファイルと、クライアント側が利用しているISPの通信ログなどを照らし合わせれば個人の特特定まで可能となります。

ウェブサーバーは、基本的にクライアントからのリクエスト1つに対して1行のログを生成します。たとえばウェブページの中にgifやjpegの画像があると、それぞれが1つのリクエストとなり、ログに記録されます。つまり、ログファイルは非常に大きいものとなり、通常はそれを直接見て何かを判断するというのは無理があります。

アクセスログは情報の宝庫 ツールをうまく使って解析を

そこで活躍するのが、ウェブのログ解析ツールです。

ログ解析ツールには、簡単な集計だけをするものから本格的なレポートを図表ともに出すものまで、また、フリーソフトから商用のソフトまで、多々あります。当然のことですが、サーバーのログの重要性をそれぞれのサイトごとに判断して、ツールを選べばよいでしょう。

図4に示したのは、商用のログ解析ツールである「サイトトラッカー」^{URL} というソフトからの出力の一例です。このソフト

では、100種類以上のサマリー(要約)やレポートの出力が可能で、たとえば利用者のサイト滞在時間や、どのページをどれくらい時間をかけて見ているかなど、さまざまな情報を得ることができます。図5のようにグラフで表示することもできます。

ウェブのログ解析は必須の作業なので、まずはフリーソフトから試してみて、必要に応じて商用ソフトやサービスを選択していくのがいいでしょう。

(本誌アドバイザー井上尚司)

^{URL} <http://www.sitetracker.jp/>

図3 アクセスログの1行の例

クライアントの IPアドレス	日時	リクエストヘッダー	終了 状態	送信 バイト数
192.168.1.54	- - [10/Jan/2003:09:51:24 +0900]	"GET /index.htm HTTP/1.0"	200	6774

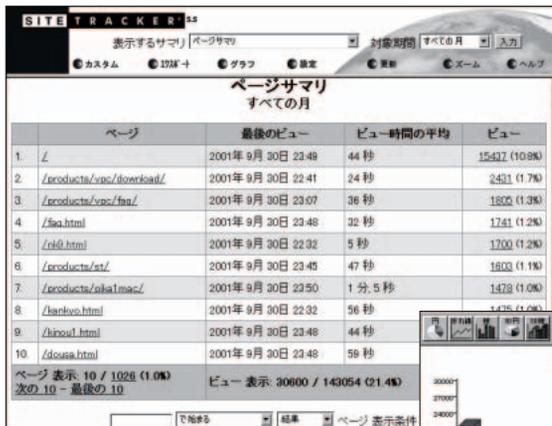
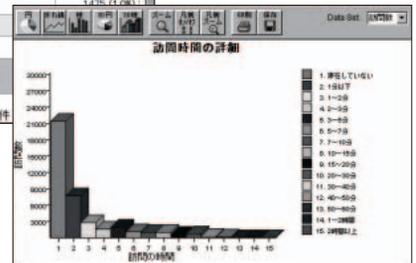


図4 サイトトラッカーのレポート画面例

図5 サイトトラッカーのグラフ画面例





[インターネットマガジン バックナンバーアーカイブ] ご利用上の注意

このPDFファイルは、株式会社インプレスR&D(株式会社インプレスから分割)が1994年～2006年まで発行した月刊誌『インターネットマガジン』の誌面をPDF化し、「インターネットマガジン バックナンバーアーカイブ」として以下のウェブサイト「All-in-One INTERNET magazine 2.0」で公開しているものです。

<http://i.impressRD.jp/bn>

このファイルをご利用いただくにあたり、下記の注意事項を必ずお読みください。

- 記載されている内容(技術解説、URL、団体・企業名、商品名、価格、プレゼント募集、アンケートなど)は発行当時のものです。
- 収録されている内容は著作権法上の保護を受けています。著作権はそれぞれの記事の著作者(執筆者、写真の撮影者、イラストの作成者、編集部など)が保持しています。
- 著作者から許諾が得られなかった著作物は収録されていない場合があります。
- このファイルやその内容を改変したり、商用を目的として再利用することはできません。あくまで個人や企業の非商用利用での閲覧、複製、送信に限られます。
- 収録されている内容を何らかの媒体に引用としてご利用する際は、出典として媒体名および月号、該当ページ番号、発行元(株式会社インプレス R&D)、コピーライトなどの情報をご明記ください。
- オリジナルの雑誌の発行時点では、株式会社インプレス R&D(当時は株式会社インプレス)と著作権者は内容が正確なものであるように最大限に努めましたが、すべての情報が完全に正確であることは保証できません。このファイルの内容に起因する直接のおよび間接的な損害に対して、一切の責任を負いません。お客様個人の責任においてご利用ください。

このファイルに関するお問い合わせ先

株式会社インプレスR&D

All-in-One INTERNET magazine 編集部

im-info@impress.co.jp