

WEBページのアクセスに関する分析とアクセスの質の向上に関する研究

鈴木浩之*¹ 内藤丈資*² 秋山 稔*² 菊池和尚*³ 大野哲晴*⁴ 匂坂 剛*⁵

Study on WEB Access Analysis and Improvement of Quality of Access

SUZUKI Hiroyuki*¹, NAITO Takeshi*², AKIYAMA Minoru*²,
KIKUCHI Kazuhisa*³, OHNO Tetsuharu*⁴, SAGISAKA Takeshi*⁵

抄録

インターネットの利用者は増加しているが、中小企業における活用は充分とは言えない状況にある。また公的な試験研究機関自身も情報発信の手段として Web ページを使用しているが、改善の余地がある。そこで、当センターの情報発信能力の向上、及び技術支援の高度化を目的として、当センターの Web ページについてアクセスログを取得、分析できる環境を構築し、分析結果に基づいて Web ページの改善を行った。改善後もアクセスログの取得、分析を継続し、改善による効果が現れていることを確認した。

キーワード：インターネット、アクセス解析、ホームページ、Web ページ

1 はじめに

インターネットの利用者は着実に増加しており¹⁾、情報を発信する手段としても Web ページ（一般には「ホームページ」とも呼ばれる。）が安価な広告媒体として企業に活用されている^{1),2)}。特に最初の情報アクセス先としてインターネットが圧倒的に多くを占めるとする調査もあり³⁾、Web ページによる情報発信はその重要性を増している。

しかし Web ページの開設状況に関する調査では、企業規模によって大きな開きがあると推定され²⁾、中小企業ではインターネットを十分に活用できていないと考えられる。公的な試験研究機関

でも Web 利用に関する研究開発が行われているが^{4),5)}、先進的な取組が多く中小企業による活用の促進に資する研究報告はあまり見られない。

一方 Web ページは、世界中に存在する膨大な情報の中からユーザーが情報を検索、取捨選択しアクセスするという点や、アクセスに関する一定の記録がアクセスログとして残る点など、従来の広告媒体と異なる特徴を持つ。アクセスログを活用した Web ページの分析は、アクセス解析と呼ばれ広く行われている。しかし開設されている Web ページの数に比較すれば十分に活用されているとは言えない。このことは当センターの Web ページでも同様であった。

このような状況を踏まえて、当センターの Web ページに対して、アクセスログを取得して統計的な分析を行える環境を構築し、実際に分析を行った。さらに分析に基づく Web ページの改良を行い、その効果を検証した。またこれらの分析、改良を通して Web ページの効果的な運用に関する

*¹ 電子情報技術部

*² 技術支援室

*³ 北部研究所 技術支援交流室

*⁴ 企画室

*⁵ 消防防災課

技術の蓄積を図った。

2 アクセス解析の概要

2.1 アクセスログの概要

アクセス解析の第一段階としてアクセスログを取得する必要がある。代表的な取得方法とその特徴⁶⁾を表1に示す。どの方法を利用するかは、各方法の特徴や Web サーバの環境などを考慮して決める必要がある。

表1 代表的なアクセスログの取得方法

| 方法 | 主な特徴 |
|-------------------|---|
| Web サーバのログを直接利用する | メリット ・画像ファイルや PDF ファイルなど全てのリクエストについてログを取得できる。 デメリット ・データ量が大きくなりやすい。 ・Web サーバの管理権限がないとログを得られない場合がある。 |
| CGI を利用する | メリット ・Web サーバとは別のサーバでもログを取得できる。 デメリット ・HTML (XHTML) 以外のファイルについてのログを取得できない。 |
| パケットキャプチャリングを行う | ・メリット 複数の Web サーバを使って負荷分散を行っている場合など大規模な環境で利用しやすい。 |

2.2 CGIを利用したアクセスログの概要

CGIを利用してアクセスログを取得する場合、アクセス時に CGI を起動させる必要がある。その主な方法を表2に示す。それぞれの特徴と Web サーバの環境を考慮して、どの方法を利用するかを決める必要がある。

2.3 リファラの概要

Web ページの大きな特徴にハイパーリンクがある。HTML 文書中に他の文書や画像などの所在情報を埋め込んで情報同士を関連付けるシステムで、一般には単にリンクと呼ばれることも多い。リファラとはこのリンクによる参照元を示す情報である。すなわち、HTML 中のリンクによって Web ページを移動した場合の直前の閲覧ページを示す情報である。

表2 代表的なCGIの起動方法

| 方法 | 主な特徴 |
|---------------------------------|---|
| SSI を利用する | メリット ・画像ファイル等をリクエストしないアクセス (検索ロボットなど) でもログを取得できる。 デメリット ・Web サーバの管理権限がないと利用できない場合も多い。 |
| HTML 中の IMG タグを利用する | メリット ・Web サーバの管理権限がなくとも利用できる。 デメリット ・取得できる情報量が少ない。 |
| JavaScript を使用して IMG タグを動的に利用する | メリット ・Web サーバの管理権限がなくとも利用できる。 ・取得できる情報量が多い。 デメリット ・ブラウザが JavaScript の実行を許可していないと利用できない。 |

近年、検索サイトを利用した検索結果から Web ページにアクセスする利用者が増えている。何らかのキーワードによる検索結果からある Web ページにアクセスした場合、リファラには検索サイトの URL だけでなく検索に用いたキーワードの情報が付随して記録される。これはアクセス解析において重要な情報である。

3 アクセスログの取得方法及び分析方法

3.1 アクセスログの取得方法

アクセスログの取得方法としては、比較的小規模な環境において使いやすい点や Web サーバの管理権限がない場合でも利用できる点を考慮して CGI による方法を採用した。

無料で使用できる CGI もインターネット上に公開されているが、使用に制約があったり、希望するアクセス情報が取得できなかったりするなどの欠点がある。また日本語文字コードの扱いが充分でない場合もある。そこで必要な CGI は、Perl を用いて自前で作成した。

CGI の起動方法としては、SSI と JavaScript を併用した。ただしブラウザが JavaScript の実行を許可していない場合には、SSI と IMG タグが併用されるようにした。

北部研究所の Web ページでは平成 18 年 5 月から、本所の Web ページでは平成 18 年 11 月からアクセスログの取得を開始した。

3.2 アクセスログの分析方法

取得したアクセスログを分析する際、初めに検索キーワードの抽出を行った。検索キーワードに日本語が使われた場合、リファラ中には URL エンコードと呼ばれる方法でデータが変換されている。これをデコードし日本語として抽出するために、Perl を用いてデコードツールを作成した。

検索キーワードの抽出が済んだログデータを、本所のデータは半月ごとに、北部研究所のデータは一月ごとに集計して分析を行った。必要に応じて種々の集計、分析を行ったが、特に当センター Web ページへのアクセス経路については、表 3 のように分類して集計を行った。

表 3 Webページへのアクセス経路の分類

| 分類名 | 分類の定義 |
|----------|--|
| ブックマーク等 | リファラが記録されていないアクセス |
| 県関連サイト | 埼玉県の Web ページからのアクセス、及び本所は北部から、北部は本所からのアクセス |
| その他のサイト | 県関連サイト以外からのリンクによるアクセス（検索結果によるアクセスを除く） |
| センター名等検索 | 当センターの名称またはそれに近い名称による検索結果としてのアクセス |
| その他の検索 | 上記とは異なるキーワードによる検索結果としてのアクセス |

4 分析結果と考察、及び改良と検証

4.1 分析結果と考察

本所及び北部研究所の Web ページについて、11 月末に分析を行った結果を図 1 に示す。両者のアクセス数には大きな差があったので、構成比で示した。両者の間には、アクセス経路に大きな傾向の違いが見られた。

本所の Web ページは主に当センターの業務に関する情報を網羅的に掲載している。そのため既に当センターを利用したことのある人や当センターを知っている人が、利用に関する情報を求めて

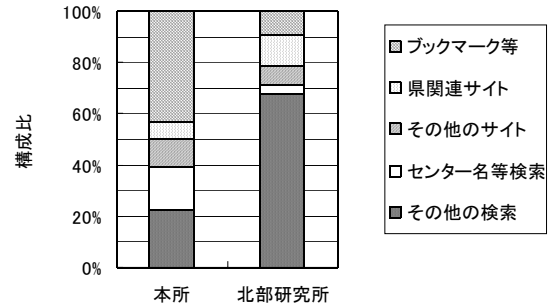


図 1 訪問経路別の構成比

アクセスすることが多いと考えらる。

これに対し北部研究所の Web ページは独自の情報に特化しており、例えば繊維関係の情報を多く紹介しているページが複数ある。その結果、繊維に関する用語を中心として多様なキーワードによる検索結果からのアクセスが多いことが分かった。

次にメールマガジンからのアクセスを捉えて、集計した結果を表 4 に示す。当センターが発行しているメールマガジンの購読者数は約 1000 件であり、それに比べれば大きい数字ではない。しかし、[受信]→[開く]→[読む]→[クリック]という一連の動作をした結果のアクセスであり、高い興味を持ったアクセスと考えられる。当センターのメールマガジンが提供する情報の幅が広いことや、メールマガジンが多種発行されている状況などを考慮すると、このアクセス数は堅実な効果と考えられる。

表 4 メールマガジンからのアクセス

| 発行日 | 記事番号 | アクセス数 | 発行日 | 記事番号 | アクセス数 |
|-------|------|-------|-------|------|-------|
| 11/ 1 | 1 | 8 | 12/ 1 | 8 | 7 |
| 11/15 | 1 | 4 | 1/ 5 | 1 | 10 |
| 11/15 | 2 | 5 | 1/ 5 | 2 | 20 |
| 12/ 1 | 1 | 9 | 2/ 1 | 1 | 5 |
| 12/ 1 | 7 | 18 | 2/15 | 1 | 12 |

表 5 は、メールマガジンのバックナンバーが、“その他の検索”によるアクセスの中で占めている割合を示している。比較のためトップページの数値も示した。バックナンバーのページ数に比べれば大きな数字ではない。しかしバックナンバーは、過去に発行したコンテンツを元にして掲載で

き、必要な作業量に対して情報の蓄積と検索によるアクセスの効果が高いと考えられる。

表5 メールマガジンのバックナンバーが「その他の検索」に占める割合 (本所のWebページ)

| ページ名 (ページ群) | 割合 |
|-----------------|-----|
| メールマガジンのバックナンバー | 15% |
| トップページ | 21% |

4.2 特筆的なアクセスログ

上記の集計のほか、外的な要因で急激にアクセス数が増加する事例があったので、幾つか挙げておく。図2は、アクセス数の変化を示したものである。

一つは、当センターが開発に関わった商品がテレビ番組中で紹介された時の例である。商品名又はそれに近い言葉での検索結果からのアクセスが、放映日に急激に増えた。

もう一例は、埼玉県民の日(11月14日)前後の例である。当センターでも県民の日にあわせてイベントを開催したが、イベント案内のページに「県民の日」をキーワードとする検索結果からのアクセスが、11/13、11/14に急激に増えた。

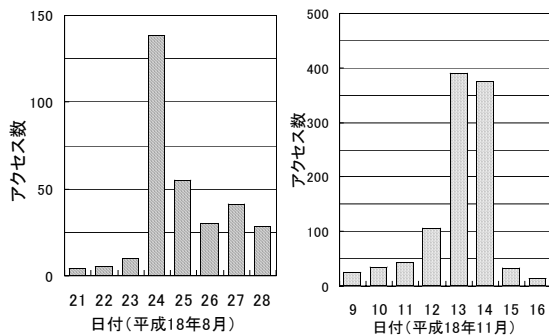


図2 急激なアクセスの増加の例 (左:テレビ放映、右:県民の日)

4.3 Webページの改良

分析結果を踏まえて Web ページの改良について検討し、当センターの関係部署と打合せのうえ、比較的着手しやすい改良について実際に変更を加えた。

4.3.1 抄録の追加

本所の Web ページに対して、検索結果からの

新規アクセス者を増やす方法を検討した。技術的課題を抱える人は、その技術に関する用語を用いて検索することが多いと推測できる。そこで当センターが扱う技術分野に関する言葉が多く含まれるように、研究報告一覧のページに研究の抄録を追加することにした。

4.3.2 PDFファイルのリンクの設定

研究報告の本文は PDF ファイルで掲載している。これらの PDF ファイルも検索の対象になるが、これらのファイルから通常の Web ページへのリンクを設定してなかった。そこでリンクを設定し、PDF ファイルから通常の Web ページへ容易にアクセスできるようにした。

4.3.3 北部研究所のWebページの改良

北部研究所の Web ページについては、検索結果からのアクセスをさらに増やすため、研究報告のうち特に重点をおいているものを PDF ファイルから HTML に書き直して掲載した。

4.4 効果の検証

4.4.1 抄録の追加

11月に行った分析と同様の分析を引き続き行い、その変化を見て効果を検証した。その結果を図3に示す。

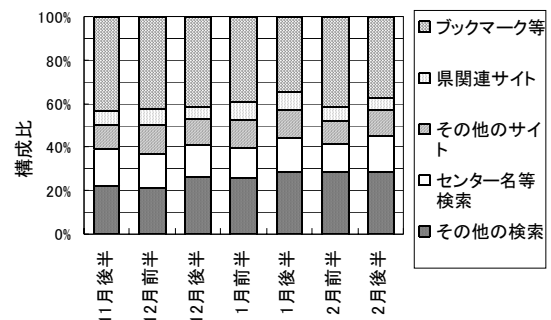


図3 訪問経路別構成比の変化

抄録の追加は 12 月中に行ったが、図3を見ると、1月以降“その他の検索”によるアクセスの比率が増加していることが分かる。アクセスログを詳しく調べたところ、抄録の追加がこの増加に大きく寄与していることが確認できた。

多種のキーワードによる検索結果からのアクセスが増えていることを検証するため、使われたキ

ワードの数を集計し、変化を調べた。その結果を図4に示す。1月以降、検索キーワードの種類が増加したことが分かる。

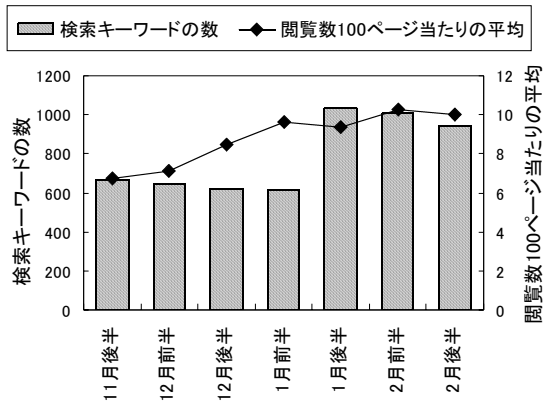


図4 検索キーワード数の変化

4.4.2 PDFファイルのリンクの設定

PDFファイルからのリンクの設定については、設定後の検証に十分な時間が得られなかったが、設定したリンクに伴うページ間の移動が2月後半で12件記録された。実際にアクセスログが記録されていることから、PDFファイルにリンクを設定することで、通常のHTMLと同様にページ間の移動を促す効果があると期待できる。

4.4.3 北部研究所のWebページの改良

北部研究所のWebページでは、顕著な差は見られなかった。この理由として、繊維関係情報のページが以前から大きな比率を占めていたことや、イベント開催などに関連するアクセス数の変動が大きく、単純に比較できないことなどが考えられた。

5 まとめ

当センターのWebページに対して、アクセスログを取得、分析できる環境を構築した。分析、Webページの改良、効果の検証を行い、技術の蓄積を行うとともに以下の知見を得た。

- (1) 研究報告一覧に抄録を加えることで、多種のキーワードで検索されるようになった。これによって新たなアクセスが増えることが期待できる。
- (2) メールマガジンは劇的な影響を与えるものではないが、堅実な反応がある。

(3) メールマガジンのバックナンバーは、検索による集客に一定の寄与をしている。既存のコンテンツを利用した効率的な運用方法と言える。メールマガジンに限らず、過去の情報であってもバックナンバーとして掲載することは効果があると考えられる。

(4) PDFファイルからのリンクを設定した方がよい。PDFファイルからHTML化するほうがさらに効果的と考えられるが、作業量を抑えて効果を上げる方法として、PDFファイルからのリンクを設定する方法にも効果が期待できる。

分析、改良、検証を行える環境が整ったので、今後も継続していきたい。特にWebページの全体的な運用は、当センター全体の運営に密接に関わることであり、データに基づいてWebページの運営に必要な情報を提供できるようにしたい。

また今回得られた知見等を活用して中小企業等の支援に役立てたい。

参考文献

- 1) 平成17年版 情報通信白書, 総務省
- 2) 平成18年度県内中小企業のインターネット利用実態調査結果(概要), WING21いばらき, (2007)
- 3) 近畿ブロック知事会 産学官連携研究会, 関西広域連携協議会 産業・科学技術部会: 企業が利用しやすいホームページを拡大するために, (2003)
- 4) 西田崇: Webアプリケーション間連携技術に関する研究(第1報), 長野県工業技術総合センター研究報告, (2005)
- 5) 瀧川明郎: LogManagementシステム開発・事業化—Webシステムにおける証跡ログ保存と監査支援—, (独)情報処理推進機構2005年度成果報告集
- 6) 石井研二: アクセス解析ログの教科書, 翔泳社, (2005)240