

WWW モニタリングシステムの構築

岐阜大学・工学部
岐阜大学・工学部
岐阜大学・工学部

*菱田 隆彰
小林 大祐
寒河江雅彦

HISHIDA Takaaki
KOBAYASHI Daisuke
SAGAE Masahiko

1. はじめに

インターネットの普及に伴って、WWW(World Wide Web)を利用したサービスが最近急増している。それと同時にWebによる情報伝達、業務処理、情報収集、販売、広報活動が、急速に普及し始めている。また、利用できる媒体も一般的なパーソナルコンピュータだけでなく、携帯電話などさまざまな通信メディアからインターネットを利用したサービスへアクセスが可能になっている。しかし、実際にその利用効果を正確に把握し、解析することは難しく、既存のマスメディア（テレビ、ラジオ等）で行われている視聴者のモニタリングや視聴率調査等に類似した利用者に関する分析は既存のメディアに比べて遅れている。

本研究では、現在インターネット上で利用できるサービスのアクセス状況等の解析に利用可能な情報の現状とその問題点を分析する。Webのアクセス状況をより正確に把握できるモニタリングシステムを構築し、そのモニタリングデータの解析例を示す。

2. インターネットの主なサービスと解析方法の現状

インターネット上で利用できる主なサービスには以下のものがあげられる。

1. 電子メールサービス
2. FTP(File Transfer Protocol)サービス
3. Telnet (仮想端末) サービス
4. WWW サービス

WWW サービスは他のサービスと比べて最も新しく、文字情報と音声や画像などのマルチメディア情報が融合され、それまでのサービスにはない様々な利用可能性を持つ。1～3に比べてより多くのログ情報が入手可能である。

WWWのアクセス状況を解析するにあたって知られている解析方法はサーバログ解析法とオーディエンス法との2種類がある。

サーバログ解析法はインターネットを利用したサービスでも一般的に用いられる方法でサーバに残る各サービスのログを項目ごとに集計する方法である。

オーディエンス法はテレビの視聴率調査に使われている方法であらかじめログをとるサンプルを決めておき、リアルタイムにその軌跡を追い、その動作を記録し、実際のアクセス先を集計する方法である。両者の特徴は以下のとおりである。

サーバログ解析法

- 1サーバ上のすべての情報に関してアクセス情報がえられる。
- ログをとるための標準的な設定はあらかじめされているので、解析のための準備が少ない。
- リクエスト単位のログのため、ページ単位のアクセス状況の把握には適さない。
- サーバ単位でログをとるのでほかのサーバとの比較が難しい。
- ユーザの属性の抽出が難しい。

オーディエンス法

- 多数のサーバに対するログを取ることが可能である。
- ユーザの属性があらかじめ把握できる。
- サンプルとなるユーザの選び方や数で傾向や精度が大きく変化する。
- ユーザ側への準備が必要なため費用や準備の時間がかかる。
- 特定のサーバに関する解析には適さない。
- ユーザ側のモニター承認が必要である。

両者の特性を簡単に述べると、サーバログ解析法は特定のサーバに関してより細かなアクセス情報を解析するのに適していて、オーディエンス法は複数のサーバを同一の指標を用いて解析するときに適している。

サーバログ解析法はサーバ単位の情報を解析するため、ソフトウェアは比較的簡単に作成できるが、オーディエンス法はサンプルとして選ばれたユーザに組み込むソフトウェアとそこから発信されるログを一手に収集し、解析するソフトウェアとの2種類の構成になり、ソフトウェアが複雑となり、サンプル数が結果に大きく左右するため、そのソフトウェアを利用した組織的な解析体制が必要になる。そのため、市販されている例は少ない。

インターネットが商用利用される以上、「どのような人が」、「どのようなサービスを」、「どのような時間に」、「どのくらいの間」、利用しているかといった情報はインターネットの効果を正確に把握する上で非常に重要である。しかし、現状ではテレビやラジオなどの既存のマスメディアに比べてアクセス状況の解析方法はかなり遅れており、実際サーバログ解析法ではログに残る情報を有効に活用できていないように思われる。

3. モニタリングシステムの基本方針

本モニタリングシステムでは、既存のモニタリング情

報に加えて、現状のサーバログを使ったシステムでは解析の難しい

- 滞留時間
- リピータ頻度

などに注目し、システムの構築を行う。ここでいう滞留時間はユーザ（クライアント）があるページを閲覧し始めてから別のページへ移るまでの時間を表し、リピータ頻度は特定のユーザが同じページにアクセスした回数を表す。また、解析方法はサーバログ解析法を利用し、クライアント側の設定は一切なく、サーバ側のみのシステムで動作することが可能なものとする。解析に利用するアクセスログに関しても、Webサーバが記録する標準的な形式のログを利用し、サーバ側に特別な設定は行わないようにした。

4. システムの構築とその解析例

WWW 上のホームページのアクセス状況がより正確に把握できるサーバログ解析法を用いたモニタリングシステムの試作を行い、実際のモニタリングデータの解析例を示す。本モニタリングシステムでは以下の情報が解析可能である。

1. アクセス数（時間別/日にち別/ドメイン別）
2. 滞留時間
 - ① html ファイル別（ページ別）
 - ② ディレクトリ別（ディレクトリ以下に滞留した時間）

3. リピータ頻度

(1)アクセス数

アクセス数を求めるシステムでは、時間別アクセス数、日にち別アクセス数、ドメイン別アクセス数をそれぞれ求めることができる。

(2)滞留時間

滞留時間を求めるシステムは、html ファイル名（ページ名）、ディレクトリ名を指定することで、それぞれhtml ファイル別滞留時間、ディレクトリ以下に滞留した時間を求めることができる。滞留時間を得ることで、アクセス数だけでは得ることができなかったユーザの利用状況が分かる。

(3)リピータ頻度

リピータ頻度を求めるシステムは、html ファイル名（ページ名）を指定することで、ドメイン名（ドメインとはある組織やサイト、ホストなどが属するドメイン（領域）を表わす文字列のこと）別にそのページに何回アクセスがあったかを示す。

5. モニタリング情報の解析例

本システムの具体的な解析例を示すために岐阜市のホームページのアクセスログ情報を 20 日分（1998 年 9 月 2 日から 21 日）解析した例の一部を示す。

図 1 岐阜市トップページの時間別アクセス数

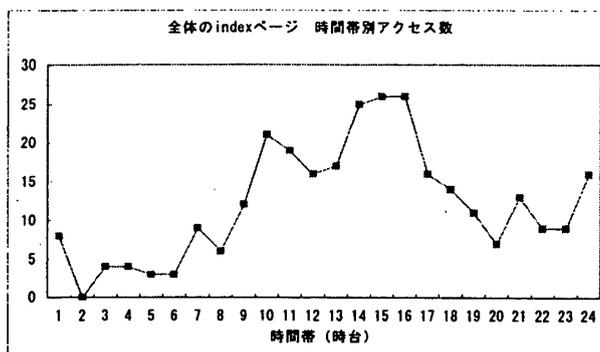


図 2 岐阜市トップページのドメイン別アクセス数

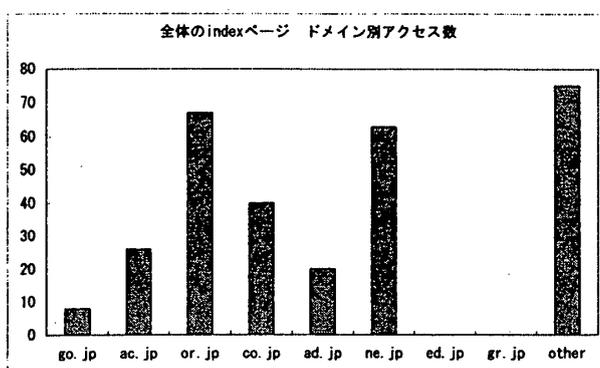
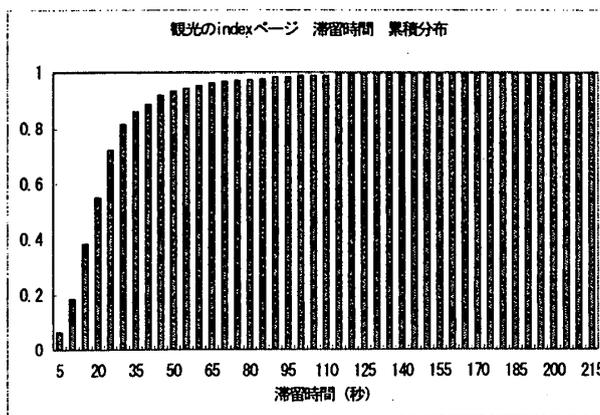


図 3 岐阜市観光案内 index ページの滞留時間累積分布



6. 解析例の考察

上記に示した解析例の考察及び他の解析結果の報告は当日発表する予定である。

参考文献

- [1] 菱田隆彰、寒河江雅彦. WWW モニタリングシステムの構築. 岐阜大学工学部研究報告, 掲載予定.
- [2] 寒河江雅彦、永坂高志. WWW モニタリングシステムとデータ解析. ソフトピアジャパン共同研究報告書, 1998, pp. 28-32.