

アクセスログをもとにしたユーザ支援の可能性について*

大垣 齊・中村 孝 (大阪産業大学工学部情報システム工学科)[†]

1 はじめに

本論文では、ユーザのウェブ情報の検索・閲覧行動を支援するための情報源を、ユーザの検索・閲覧行動の履歴そのものに求めることを提案する。すなわちユーザがウェブ情報に対してどのようにアクセスしたかという情報を用いて以降のアクセスの支援・補助を行うのである。HTTP Proxy が記録するアクセスログを利用した筆者らの研究事例を二つ報告し、その可能性について検討する。

2 アクセスログの利用例

ここでは HTTP Proxy のアクセスログを利用したシステムとして、ユーザ向けの“アクセス履歴の表示”およびコンテンツ管理者向けの“アクセスパターンの抽出”について述べる。

2.1 アクセス履歴の表示

ユーザがブラウザの機能を利用して、自発的にブックマークを記録するのではなく、アクセスログをもとにブックマークの代わりとなるものを自動的に作成しユーザに提供するシステムを構築した。

本システムは HTTP Proxy のログをあらかじめ加工し、ユーザごとに取り出したアクセス履歴を CGI を用いて各ユーザに提供するものである。ユーザの識別には ident デーモンを用いた。本システムでは、定期的にアクセスログを加工することにより、当日のアクセス履歴を表示することも可能である。アクセス履歴の提供方法として、アクセスログからユーザごとの履歴を整理し表示するページと、HTTP Proxy がキャッシュしている情報を元に過去に閲覧したコンテンツからキーワ

ード検索を行うページを作成した。

アクセス履歴は、過去 1 週間分のアクセスログをもとに URL ごとのアクセス回数を点数化し当日のアクセスほど高い点数、過去になるにしたがって得点が低くなるように点数付けを行い、合計点数の高いものから表示する方法を採用した。アクセス履歴の表示例を図 1 に示す。表示はこのシステム専用の URL にアクセスするだけで、ユーザ識別を行い、そのユーザの過去のアクセス履歴を表示する。表示される URL はクリックブルで、各 URL の示すページへのリンクとなっている。

現在は詳細表示 [標準] [簡素]

```
(362) www.estyle.ne.jp
□ (335) www.estyle.ne.jp/%7Eandrew/
  ☆ (324) www.estyle.ne.jp/%7Eandrew/style.css
  ☆ (10) www.estyle.ne.jp/%7Eandrew/a/link.css
  ☆ (4) www.estyle.ne.jp/%7Eandrew/
□ (16) www.estyle.ne.jp/control/
  ☆ (12) www.estyle.ne.jp/control/url'update2.phtml
  ☆ (2) www.estyle.ne.jp/control/url'update.html
  ☆ (2) www.estyle.ne.jp/control/index.html
```

図 1: 履歴表示の例 (詳細表示)

また、HTTP Proxy は HTTP Cache であるため、過去に閲覧したウェブページの内容をハードディスクなどに記録してある。Namazu[1] を用い、HTTP Proxy (Cache) がハードディスクに保存しているキャッシュファイルから検索用インデックスを作成することにより、過去に閲覧したページを対象とした検索システムを構築した。

2.2 アクセスパターンの抽出

筆者らはプログラミング演習等の授業に関する情報や資料等をウェブを用いて学生に提供してきた。また、これらのコンテンツを用いたウェブ教材改良のため、アクセスログの解析によりユーザ

[†] Possibility of user support using access log.

[§] OHGAKI Hitoshi, NAKAMURA Takashi(Osaka Sangyo University, Department of Information System Engineering)

のアクセス状況をアクセスパターンとして抽出するシステムを構築してきた [2, 3]。

本論文におけるアクセスパターンとは、ユーザがウェブサーバ内のページをどのような順番でたどって閲覧したかを示す情報である。アクセスログから学習者がどのようにそれらのページを閲覧していたかを“アクセスパターン”として抽出するシステムを構築した。

ユーザの識別を利用せずに全体のアクセスパターンを得ることで、

- 1) 学習者全体の傾向を知る
- 2) 教材提供側のシナリオとのマッチングをはかることが可能となり、これらから得られる情報をもとに、
 - 1) ユーザへのナビゲーションの変更
 - 2) コンテンツの(自動)改良・変更を行うことが可能となる。

またユーザを識別し、学習者ごとのアクセスパターンを得ることで、学習者一人々の学習傾向を知ることが可能となり、学習者個々へのアドバイスが可能となる。

3 アクセスログ利用の問題点

本論文で紹介したアクセスログの利用例では以下の様な問題が存在する。

- 1) システムの利用が特定の HTTP Proxy を利用するユーザに限定される。
- 2) セキュリティ対策が必要とされる。

1) については、HTTP Proxy のアクセスログを利用するというシステムの性質上、現時点では問題回避方法を提案することはできないが、履歴の表示機能や検索機能を充実させることにより、ユーザにブラウザの設定の負担以上のメリットを与えることができれば、この問題は解決するのではないかと推測される。

2) については、ユーザ識別を `ident` に頼っているため `ident` で嘘のユーザ情報を返すことで、他のユーザになりすましそのユーザの履歴を見ることができてしまう。ブラウザのクッキーとユーザ認証を組み合わせる等の `ident` を利用しない方法を検討する必要がある。

4 今後の課題

本論文では、HTTP Proxy のアクセスログを利用した“アクセス履歴の表示”および“アクセスパターンの抽出”について報告した。今後の課題として、以下のものがあげられる。

4.1 履歴表示方法の検討

現時点では、アクセスログに記録されたすべての記録をもとに履歴を表示しているため、ユーザが必要としている履歴が見つげにくい。また、間違っていてアクセスしたものや、コンテンツの構成上ユーザが意図しないアクセスによるものなどを除外できない。また、履歴から不要な URL をユーザが能動的に消去したり、表示順を変更することができない。

さらに、ユーザが過去にアクセスした URL をどの程度再利用するかを解析し、履歴の表示の対象期間や点数付けの方法を検討する必要がある。

4.2 アクセスパターンの表現方法の確立

現状では、アクセスパターンは単純なテキストなどで表示している。これを

- 1) コンテンツの管理者がより理解しやすい形式
- 2) コンテンツの自動改良のためのパラメータとして利用できる形式

で表現する方法を検討する必要がある。

コンテンツの自動改良のために、コンテンツに専用のマークアップを行うなど自動改良を行う際にプログラムが利用できる手がかりを埋め込む必要がある。また、アクセスパターンを元にコンテンツ改良の情報を得る手法や、自動改良の手法を確立する必要がある。

参考文献

- [1] Namazu Project: Namazu, <http://www.namazu.org/>
- [2] 大垣 斉・中村 孝: ウェブ教材改良のためのアクセスパターンの解析, 情報処理学会第 58 回全国大会講演論文集 4X-11 (1999)
- [3] 大垣 斉・中村 孝: ウェブ教材改良のための情報抽出 -アクセスログの解析から-, 人工知能学会第 27 回知的教育システム研究会 (1999)