

WWWアクセス活動とWebコンテンツの情報統合における 履歴抽出精度の向上とその応用

8W-03

石川雅弘 後藤文太郎

北見工業大学 情報システム工学科

1. はじめに

ユーザはWWW利用時において、自身が意識しているかどうかに関わらず、個々の持つ様々なノウハウを活用している。履歴情報にはノウハウに基づいたユーザ独自の特徴が現れていると考えられる。この特徴からノウハウを抽出し道具として提供できれば、WWWを利用するユーザにとって大きな支援効果があるものと考えられる。

我々は履歴情報の中にあるユーザのWWW利用時に用いるノウハウの知識を抽出し、WWW利用支援への適用を試みてきた[1]。しかし履歴情報をプロキシサーバで取得する方法[1][2]ではブラウザからWWWに対するアクセス履歴以上の履歴情報を獲得することができない。アクセス履歴だけでなく、あるWebページをアクセスするまでに至った経過や、アクセスしたWebページに対するユーザの反応、ブラウザ上でのアクションを含めた詳細な履歴情報にこそ、WWW利用支援効果を高めるための手があると言える。

そこで本研究では、ユーザの履歴情報を独自に開発したブラウザで取得することで履歴抽出精度を向上させた。そして履歴情報から抽出したURL毎の統計情報やアクセスパターン等のWWWアクセス活動とWebコンテンツとの情報統合を行った。情報統合により、リンク選択判断材料やユーザにとって重要度の高いWebページへの直接リンクの提供によるWWW利用支援を実現した。

2. 履歴情報とWWWアクセス活動

WWWを利用する場合、ユーザは通常ブラウザを用いる。そしてユーザはある目的に対して自身のノウハウに従ったWWWアクセスを行う。ブラウザ、WWW間におけるHTTPプロトコルによるリクエスト、レスポンスの一連の記録をアクセス履歴と呼ぶことにする。アクセス履歴だけでなく、ブラウザの

利用状況やブラウザ上でのアクション等、ブラウザでの活動状況も含めた利用記録を履歴情報と呼ぶことにする。ユーザがある目的のためにWWWに対して行う自身のノウハウに基づいた活動手段のことをWWWアクセス活動と呼ぶことにする。本研究では履歴情報からWWWアクセス活動の抽出を行いWWW利用支援への適用を試みる。

3. 履歴抽出精度の向上

本研究ではユーザの履歴情報を独自に開発したブラウザ(I3ブラウザ)で取得する。I3ブラウザを用いることにより、プロキシサーバで取得可能であったアクセス履歴に加え、ブラウザの起動、終了状態やブラウザ上での正確な表示ページの取得、ブラウザ上でのユーザのアクション等を含めた詳細な履歴情報の記録が可能になる(図1)。

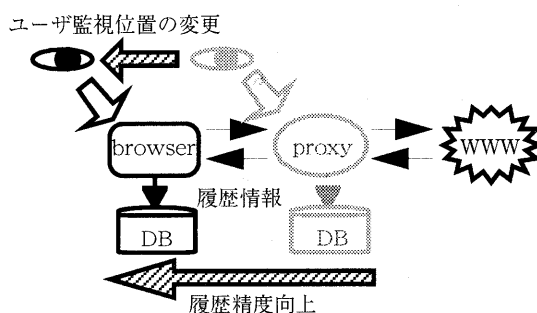


図1. 履歴取得位置の変更による精度向上

4. WWWアクセス活動を用いた利用支援

履歴情報から抽出したWWWアクセス活動とWebコンテンツの情報統合により実現したユーザに対するWWW利用支援手段を以下に示す。

リンク選択判断材料の提供

URL毎のアクセス回数、滞在時間等の統計情報から得られるユーザのWWWアクセス活動をWebコンテンツと情報統合する。情報統合によりユーザはWebコンテンツ中に含まれる全てのリンクに対するリンク先Webページの利用頻度や状態をリンク選択判断材料として利用可能になる。

重要度の高いWebページへのJumpリンクの提供

あるWebページからの連続したリンクの辿り方をアクセスパターンと呼ぶことにする。ユーザのアク

セスパターンとブラウザでのWebページ閲覧時の振る舞いから得られるWWWアクセス活動を用いたWWW利用支援としてJumpリンクを提案する。

Jumpリンクとは、あるWebページからリンク構造を辿って到達可能な重要度の高いWebページに向けた、起点となったWebページからの直接リンクである(図2)。ここで重要度の高いWebページとは、ユーザが閲覧時に次の動作を行ったWebページを指す。

- (1) ブックマークに登録した。
- (2) 閲覧に時間(2分以上)を費やした。

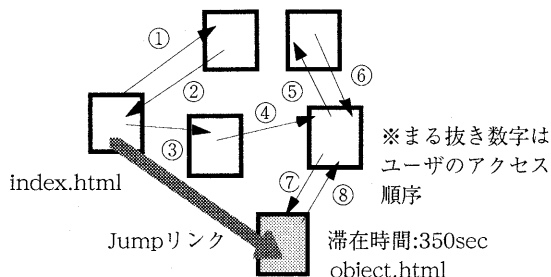


図2. Jumpリンク生成例(滞在時間による場合)

JumpリンクとWebコンテンツの情報統合により閲覧するWebページからリンクを辿って到達可能な、重要度の高いWebページへの直接リンクが利用可能になる。

5. システムアーキテクチャ

本研究で構築したシステムのプロットを図3に示す。システムを利用する際には独自に開発したI3ブラウザを使用する。利用者のブラウザ上での振る舞い及びWWW利用状況は閲覧するWebページが切り替わる度に逐次イベント管理サーバ(EMS)に送信される。EMSに蓄積された利用者の履歴情報からのWWWアクセス活動の生成は、analysis moduleで行なう。I3ブラウザからのWWWアクセスは監視プロキシサーバ(WPS)を介して行う。WPSはI3ブラウザからURL部分が(a)のように変更されたHTTPプロトコルによるリクエストを受信する。

/Function/BrowseID/URL ... (a)

(a)におけるFunctionはブラウザに返す情報の形式であり、URLに対する通常のWebコンテンツを返すnomal、Webコンテンツに統計情報を統合して返すstatistics、Jumpリンクを統合して返すjump等がある。BrowseIDはシステム内で履歴情報とコンテンツ情報の対応を取るために用いる。

WPSではリクエストに含まれているFunction、BrowseID等を情報源関連管理サーバ(RMS)に送信することでWWW及びシステム内の各種サーバを結び付けて活用するための処理プロセス、つまり情報統合プロセスを受け取る。ここで得られるプロセス

に従ってWPSはシステム内の処理を進める。

例えばfunction=statisticsでは、まずリクエストに基づいてWWWをアクセスし(①)、取得したレスポンスをコンテンツ管理サーバ(CMS)に送ってDBに保存する(②)。更にWPSはEMSより、レスポンス内のオリジナルHTMLソースに含まれる全てのリンクに対するアクセス回数、滞在時間等のXMLデータを取得し(③)、オリジナルのソースに埋め込んでI3ブラウザに返す(④)。I3ブラウザではXMLデータを解析し必要に応じて解析結果をWebページにダイナミックに挿入することで統計情報とWebコンテンツの情報統合を実現する(図4)

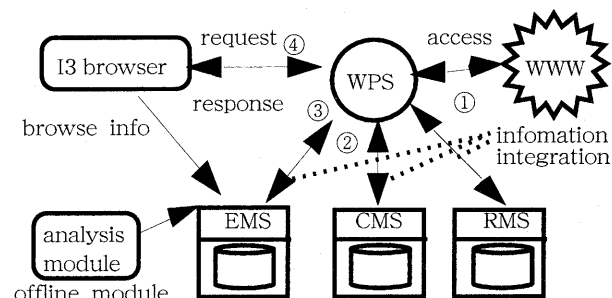


図3. システム概念図

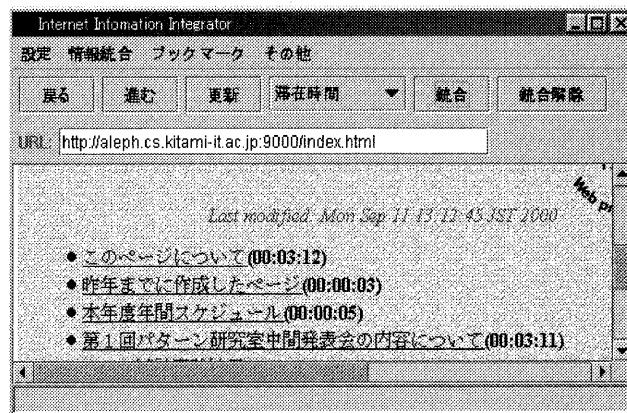


図4. I3ブラウザにおける滞在時間の情報統合

6. おわりに

利用者の履歴情報の獲得を独自に開発したブラウザで行うことで、プロキシサーバで行うシステム[1]と比べて正確で詳細な履歴情報の取得が可能になった。更に精度の向上した履歴情報から抽出したWWWアクセス活動とWebコンテンツの情報統合によりWWWによる情報獲得支援を実現した。

参考文献

- [1] 石川雅弘、後藤文太郎: "WWWアクセス活動とWebコンテンツの情報統合", 第60回情報処理学会全国大会講演論文集(3)p157-158, 2000
- [2] Alan Wexelblat et al: "Footprints: History-Rich Tools for information Foraging", Proc. of IEEE's 32nd Hawaii International Conference on System Science(HICSS'99)