

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

WWWのダイナミック性は最新の情報を即座に入手可能としたが、一方で情報の損失という問題の原因となる。情報の損失とは、ブックマーク機能のように獲得した情報をURLのみで保存した場合、再アクセス時に、内容の更新、削除によって、以前には得られた情報を失うことである。情報の損失を回避するためにはURLのみの保存ではなく、アクセス情報全体、つまり利用者のWWWにおけるアクセス活動の保存が必要と言える。SPIRAL^[1]は、Webページにアクセスした時点でのWebページそのものの情報を自動的に保存し、アクセス時刻に応じた保存済みWeb情報へのアクセス機能を提供している。しかし、SPIRALにおけるDB検索機能は、WebページあるいはWebページ間のグループに限定されている。

そこで我々は、SPIRALにおけるDB検索機能を強化し、保存したWebページだけでなく、アクセス先使用状況の検索を可能にしたのに加えてWebコンテンツとの情報統合を行った。情報統合により、WWWブラウジングの流れの中で以前に使用したWebページを利用することを可能にした。更にアクセス先の使用状況をWebページに反映させることで、WWW上の情報取得支援を実現した。

2. WWWアクセス活動のDB化

WWWアクセス活動とは利用者が「いつどのようにWWWをアクセスし、レスポンスを受け取ったか」という一連の情報である。この情報を管理し効率的に利用するためにDBに蓄積する。利用者のア

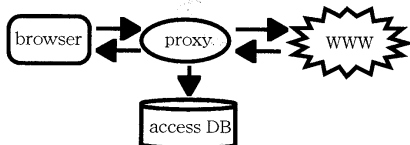


図1. 利用者のアクセス活動の取得

Information Integration of
WWW Access Activities and Web Contents
Masahiro ISHIKAWA and Fumitaro GOTO
Department of Computer Sciences
Kitami Institute of Technology
165, Koen-cho, Kitami 090-8507, Japan

クセス活動の監視、DB格納処理はProxyサーバで行う(図1)。Proxyサーバでは次の情報を1つのリクエスト毎にDBに保存する。

```

access(User,Time,URL,Method,Response)
User      : 利用者
Time     :  アクセス時刻
URL      :  アクセス先URL
Method   :  使用メソッド
Response :  取得したWebコンテンツ
  
```

構築したWWWアクセス活動DBを用いることで以前に使用したページを取り出し利用することや、各Webページの利用頻度を知ることが可能となる。しかし、WWWアクセス活動DBを単独で用いるのみの利用方法の提供だけでは、インターネット環境での利用しやすいインターフェースとは言えない。そこで、異種情報源の統合利用、即ちWWWアクセス活動とWebコンテンツの情報統合を行う。WWWアクセス活動DB検索、Webコンテンツとの情報統合をシステムが行うことで、利用者はDBを意識せずに、DBの内容が反映されたWebページの閲覧が可能になる。

3. 情報統合

情報統合とは分散しているばらばらな情報源から、個々の経路が持つ特性を相補的に生かしながら全体として整合性のある表現にまとめ上げることを言う^[2]。情報統合には様々な形態があるが本稿における情報統合は複数の情報処理過程からの出力の統合、即ち、水平型の統合を指すものとする。情報統合によって特定の情報伝達機能の性能の向上が期待できる。

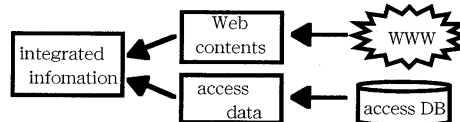


図2. WWWアクセス活動とWebコンテンツの情報統合

本研究における情報統合の対象はWWWアクセス活動とWebコンテンツである(図2)。WWWアクセス活動とWebコンテンツの情報統合により通常のWeb

ブラウジングの流れの中で以前に使用したページの内容が利用可能となる。また利用者のアクセス状況をWebブラウジングにおけるリンク選択の判断に利用することでWWW上の情報取得支援が可能となる。

4. 情報統合システム

システム概念図を図3に示す。システム起動中はブラウザからcgiを介してWWWをアクセスする。

cgiではまず受け取ったリクエスト先URLに対応するWebコンテンツをWWWから取得する(①)。

次に取得したWebコンテンツ内の全てのリンク先URLを抽出する(②)。②で得られたURLの集合をkeyとしてDBをアクセスし次の情報を得る(③)。

アクセス先の使用状況：

- アクセス回数
- 総滞在時間
- アクセスまでにかかる平均時間

また各URLに対するヘッダ情報をWWWから取得し次の情報を得る。

アクセス先の状態：

- ファイルサイズ
- 接続可否情報

以上で得られたリンク先URLに対する情報をcgiではXML形式に変換しXMLを操作する javascript とともにブラウザに返す。またcgiではWWWからのコンテンツに対して次の変換、追加を加えてブラウザに返す(④)。

(1)AタグにおけるHREF属性値の変換及びID属性、onClick属性の追加。

(2)その他のタグの属性値における相対アドレスから絶対アドレスへの変換。

(1)におけるHREF属性値はWWWアクセスを連続してcgi経由で行うために変換する。ID属性はリンクとXML形式の情報の対応付け、情報の挿入の際に用いる。onClick属性はクリック後に新しく表示されるWebコンテンツとXMLの同期を取るために用いる。また、cgiを介してWWWをアクセスしているため、相対アドレスはそのまま利用できない。このため(2)に示した相対アドレスから絶対アドレスへの変換が必要となる。システムでは以上に示した①から④の処理がシステム終了まで繰り返し実行される。

システムの実行画面を図4に示す。システム実行時には、ブラウザ上部にボタンが配置される(図4(a))。書き込みボタンのクリックによりWebページ上にイメージが埋め込まれ、利用頻度等に応じて色分けされる。イメージをなぞると実際のデータが表示され、イメージをクリックすると以前に使用したWebページの内容へのリンクが利用できる(図4(b))。

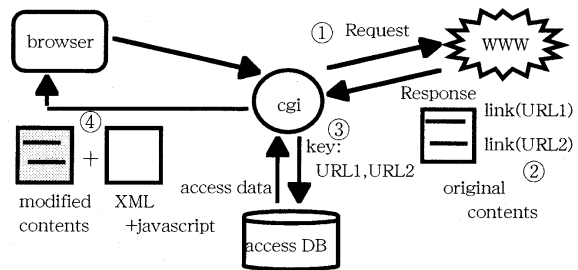
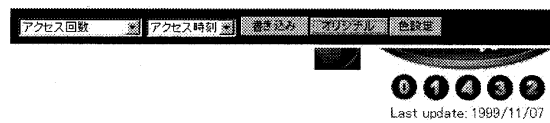


図3. システム概念図



このページは石川隆弘のホームページです。
このページは、「WWWと論理プログラム」の内容を中心としたページです。只今、WWWにては、研究内容を御覧下さい。

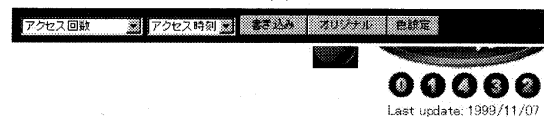
項目

初めて御覧の方は、What's newを見る前に研

● [What's NEW](#)

● [プロフィール](#)

(a)



このページは石川隆弘のホームページです。
このページは、「WWWと論理プログラム」の内容を中心としたページです。只今、WWWにては、研究内容を御覧下さい。

項目

present...	http://zeta.cs.kitami-it.ac.jp/8080/
1999-7-11-18 28:53...	http://zeta.cs.kitami-it.ac.jp/8080/
1999-7-20-9 41:56...	http://zeta.cs.kitami-it.ac.jp/8080/
2000-1-11-20 9:52...	http://zeta.cs.kitami-it.ac.jp/8080/
2000-1-11-20 10:16...	http://zeta.cs.kitami-it.ac.jp/8080/
2000-1-11-20 11:17...	http://zeta.cs.kitami-it.ac.jp/8080/
2000-1-11-20 12:38...	http://zeta.cs.kitami-it.ac.jp/8080/

● [What's NEW](#)

● [プロフィール](#)

(b)

図4. システム実行画面

5. おわりに

WWWアクセス活動とWebコンテンツの情報統合によりWWWのダイナミック性が引き起こす情報損失の回避及びアクセス先の利用状況、状態の把握をWWWブラウジングの流れの中で実現した。

参考文献

- [1] 松井英昭、後藤文太郎：“WWWのシームレスな利用環境の構築”,人工知能学会研究会資料SIG-KBS-9803,pp.37-41.
- [2] コミュニケーション社会の情報技術の調査・提言委員会(電子技術総合研究所)：“情報主導社会における情報学研究のあり方”, <http://ci.etl.go.jp/teigen/main.html>.