

# WWW 更新ページ報知サービスシステムの開発

4 S - 3

藤田 義之      遠藤 裕英      武藤 英男      松尾 修      小川 和美

(株)日立製作所 システム開発研究所

## 1 はじめに

現在インターネット上には膨大な数の WWW(World Wide Web)サーバが存在する。研究者はこれらの WWW サーバから、仕事に役立つ情報を掲載しているページを探索し利用している。しかし、それらのコンテンツは情報提供者により不定期に更新されるため、研究者は当該ページの更新有無を常に監視している必要が生じている。

今回我々は、WWW ページの更新情報を自動的に報知するシステムを開発した。本システムの概要、及びその効果について報告する。

## 2 システムの概要

本システムのシステム構成を図 1 に示す。

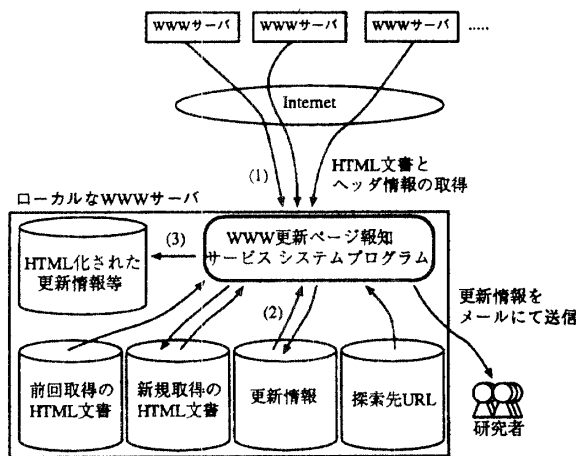


図 1: システム構成図

本システムの機能を大きく分けると、次の3つになる。

- (1) HTML 文書の取得
- (2) 更新情報の作成
- (3) ローカルな WWW サーバへの情報掲載

以下、上記機能についての概要を説明する。

### 2.1 HTML 文書の取得

この機能は、予め探索先として登録してある URL(Uniform ResourceLocator) のリストを調べ、該当する HTML(Hyper Text Markup Language) 文書を定期的にアクセスし、ローカルにキャッシュする。

HTML 文書の取得は、まず該当する HTML 文書のヘッダ情報を取得する。この情報から最終更新日等の情報を取り出し、HTML 文書を取得するか否かを決定する。しかし WWW サーバによっては、最終更新日の情報が無い場合がある。このような時は無条件に HTML 文書を取得する。

### 2.2 更新情報の作成

本機能は、今回キャッシュした新 HTML 文書内容と、前回取得していた旧 HTML 文書内容を比較し、更新情報を出力する。

更新情報作成の流れを次に示す。

1. 新規に取得した HTML 文書中から、全てのリンク先 URL を取り出す。
2. 1 と同様に、前回取得した HTML 文書中から全てのリンク先 URL を取り出す。
3. 1 と 2 で求めたリンク先 URL の比較を行い、新たに追加されたリンク先 URL、削除されたリンク先 URL の情報を作成する。

4. 追加されたリンク先 URL の情報を基に、その URL を示す語句を取り出し対応表を作成する。例えば HTML 文書内で、

```
<A HREF="/Link"> Text </A>
```

のように書かれていれば、リンク先 URL が /Link 語句は Text となり対応表は

```
/Link    Text
```

となる。この対応表が更新情報となる。

### 2.3 ローカルな WWW サーバへの情報掲載

本機能は、作成された更新情報等を全て HTML 文書に変換し WWW サーバ上に掲載する。(図 2)。また、研究管理者にはこの情報をメールにて通知する。

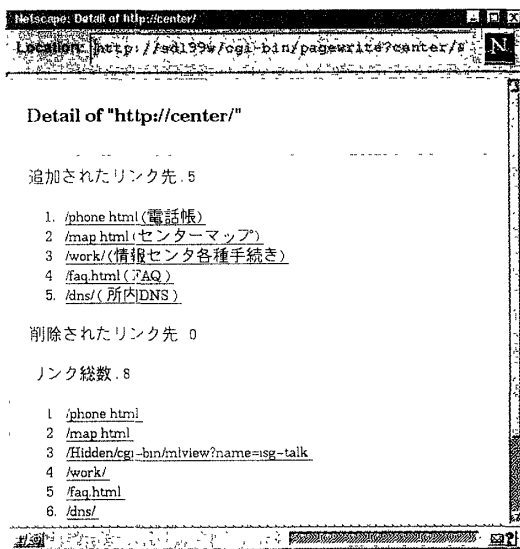


図 2: Netscape Navigator<sup>†</sup>による更新情報の表示例

しかし、図 2 のような更新情報だけでは、何についての情報について書かれているのかを理解するのは難しい。WWW ページ全体を見て、その中で何処が変更になっているのかが分れば、前後の文章を読み何についての情報であるかを容易に理解できる。

そのため本機能には、作成した更新情報を基にオリジナルの HTML 文書中の更新箇所に [New] マークを挿入し、WWW サーバに掲載する機能を持たせている (図 3)。

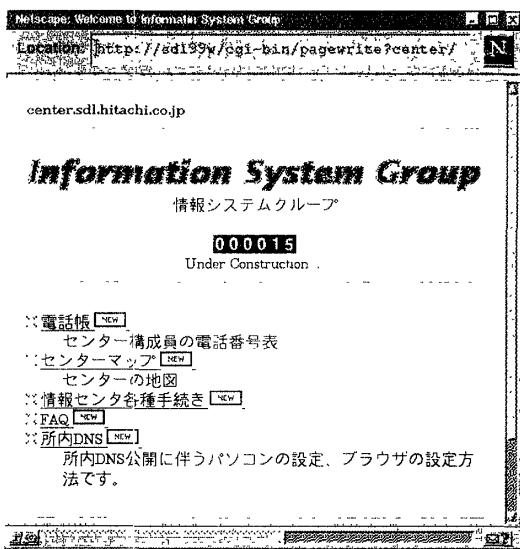


図 3: Netscape Navigator<sup>†</sup>による更新箇所の表示例

### 3 システムの運用効果

業界トップメーカー数社の WWW トップページを対象に、本システムを試行運用してみた。約 3 ヶ月間のアクセス状況の統計を取ってみたところ、頻りに内容が更新されているサイトへのアクセスは、約 30% 程の増加が確認できた。しかし、あまり内容の更新がされていないサイトへのアクセスは、殆ど変化は見られず、むしろアクセス数が減少している期間もあった。

このように、更新頻度が多いサイトへのアクセス数が多く、更新頻度が少ないサイトへのアクセス数が少ないという結果から、本システムの効果を確認することができた (図 4)。

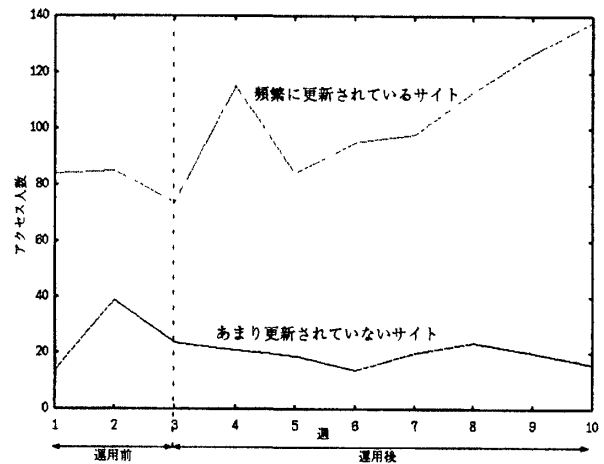


図 4: 本システム運用によるアクセス状況の変化

### 4 おわりに

現在、本システムは管理者により探索先の URL を決定している。しかし、この探索先 URL が示すページに、研究者が本当に必要としている情報を掲載しているかは定かでない。

個人レベルで探索先の URL を決める事ができれば、研究者にとって情報探索の時間を大幅に短縮する事が可能になると思われる。そのためには、研究者が自由に探索先 URL を入力できるユーザインタフェース部の作成が必要である。このユーザインタフェース部の作成が今後の課題である。

### 参考文献

- [1] Laura Lemay 著, 武舎広幸, 久野禎子, 久野靖訳: "HTML 入門 - WWW ページの作成と公開", トッパン, 1995
- [2] Laura Lemay 著, 武舎広幸, 久野禎子, 久野靖訳: "続・HTML 入門 - 新機能, CGI, Web の進化", トッパン, 1995

<sup>†</sup>Netscape Navigator は、Netscape Communications Corp. の商標です。