

WWWにおける情報検索の高度化について

3 F - 6

本田 栄司

株式会社 インテック・システム研究所

1. はじめに

近年、クリントン政権の「情報スーパーハイウェイ」構想などをきっかけにインターネットの利用者が爆発的に増加している。利用者の多くは電子メール、FTP、Telnetなどの従来からのインターネット上のサービス利用に加え、最近ではWWW (World Wide Web) などの情報検索サービスも盛んに利用されている。現在、全世界で公開されているWWWサーバ数も、利用者と同様急激に増加している。

WWWサーバにアクセスするためには、通常MosaicやNetscapeなどのWWWブラウザを用いる。WWWブラウザ上では、すでに知っているURL (Universal Resource Locator) を直接指定するか、あるいはリンクをたどって行くという方法でアクセスできる。つまり、世界中に数多く存在するWWWサーバの中から目的とするサーバを見つけるためには、非常に時間と労力を要する状況となってきた。

そこで、本稿では1度アクセスした情報を中継サーバに分類し蓄積しておき、これらの蓄積情報を使って現在利用者が見つけようとしているサーバと類似のサーバの情報を提供する方法を説明する。

2. アクセス情報の蓄積

現在、利用者がアクセスしたURLの保存はクライアント側のWWWブラウザでホットリスト (Hot list) あるいはブックマーク (Bookmark) などの機能で行うことができる。しかし、アクセスしたURL情報を加工・処理したいような場合は、クライアント側でのURLの保存では保存情報を加工するためにWWWブラウザに手を加える必要があること、他の利用者の保存情報が利用しにくいこと、セグメントのすべてのアクセス情報を統一的に扱えないことなどの欠点がある。そこで、下図のようにURLとタイトルなどのアクセス情報 (マップと呼ぶ) を、アクセスがある度にマップマネージャが利用者別 (個人マップ) と全体 (アクセスマップ) とに分けて蓄積することを考えた。

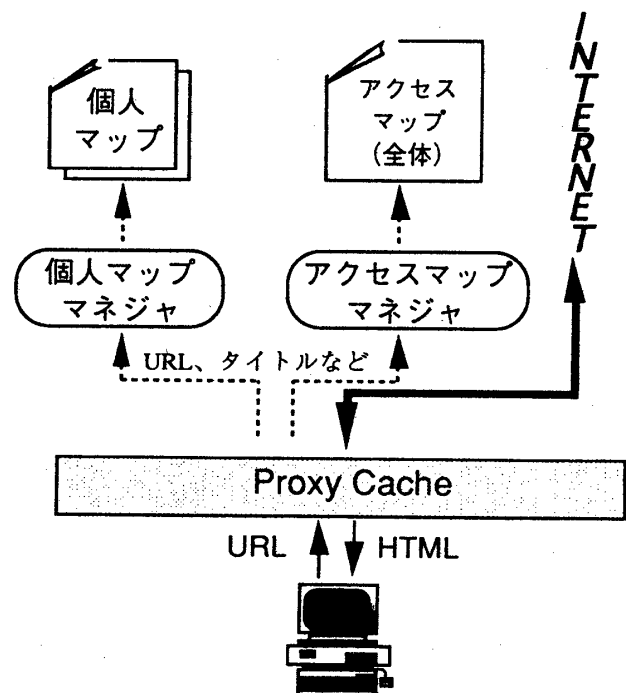


図2-1 アクセス情報 (マップ) の蓄積

An Enhancement of Information Retrieval for WWW

Eiji Honda

Intec Systems Laboratory Inc.

3-23 Simoshin-Machi, Toyama 930, Japan

E-Mail: honda@isl.intec.co.jp

(1) マップの構造

マップは個人マップ、アクセスマップとも同じで、1つのURLに対してカテゴリ、URL、タイトル、アクセス回数の構造である。これらのデータの組をマップデータと呼ぶ。

なお、個人マップは利用者の特性を表現していると言える。従って、個人マップ間の類似度（例えばマップに含まれる共通のカテゴリ数など）を求めることにより同じことに興味を持っている個人やグループなどを特定できる。

(2) マップマネージャ

マップマネージャはアクセスされたURLに対して先に述べたマップデータを作成する働きをする（常駐プロセス）。マップデータ作成で一番問題となるのがカテゴリの割り当てである。我々のアプローチでは、予めファジィシソーラス（類義語辞書）を用意しておきマップデータのURLとタイトルに含まれるキーワードとファジィシソーラスの類似度[1]からカテゴリの割り当てを行う。ただし、キーワードがどれもみな類似しないなどマップマネージャが判断を下せない場合またはマップマネージャの結論が人間から見ておかしいと判った場合などは人間がカテゴリの割り当てを行う。

3. 類似サーバのURL情報の提供

ある利用者のアクセスした情報と類似するサーバのURL情報を提供できれば、探索のための労力やネットワークや計算機資源などの節約に非常に有効であると考えられる。

我々の類似するサーバのURL情報を求める方法は、利用者の個人マップのカテゴリと全体のアクセス情報として蓄積されているアクセスマップのカテゴリのマッチングを行い、もしマッチしたカテゴリがあればアクセスマップ内のそのカテゴリを持つURLが類似するサーバ情報である。

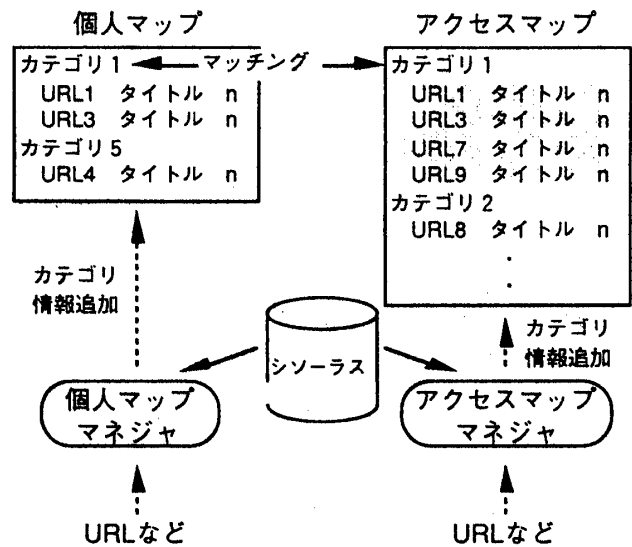


図3-1 類似サーバのURL情報の抽出

4. むすび

本稿では、マップマネージャによるURL情報の蓄積と個人マップとアクセスマップを利用した類似サーバ検索の方法を説明した。WWWサーバが爆発的に増加している状況において、ネットサーフィンで溺れないようにするためにも早期にこのような機能が必要になってくることが予想される。

今後はカテゴリの割り当てを再検討するとともに、プロトタイプを試作し評価する予定である。

参考文献

- [1]宮本ら:ファジィ命題索引にもとづく文書検索と画像検索、日本ファジィ学会誌 Vol.3 No.1、1991.
- [2]平岩、神田:Info-Plaza:進化的クラスタリング機能を備えたWWWマルチブラウザ、グループウェア11-10、1995.
- [3]D.Wessels: Intelligent Caching for World-Wide Web Objects、INET'95、1995.
- [4]A.Luotonen: Proxy caching、<http://www.w3.org/hypertext/WWW/Daemon/User/Config/Caching.html>、1994.