

WWWアクセスアドバイザーエージェント ～興味に基づくWWWウォッチング～

4C-1

朝倉敬喜、喜田弘司、垂水浩幸
{asakura,kida,tarumi}@obp.cl.nec.co.jp

NEC 関西C&C研究所

1 はじめに

インターネットの普及に伴ってネットワークから獲得できる情報量も急激に増加している。そこから所望の情報を抽出するため我々は電子メール、ネットニュース等のテキスト情報を対象とした興味ベースフィルタリング[1]について研究している。個人の興味を個人興味ベースで管理し、ユーザの興味にあった情報を優先的に提示するものである。一方、新しい情報伝達手段であるWWWでは情報提供サーバへ接続するために、そのURLアドレスを知らなければならないという問題が存在する。本研究ではこれまで検討してきた電子メール、ネットニュースのフィルタリングをWWWに拡張したWWWアクセスアドバイザーエージェント(WWWAAA=WWW Access Adviser Agent)について報告する。これはWWWのページ新設情報をネットニュースから抽出し、ユーザマシンに存在する興味ベースを利用してユーザの興味にあったページを通知するとともに、ページの継続的な監視を行い、その変化を通知するエージェントである。

2 WWWアクセス支援サービス

世界中の多数のWWWサーバから情報を獲得するための検索支援を行う代表的なサーバとしてLycos[2]等がある。ユーザはこれらへアクセスし、単語を入力して所望するページのリストを獲得し、そこから各ページにアクセスする。これらのサーバではデータベース内のページ情報が様々な分野のカテゴリにあらかじめ分類されている場合が多く、そこから所望するページを探し出すことも可能となっている。

またURL-Minder[3]はWWW監視サービスである。これはユーザが登録したURLを監視し、ページのチェックサムで変化を認識した場合、ユーザに電子メールで通知するサービスである。

これらWWWアクセス支援サーバを利用すると簡単に所望の情報ページが見つかる場合が多いが以下の問題も存在する。

- (1) WWWページを発見するためにサービス提供サーバへ接続しなければならない。つまり受動的サービスであり能動的にユーザに情報を提供しない。
- (2) WWWページの発見はできるが、URL-Minderを除いては継続して変化を監視することができない。
- (3) ローカルに保持している興味(例えばBookMark等)との連携手段が提供されていないため、興味を表現する単語(以下興味単語と表現する)の入力が手間となる。

3 WWWアクセスアドバイザーエージェント

3.1 概要

WWWへのアクセス支援のために2章で述べた問題点を解決するために以下の特徴を持つエージェントを提案する。

- ネットニュースからの興味にマッチした新設WWWページの発見・通知
- 特定WWWページの変化監視
- 興味ベースフィルタリングモジュールが生成した個人興味ベースの利用

3.2 構成

WWWAAAの全体構成を図1に示す。ネットニュース、電子メール、外部WWWの各サーバへアクセス可能なローカルサーバ上にWWWAAAを配置する。

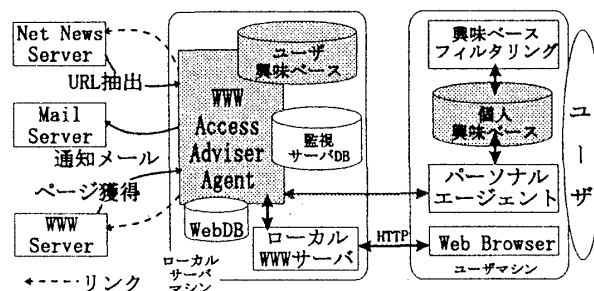


図1：全体構成

WWW Access Adviser Agent
Takayoshi ASAKURA, Koji KIDA, Hiroyuki TARUMI
Kansai C&C Res.Labs., NEC Corporation
1-4-24 Shiromi, Chuo-ku, Osaka, 540, Japan

WWWAAAは、ユーザが登録した特定のネットニュースグループから新設WWWページURLをパターンマッチングで抽出、アクセスしてページ(テキストのみ)を獲得し、獲得したHTMLファイル、URL抽出元News記事、Subject、URL、日付を1レコードとしてWebDBへ格納する。

また、ユーザから入力された興味単語をユーザ興味ベースに登録/削除するとともに、ユーザマシンのパーソナルエージェント[4]から興味単語を獲得する。興味ベースフィルタリングモジュールが生成し、パーソナルエージェントが管理する個人興味ベースは2つのプロフィール～短期興味と長期興味～でユーザの興味を表現しており[1]、これらのプロフィールから興味単語のみを抽出してWWWAAAが管理するユーザ興味ベースに登録する。パーソナルエージェント経由で登録された興味単語は、随時個人興味ベースと同期させる。新規にWebDBへ登録されたHTMLファイルにおける興味単語の含有率からスコアを算出し、ユーザの興味にあったページの判定を行う。ユーザの興味に合致したレコードをWebDBから抽出しユーザ通知ページを作成するとともに電子メールでの通知も行う。

監視サーバDBに登録されたサーバやページについては実際にアクセスしてページの変化を確認し、変化があった場合に通知する。前回アクセス時のHTMLファイルとの比較を行単位で実行している。

WWWAAAによって作成されたユーザ固有の新設ページリストの一例(画面)を図2に示す。各リスト要素は元ニュース記事のSubject、URL、WWWAAAが獲得し独自に保持しているページへのポインタ、元記事作成年月日(元記事へのリンク)から構成する。

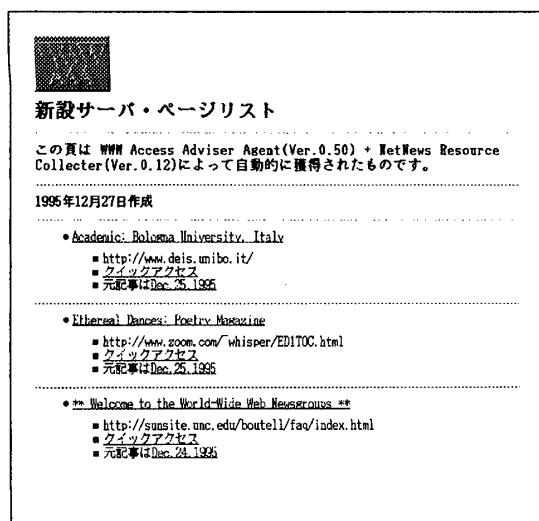


図2：画面例

4 検討

4.1 試験運用結果

一週間の試験的運用(現在は定常的に運用)の結果を表1に示す。

表1：試験運用結果

対象: comp.infosystems.www.announce	
全ニュース記事数	120件
URL全抽出件数	115件
誤ったURLの抽出件数	115件中3件
利用者興味ベースに基づいた抽出件数	115件中5件
興味に合致した件数(利用者の判定)	5件中3件

4.2 課題

ローカルWWWサーバ上にリンク情報を生かしながらページを保持しているためページ内容の確認を迅速に行うことができるが、以下の課題がある。

• 情報の階層構造

WWWではHTMLを用いて情報を階層的に表現している。そのため所望の情報が獲得したURLの数階層下に存在することがあるため、URLの示すページだけでなく、その1、2階層下までのHTMLを獲得してフィルタリング、提示する必要がある。

• ユーザフィードバックの獲得

ユーザのWeb Browserに対する操作履歴から、フィルタリング結果に対するユーザの評価を獲得し、興味ベースへフィードバックすることでより精度の高いフィルタリングを実現できる。

• ネットニュースからのURL抽出

発信者のURL等を誤抽出することがある。出現位置等を考慮した抽出が必要である。

5 おわりに

本稿ではWWWのアクセス支援サービスの一つとして、WWWアクセスアドバイザーエージェントの提案、試作、実験結果について述べた。フィルタリング技術の一応用場として今後も研究を進めてゆく。本成果は無償で提供しますのでご希望の方は電子メールでご連絡下さい。

参考文献

[1] 朝倉ほか: “エージェントによる情報フィルタリング”, 情報処理学会研究報告, IM-20-7, pp49-55, 1995

[2] <http://www.lycos.com/>

[3] <http://www.netmind.com/>

[4] 石黒ほか: “マルチエージェントオフィスシステムとその実現”, MACC'95, 1995