

逐次興味抽出を適用したブラウジング支援¹

5K-10

村本 達也 鷲崎 誠司²NTT サイバースペース研究所³

E-mail : {muramoto, suzaki}@isl.ntt.co.jp

1 はじめに

WWW(World Wid Web)上に無数に存在する情報をハイパーリンクに従って次々とブラウジングするうちに、利用者は本来の目的とはかけ離れた情報を参照していることがある。また、具体的な目的無しに、そぞろ歩きのようにWWW上の情報を巡っているうちに、興味のある情報に辿り着くこともある。このように目的としていないブラウジング状況に陥った際に、利用者が適切な情報を得ることを可能にするブラウジング支援として、WWW上の情報の隣接関係や意味の繋がりを視覚的に提示する方法 [1] や、選択肢多寡の状態から選択候補をフィルタリングする方法 [2] などが提案されている。

本稿では、利用者に現在ブラウジングしている情報が属する分類と、分類情報から推定した利用者の興味をディレクトリ型検索エンジンのカテゴリの形で提示することで、利用者が現在ブラウジングしている情報とディレクトリ型検索エンジンの該当カテゴリの情報との間をシームレスに相互移行することを可能にするブラウジングの支援方法を提案する。

2 興味抽出方法

本稿で利用する興味抽出エンジンは、(1)ブラウジングしている情報が属する分類の推定モジュールと(2)分類結果に基づく興味の推定モジュールとに分かれる。情報の分類の方法は、文書内の単語の頻度分布を基に行う方法、文書中のキーワードを抽出する方法などがあるが、本稿では利用者がブラウジングしている当該情報と当該情報が参照している情報から単語を抽出し、抽出した単語を分類体系にマッピングする方法 [3] を用いる。利用者がブラウジングしている情報を分類体系のカテゴリ(ジャンル)に抽象化して分類することにより、利用者がブラウジングしたカテゴリを利用者の興味のあるカテゴリと仮定すると、分類した結果を直接利用者の興味推定に利用できる。

利用者の興味は、前述の分類結果のカテゴリに重みを与えて、利用者のブラウジングの度に累

積計算する方法 [4] にて推定する。すなわち、あるカテゴリの重みを W とすると興味推定の累積計算を、以下の様に行う。

$$W_{new} = Const \cdot W_{old} + W_{add}$$

この式により、直近にブラウジングした情報が分類されたカテゴリにより興味があり、以前にブラウジングした情報が分類されたカテゴリへの興味は相対的に少ないと推定される。

3 ブラウジング支援の構成

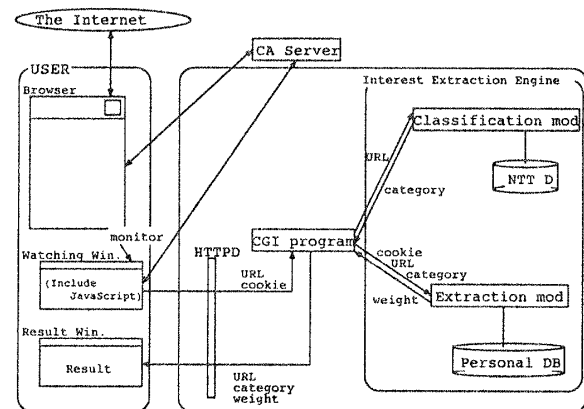


図1：機能構成の外観

我々は、利用者のブラウジング情報とディレクトリ型検索エンジン中の情報とをシームレスに移行するブラウジング支援システムの枠組として、図1で表されるWWW情報を対象としたプロトタイプを構築した。本方法では、定期的にブラウザが表示しているURLを監視することで、逐次的な興味抽出を実現している。

しかし、代表的なWWWブラウザの多くは、セキュリティ上の要請により、利用者がブラウジングしているURLを他のシステムに送ることを許していないので、何らかの方法でURLを取得する必要がある。これを解決するにはProxyログを利用する方法もあるが、本プロトタイプでは、CAサーバ(認証サーバ)を仮想的に動作させることに

¹A Browsing Support System Based on Successive User's Interests Extraction

²Tatsuya MURAMOTO, Seiji SUSAKI

³NTT Cyber Space Laboratories

よって解決した。具体的には、署名付き JavaScript を内包した監視ウインドウを立ち上げ、この署名をもって CA サーバより認証を受ける。監視ウインドウは主ブラウザを定期的に監視し、利用者が別の情報をブラウジングした際に、これを契機として httpd を通じて cookie と共に、利用者がブラウジングしている URL を取得する。

分類の推定モジュールは、取得した URL の情報を分類体系 (今回のプロトタイプでは NTT DIRECTORY[5] を使用) にマッピングし、適合度の高いカテゴリから 2 つを採用して返す。この分類結果を利用者興味推定の推定に利用する。

興味推定のモジュールでは、cookie により利用者を識別し、個人別に管理する DB に問い合わせる。個人別の DB は具体的には、(1) 個人ごとの過去のブラウジング履歴 (URL) と、(2) URL をキーとし、分類されるカテゴリを値とするハッシュテーブルから成っている。(1) に対して前述の式を適用し、利用者の興味を表す各カテゴリの重みを計算する。本プロトタイプでは、ブラウジング履歴として 30 件保持し、式中のパラメタ *Const* を 0.6、閾値を 0.1 と定め、重みがこれに満たないカテゴリは削除することにした。(2) のハッシュテーブルは主にカテゴリサーバのキャッシュ情報として利用する。これにより逐次興味抽出の高速化を図っている。

4 動作例

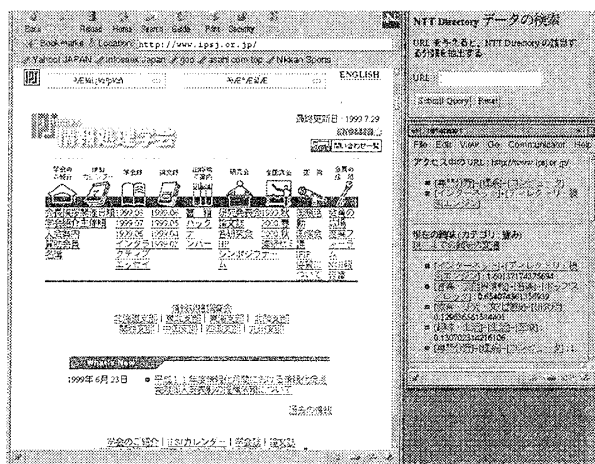


図 2 : システムの例

利用者がブラウジングしている情報の分類先と、推定された利用者の興味を結果表示ウインドウに表示した例が図 2 である。図中右上の監視ウインドウ中の JavaScript で主ブラウザの URL を監視して、解析した興味を右下の結果表示ウインドウに表示している。

個人の興味情報はカテゴリで表現してあるので、結果表示ウインドウ中に分類体系のカテゴリのうちの利用者が興味のある (頻繁に利用すると期待できる) 部分集合を保持して更新しているとも考えられる。利用者がこの分類先と利用者の興味を表現しているカテゴリをクリックすると、結果表示ウインドウに当該カテゴリのページを表示する。これにより利用者はブラウジング途中で、結果表示ウインドウを経由して利用者の興味のあるカテゴリの情報を参照することが可能になり、これらの移行もスムーズに行うことが出来る。

5 まとめ

本稿では、定期的にブラウザが表示している URL を監視することで、逐次的な興味抽出を行うブラウジング支援の方法を提案した。また、分類体系に NTT DIRECTORY を用いたプロトタイプの詳細を説明した。

本プロトタイプは WWW 情報を対象としているが、ローカルで管理している文書を対象とすることも可能である。この場合は CA サーバは必要がなくなるので、実現はより容易になる。このほかに本ブラウジング支援の利用方法として、個人適応型検索や情報推薦システムのブラウジング支援などが考えられる。今後の課題として、本方法では、利用者のブラウジングを発散させてしまう可能性があるため、これを適切に誘導する方式を考察する予定である。

参考文献

- [1] 沢井他 : WebMap:Hyperbolic Tree を利用した WWW ブラウジング支援, 第 58 回情報学大会 vol.3, pp.67-68, (1999).
- [2] 平嶋他 : 文脈情報を用いたハイパーテキストのブラウジング支援, 人工知能学会誌 Vol.12 No.5, pp.724-733, (1997).
- [3] 村本他 : 階層型知識体系を用いた WWW 情報の自動カテゴリ推定方法, 第 56 回情報学大会 vol.3, pp.205-206, (1998).
- [4] 村本他 : 階層型知識体系を用いた自動情報推薦方式, マルチメディア, 分散, 協調とモバイル (DICOMO) シンポジウム, pp.219-226, (1998).
- [5] <http://navi.ocn.ne.jp/>.