

個人履歴情報を用いた Web 情報検索方式の提案

高橋 英史朗[†] 辻 秀一[‡]

東海大学大学院工学研究科[†] 東海大学電子情報学部[‡]

1 はじめに

近年、インターネットの急速な普及により利用者が増加するにつれ情報量も多くなってきた。そのため、個人の情報の要求がしっかりしていない場合の検索では要求に関係のない情報も膨大に検索結果として獲得してしまう。従来の検索エンジンの場合、個人要求から独自に具体化、類義語の作成などでより検索精度の高い検索エンジンをめざすが個人要求から離れてしまう場合があった。

本稿では、個人利用履歴を用いてより個人要求に適応した検索方式を提案する。

2 従来の検索方式

従来の検索の仕組みは、検索キーワードを入力し検索エンジンで検索処理が行われ検索結果を返すというものである。ここでは検索処理の部分が検索エンジンによりそれぞれ異なる特徴をもっている。Google¹、Infoseek²などの検索エンジンはロボット型の全文検索方式をとっている。

また、MetaCrawler³のような複数の検索エンジンに並列に問い合わせを行い、その結果を統合して提示してくれるメタ検索エンジンも存在する[1]。しかしこれらは全世界で8億ものWWWページが存在すると推定される膨大な情報の中からより多くの検索結果を提示してくるため個人の求める情報を獲得できないことが多い[2]。そのためにより個人の要求に近いもので検索をかけようと考えたのが類似検索である。

これは検索キーワードの類似語をいくつかあげて複数の語句による検索をすることでより個人

の要求する情報を獲得しやすくするものである。しかしこの検索方式も類似語の提示によりまったく異なる情報へと広がる可能性もあり全文検索などと同様に精度が低い部分がある。

そこで本稿では個人の履歴情報を用いた検索方法を提案しより個人の求める情報を獲得することを目指す。

3 個人利用履歴を用いた検索方式

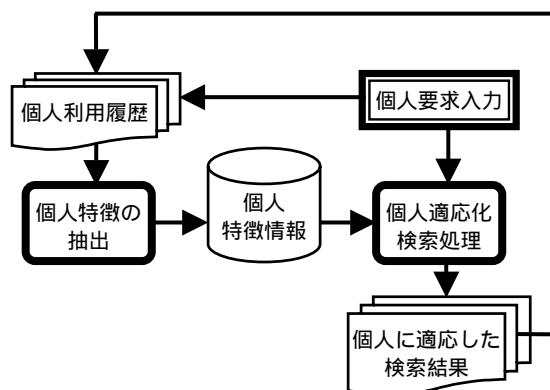


図1 提案の検索処理

図1に個人利用履歴を用いた検索方式の概要を示した。従来の検索法との大きな違いは個人に特化した検索法であり個人利用履歴から抽出した個人特徴情報を用いて検索処理を行うことにより、個人に適応した検索結果を獲得ができる。

3.1 個人特徴の抽出

個人特徴とは個人利用履歴から得られる個人特有の情報であり個人利用履歴が多いほどその特徴が明確に見えてくるものである。

従来の研究ではこの個人特徴を興味情報と呼び文章中のキーワードの出現頻度の累積計算により高い結果を示したものを興味情報としたものがある[3]。本提案では頻度に加えキーワードの関連付けによるクラス分け抽出法を提案する。

Proposal of the Web information searching method using private record information.

[†] Eishiro Takahashi [‡] Hidekazu Tuji

[†] Graduate School of Engineering, Tokai University

[‡] School of Information Technology and Electronics, Tokai University

¹ <http://www.google.co.jp/>

² <http://www.infoseek.co.jp/>

³ <http://www.metacrawler.com/>

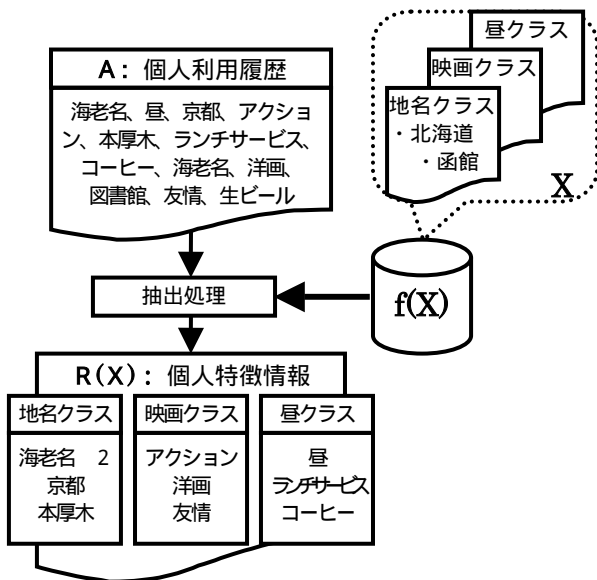


図2 提案の抽出法

図2で示したようにクラス分けし抽出したものを個人特徴情報と呼ぶ。抽出計算式は A:個人利用履歴, X:分野クラス, f(X):分野クラス内特徴情報の集合, R(X):個人特徴情報とすると以下になる。

$$R(X) = (A \cap f(X)) \quad (1)$$

以上の抽出処理はオフラインで個人の管理する中で実行するので個人情報個人で管理することになる。次節でこの個人特徴情報を利用した個人適応化検索処理を説明する。

3.2 個人適応化検索処理

この処理は前節で獲得した個人特徴情報を元にユーザの個人要求から検索式を作成する。ここでは検索式作成法について説明する。

例えばユーザ個人要求が「神奈川で昼に映画を見たい」であったとする。この場合、特徴情報は「神奈川」「昼」「映画」である。このときのユーザ個人要求をW(X)とする。

そこで、本提案では前節で個人利用履歴から抽出した個人特徴情報 R(X)を利用した検索式作成を行う。以下の検索式作成法の説明を図3に示す。ユーザ個人要求W(X)と前節の R(X)から分野クラス X の検索キーワード集合を作成することができる。以上を式で表すと以下になる。S(X):分野クラス X の検索キーワード集合とする。

$$S(X) = R(X) \cap W(X) \quad (2)$$

特徴情報となる分野クラス X の数の S(X)だけ AND 計算し、これらを全て AND 計算したもの

が検索式となる。この節での場合は S(地名)=(海老名 OR 本厚木)などを AND 計算した場合以下の検索式になる。

$$\text{「(海老名 OR 本厚木) AND (昼 OR ランチサービス) AND (洋画 OR アクション)」} \quad (3)$$

この検索式は個人要求に個人特徴の含まれたものになる。よってユーザの個人要求の「神奈川で昼に映画を見たい」に対しより精度の高い情報をユーザに提供できる。

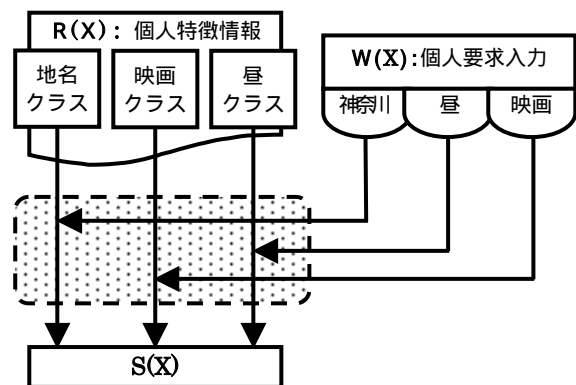


図3 検索式作成法

4 まとめ

本提案の特徴としては検索式作成までの工程はオフライン処理なので個人情報を外部に漏らすことなく個人で管理することができるので従来のようにネット上での個人情報が漏れるという心配もなくなることである。

また今後、個人情報は個人で管理する時代が来ると考えられ高セキュリティの記憶媒体として IC カードなどに管理することで安全に管理し個人利用履歴を利用する機会も増えることであろう。

参考文献

- [1] 北村泰彦：「インターネットエージェント 知的情報検索と自動情報配信」 bit, February, Vol.31, No2 pp.7-14, (2001)
- [2] 福島週一：「WWW サーチャエンジン」, 人文学と情報処理, No.28, pp.30-43, (2000)
- [3] 砂山渡・大澤幸生・谷内田正彦：「ユーザの興味の構造を用いて関連検索キーを提示する検索支援インターフェイス」, 人工知能学会誌, 15巻6号, (2000/11)