2P-1

ユーザのブラウザ操作情報を利用した検索語推薦システム

佐々木 直哉 †　　中山 泰一 †
† 電気通信大学情報工学科

1　はじめに

インターネット上に Web ページは無数にあり，その中から自分の目的にあった情報を効率的に得るための様々な方法が提案されている．一般に用いられているものとして，検索エンジンに自分の知りたい情報と関連する語を入力し，調べるという方法がある．

検索エンジンは，入力された語を含むページについて全文検索を行い，その結果を表示する．したがって，適切な語を入力することが，目的の情報が存在するページを見つける上で必要となる．

中島ら [1] は検索の熟練度と目的についての知識量が，ユーザの検索行動やその成果に影響を及ぼすことを指摘している．特に，目的の情報に関する領域知識がパフォーマンスに与える影響は検索経験より大きいことを示している．

中島らによれば，検索熟練者は検索結果を詳しく見るより再検索を繰り返し，閲覧したページから抽出した語を検索に用いる傾向がある．また，検索熟練者はページ内検索などの機能を用いて，特定のキーワードを探し出したりハイライトすることがわかっている．一方，領域知識のあるユーザは，1 語が長く具体的な単語を入力する傾向があるとしている．

本研究では，ユーザのブラウザ操作情報から目的を判断し，適した検索語を推薦し再検索を促すことで，初心者と熟練者間，領域知識のある者とない者の差を埋めるシステムを提案する．

2　関連研究

情報システムでは，ユーザの目的を判断するために様々な情報が用いられている．

入力された検索語をユーザの情報として利用している例としては，堀ら [2] の研究がある．閲覧履歴中の単語を自己組織化マップを用いてクラスタリングを行い，それに検索語を当てはめ拡張検索語を得ている．

各ページの閲覧時間をユーザの短期的なプロファイルに利用している例としては，杉山ら [3] の研究がある．杉山らは各ユーザに適応させた検索結果を提示するため，ユーザの閲覧履歴を用いている．このとき単語数によって正規化された閲覧時間が，一定以下のページの内容を反映しないようにしている．

しかし，入力された検索語は必ずしもユーザの目的を表しているとは限らず，閲覧時間はページ中のコンテンツを探しているだけで読んでいない場合などを考慮していない．

杉田ら [4] は，ブラウザのスクロールおよびタブ操作とユーザの興味についてユーザ間に共通する特徴を確認している．本研究では，ユーザのスクロール操作に関する特徴を実際の情報推薦システムに応用し，評価することにした．

3　システムの概要

本システムは，ユーザの閲覧したページを解析しユーザの目的を判断する．そして，検索上位のページから目的に合った語を抽出し，ユーザに推薦する．

システム構成は図 1 の通りである．次のように動作する．

1. ユーザは Google 等の検索エンジンに検索語を入力し，検索結果を得る．
2. ブラウザは入力された検索語を解析エージェントに送信し，解析エージェントは検索結果上位のページを獲得する．
3. ユーザは得られた検索結果から目的に合致するページを選び閲覧する．
4. ブラウザは，閲覧ページのテキストやスクロール速度等のブラウザ操作情報を解析エージェントに送信する．
5. 解析エージェントは，ブラウザから受信した操作情報を解析し，ユーザの目的を判断する．
6. 解析エージェントは，ユーザの目的に合わせて検索結果上位のページから推薦する語を抽出し，ブラウザに送信する．
7. ブラウザは，受信した検索語をユーザに提示すると同時に，ページ中の該当する語をハイライトする．

Query suggest system using user's browsing behavior
†Naoya SASAKI　†Yasuichi NAKAYAMA
†Department of Computer Science, The University of Electro-Communications

Copyright ©2013 Information Processing Society of Japan.
All Rights Reserved.

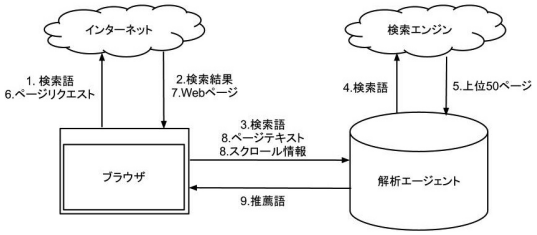

図 1: システム構成図

4 ユーザの目的の判断

ユーザの目的の判断には，ユーザが検索中に閲覧したページを利用する．スクロール情報を考慮することで，ユーザがそのページにどれほど興味を持ったか判断する．

4.1 特徴ベクトルの生成

TF-IDF 法により各ページの単語について重み付けを行う．各変数を次のように定義する．

N 検索後にアクセスしたページ数

p_i ページ i

$f_{t,d}$ ページ d での単語 t の出現回数

M ページの d 中の単語の総数

df_t 単語 t を含むページ数

このとき，TF-IDF 値 $w_{t,d}$ は

$$w_{t,d} = tf_{t,d} \times idf_t$$
$$tf_{t,d} = \sum_{t=1}^{M} f_{t,d}$$
$$idf_t = \frac{logN}{df_t}$$

となる．

4.2 スクロール情報の利用

杉田らはユーザの各スクロール操作の特徴を次のように説明している．

ゆっくりとしたスクロール ページの内容に注目

逆スクロール ページ上方の内容の再確認

上下の繰り返し 目的のコンテンツの探索

ユーザが一定以下のスピード k でスクロールして場合，ユーザの興味にあったページとする．k の値は各ユーザごとに事前に調査を行い設定する定数である．さらに逆スクロールしていた場合，ユーザが必要と感じた情報を読み返していると判断し重みをつける．

5 推薦語の抽出

推薦する検索語は，検索結果の上位 50 ページより取得する．上位 50 ページにそれぞれついて，TF-IDF 法により特徴ベクトルを作成する．Rocchio の式により適合性フィードバックを行う．

$$Q_{i+1} = Q_i + \frac{1}{N^+}\sum_{j=1}^{N^+} D_j^+ - \frac{1}{N^-}\sum_{j=1}^{N^-} D_j^-$$

Q_i は i 回目の検索質問を表し，Q_{i+1} は修正された検索質問を表す．D_j^+ および D_j^- はそれぞれ N^+，N^- 個ある適合文書と不適合文書を表す．上位 50 ページの内，ユーザの特徴ベクトルとの類似度が高いものを適合文書とし，それ以外を不適合文書とする．

得られた検索結果ベクトルの要素の内，重みが上位 30 位以内のものを推薦語とする．

6 まとめ

本研究では，ユーザのブラウザ操作情報を利用した検索語推薦システムを提案した．これにより正確にユーザの目的を把握し，領域知識を補うような語を推薦することで，ユーザの検索パフォーマンスを向上させた．

今後は，実際に検索初心者を対象に実験を行い，システムを評価する．

参考文献

[1] 中島悠, 土方嘉徳, 西田正吾：検索経験と領域知識の WWW 情報検索行動に与える影響, ヒューマンインタフェース学会論文誌, Vol. 7, No. 2, pp. 309–319 (2005).

[2] 堀幸雄, 今井慈朗, 中山堯：ユーザの Web 閲覧履歴を用いた検索支援システム (< 特集 > 第 15 回 (2007 年度) 年次大会 (研究報告会&総会)), 情報知識学会誌, Vol. 17, No. 2, pp. 95–100 (2007).

[3] 杉山一成, 波多野賢治, 吉川正俊, 植村俊亮：ユーザからの負担なく構築したプロファイルに基づく適応的 Web 情報検索 (コンテンツ技術,Web 情報システム), 電子情報通信学会論文誌. D-I, 情報・システム, I-情報処理, Vol. 87, No. 11, pp. 975–990 (2004).

[4] 杉田賢治, 福原知宏, 増田英孝, 山田剛一：ブラウザ操作ログ収集ツールを用いた Web ページ閲覧行動分析, 情報処理学会研究報告. データベース・システム研究会報告, Vol. 2011, No. 1, pp. 1–9 (2011).

Copyright ©2013 Information Processing Society of Japan.
All Rights Reserved.