

Web 検索結果のキーワードマップにおける 滑らかな重み付け操作による再ランキング

吉田 大我[†] 中村 聡史[†] 田中 克己[†]

[†]京都大学大学院情報学研究科社会情報学専攻 〒606-8501 京都市左京区吉田本町

[†]E-Mail: {yoshida, nakamura, tanaka}@dl.kuis.kyoto-u.ac.jp

1 はじめに

検索エンジンを用いる際、ユーザは検索に用いるクエリを入力する必要があるが、目的のページを発見可能なクエリを思いつづことが困難である場合もある。例えば、ユーザが京都の東西線について知りたいと思い、クエリとして「東西線」と入力したとする。すると、表示される結果は東京の東西線に関するページが大半を占め、京都の東西線に関するページはほとんど含まれない。これは東西線という語が曖昧性を持つことが原因である。曖昧性を回避するためには「京都」や「JR」などの曖昧性を解消するキーワードをクエリに追加する必要がある。このように、一度検索した結果を反映して、目的の話題に関係するページが表示されるように検索キーワードを修正しながらクエリを組み立てることは多くのユーザにとって容易なことではない。

ユーザが検索結果中にどのような話題が含まれているかを一覧的に閲覧し、再検索することを支援するための手法として松生ら[1]による手法があるが、検索結果ページを再ランキングして目的のページを発見するためには利用できない。

また、ユーザの操作を検索結果に反映させるものとして、Yahoo! Mindset[2]などが挙げられる。このシステムは検索結果中のページについて買い物 (shopping) と情報収集 (researching) のどちらに適したページなのかスコアリングを行う。ユーザは、どちらに適したページを検索するかをスライダーを用いて操作することでシステムが検索結果を動的に再ランキングする。しかし、再ランキングに利用する評価軸が限られているという問題点があった。

本研究では、検索結果がどのような話題を含んでいるかを一覧的に見ることを可能にするため、閲覧ページ中や検索結果中に出現する重要な語を話題語として定義し、キーワードマップと名付けた手法を用いて可視化した。キーワードの可視化により、ユーザは求める情報と検索結果が合致しているかどうかを判断し、合致していないならばどのような語を重視することで目的の情報を含むページを見つけることができるかを知ることができる。

また、提案したシステムでは、複雑なクエリを生成する代わりにマウスによってグラフ上に表示された話題語の関係を操作することにより、検索結果の再ランキングを行うことができる。

Relanking Web Search Results Using Term Weighting Operations with Keyword Map

[†]Taiga Yoshida, Satoshi Nakamura and Katsumi Tanaka
Department of Social Informatics, Graduate School of Informatics, Kyoto University Yoshida-Honmachi, Sakyo, Kyoto 606-8501 Japan

2 提案システムの概要

2.1 話題語とは

本研究では、複数の話題が混在する検索結果に対し、ユーザが検索結果集合における話題傾向を発見することの支援を目的としている。検索結果ページのスニペット中に出現する語の中には、検索結果中のある話題と強い相関性を示す語がある。例えば、「東西線」というクエリにおける「京都」や「メトロ」といった語がそれにあたる。そのような語を、話題を構成する語であることから話題語と定義し、抽出した話題語を図 1 のようなキーワードマップと名付けた手法により可視化する。

キーワードマップ上に表示する話題語として、外在語および内在語という二種類の語を定義した。外在語とは、ユーザが入力したクエリによる検索結果リスト中から抽出した話題語である。そして内在語とは、ユーザが検索結果中のページに興味を示し、ハイパーリンクをクリックしてそのページの閲覧を開始した際、ブラウザ領域に表示されたページ中から抽出した話題語である。

外在語および内在語は、閲覧ページ中および検索結果ページ中において出現頻度の高いものからそれぞれ20語および10語が選択される。ユーザはこれらの語の重み付けを変更することによって閲覧する検索結果を再ランキングすることができる。

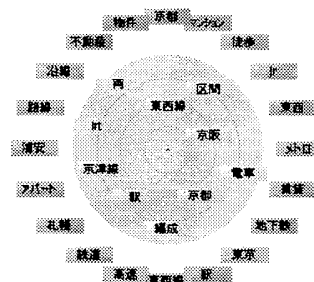


図 1 「東西線」検索時のキーワードマップ

2.2 可視化手法

抽出した外在語および内在語は、単純に並べて表示するだけではユーザとのインタラクションにより重み付けを変更することが困難である。そこで本研究では、外在語および内在語を、二次元平面上に可視化するキーワードマップにより、マウス操作による重み付けを可能とした。本稿では、キーワードマップ上において語を表すものをノードと呼ぶ。

キーワードマップには、外在語および内在語がそれぞれノードとして描画され、初期状態では外在語は円の外側に、内在語は円の内側に配置される。

2.3 ユーザ操作による検索結果の再ランキング

提案するシステムでは、複数のキーワード候補をユーザが操作し、ゆるやかな重み付けにより再ランキングした結果を閲覧することができる。外在語および内在語はそれぞれ固有のスコアをもつ。円の半径を r 、円の中心とノード間の距離を d とすると、内在語のスコアは $r-d$ で計算され、外在語のスコアは円の外側にある場合は0、円の内側にある場合は $(r-d) \times 2$ で計算される。つまり、円の中心に近い語ほど高いスコアを持ち、円の内側では内在語よりも外在語の方が高いスコアを与えられる。

検索結果リストに含まれる各ページも、ノードと同様に個別のスコアを持つ。ページのスコアは、そのページの検索結果におけるタイトルおよびスニペットに含まれる語のスコアの合計によって計算される。システムは各ページのスコアを計算した後、スコアが高いものから順に並べて検索結果を表示する。

このようにして、ユーザはキーワードマップにおいて円の中心とノードとの間の距離を変えることにより、検索結果の再ランキングを行うことが可能である。再ランキング後のページは検索結果リスト中に表示され、ユーザはリンクをクリックすることで各ページを閲覧することができる。ユーザが新しいページを閲覧するたびにシステムが新しい内在語を抽出することで閲覧中のページと内容の似た検索結果ページが上位に再ランキングされ、ユーザは関連したページを連続して閲覧することができる。

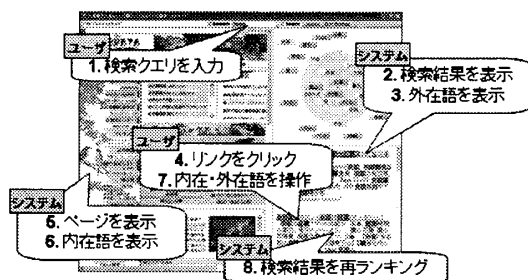


図2 システムイメージ

3 システムの動作イメージ

ユーザがシステムを用いて検索する流れは、以下のようになる。

1. ユーザがクエリを入力
2. システムは検索結果ページのリストおよびキーワードマップ上の外在語を表示
3. ユーザは重視する語についてノードを操作
4. システムはキーワードマップのノード配置に応じて検索結果を再ランキング
5. ユーザは検索結果リストから興味のあるページを閲覧
6. システムは閲覧ページ中から内在語を抽出し、それらをキーワードマップに表示
7. ユーザは目的のページが見つかるまで外在語および内在語の操作を繰り返す。

実装したシステムのイメージを図2に示す。システムでは、キーワードマップと共に検索結果リストが表示され、再ランキングされた検索結果における、各ページのタイトルやスニペット、URLなどが表示される。

4 システムの評価

本システムを利用することにより、検索結果および閲覧ページに対し、高い精度で外在語および内在語を提示することができた。

システムによって提示される語は検索結果や閲覧ページ中において出現頻度の高い語である。そのため、検索するクエリについて詳しく知らないユーザはキーワードマップ上の語を一覧することにより、一件ずつ検索結果ページを見なくてもキーワードの意味を把握することができる。と考えられる。

また、「東西線」などの多義語や「yahoo」のように多数の話題を持つキーワードでは、話題の傾向を絞り込むために有用な語が外在語・内在語として抽出され、興味のある話題を優先的に閲覧することが可能となった。

最初から閲覧したいページが決まっている場合や、検索結果中にわずかしかが含まれていない話題を扱ったページを探す場合には本システムは向いていないと考えられる。

5 まとめと今後の課題

本研究では、検索結果中のページをキーワードマップ上に配置されたノードを移動することによりキーワードの重み付けを行い、検索結果を再ランキングすることによりユーザの検索を支援するシステムを提案した。

本システムを用いるとユーザは閲覧ページや検索結果ページにおける話題傾向を一覧的に把握することができ、ユーザはキーワードマップによってどのような話題がページ中に含まれているかが分かる。しかし、ユーザはどの話題が一般的な話題で、どの話題がマイナーな話題なのかといった情報をキーワードマップから得ることができない。そこで、キーワードマップ上のノードの大きさを変えるなどして、話題の一般性をユーザに提示できるようにすることが今後の課題である。

また、システムの利用時にキーワードマップ上に操作を行いたい語のノードが配置されていないことがあるため、ユーザにより自由にキーワードマップ上に語の追加を行えるための手法を考案する必要がある。今後は任意の語について重みを変更し、ユーザが求めるページの検索に利用することができるよう、システムを改良していく予定である。

謝辞 本研究の一部は、グローバルCOE 拠点形成プログラム「知識循環社会のための情報学教育研究拠点」、文科省研究委託事業「知的資産の電子的な保存・活用を支援するソフトウェア技術基盤の構築」、科研費：計画研究「情報爆発時代に対応するコンテンツ融合と操作環境融合に関する研究」(課題番号18049041)、文部科学省科学研究費補助金若手研究(B) No. 18700129 によるものです。ここに記して謝意を表すものとします。

参考文献

- [1] 松生泰典, 是津耕司, 小山聡, 田中克己: “検索結果の概要を表すキーワード式生成による質問修正支援”, 電子情報通信学会第16回データ工学ワークショップ (DEWS2005) 1C-i9 (2005)
- [2] “Yahoo! mindset. <http://mindset.research.yahoo.com/>”