

キーワード利用傾向に基づく発想検索支援

5K-4

織田 充^{1,2}南 俊朗³¹(株)富士通研究所, ²(財)九州システム情報技術研究所, ³九州大学附属図書館研究開発室¹oda@flab.fujitsu.co.jp ²oda@k-isit.or.jp ³minami@lib.kyushu-u.ac.jp

1 はじめに - 発想的キーワード探索

考え, およびシステムの概略を述べる。

現在インターネット, イン트라ネット等のネットワークを通じて提供される Web 文書は, 既に膨大な量となりさらに増加し続けている。膨大な量の Web 文書の中から有益な文書を探しだすために, キーワード検索エンジンが多数提供されている。そこでの問題の 1 つは, 適切なキーワードを発想することの困難さである。

検索時に入力されるキーワードは, どのような文書を参照するかという検索者の興味, 検索意図を表す端的な表現であると共に, 多量な文書群の中から参照する文書を絞り込む検索条件の表現となる。この観点から, 入力キーワードの決定過程には, 発想支援システム [1] における発想過程と同様, 検索者の興味, 検索意図の発想過程という側面と, 同時に, それらを満足する文書を検索するに十分な検索条件の発想過程という側面がある。適切なキーワードを思いつくことの困難さは, これら 2 発想過程の困難さに起因する。

入力キーワードを自ら思いつくことが困難であっても, 示された候補の中から自分の要求に最も近いと思われるキーワードを選択することは比較的容易である。このことを利用し, キーワード推薦システムが開発されている。しかしそれらの多くは, 文書中出现する単語からその出現頻度を用いて重要単語を決定し, 推薦キーワードとして用いる。この点で検索結果の絞り込みに適し, 後者の発想過程の支援に向けたキーワード推薦方式である。

これに対して, 検索に用いられたキーワードは, 提供者に比べより検索者に類似な立場にある主体, すなわち検索者の興味や検索意図を何らかの意味で反映するコンパクトな表現である。この点に注目し, 検索者の検索履歴を基に検索者のキーワード発想過程を支援する, 発想的キーワード推薦システム SASS (Searching Assistant with Social Selection) を提案した [2,3]。以下, その基本的な

2 検索履歴に基づくキーワード関連性

SASS システムは, 検索者が入力したキーワード列に関連したキーワードを推薦する。検索者の興味に基づいて変化するであろう「関連性」として何が適切かを一般的に定めるのは困難である。SASS では複数の「関連性」選択肢を検索者に提供し, 選択された関連性を元に検索履歴からキーワードを推薦している。本稿ではその「関連性」の一つである「検索関連性」に焦点を当て紹介する。絞り込み検索においては, キーワードを補足や, 修正が行なわれるが, その結果出された最終的な検索式を以下では単に「検索」と呼ぶ。

検索は検索者の興味の端的な表現と見なすことができる。従って, 同じキーワードを含む二つの検索があれば, それが表す二つの興味も類似していると推定される。そこで,

仮定 検索者が入力したキーワード k を含む他の検索 r がある時, その r に含まれる他のキーワード k' を推薦することは妥当である。

を前提とすることができる。しかし, 実際に検索履歴を解析した結果, 同じキーワードを含む検索はそれほど多くなく, 興味が多様なそれぞれの検索者に対して推薦するに十分な量のキーワードを見つけられないことが分かった。この問題を解決するために, 仮定を推移的に拡張し, それに基づいて下記「推移関連性」を採用している。

推移関連性: キーワード k に対し, i) k を含む検索 r 上の任意のキーワード k' は k に関連する。また, ii) k と k' が関連している時, k' を含む検索 r' 上の任意のキーワードは k に関連する。

3 SASS システム

以上の準備の基で, 推移関連性を用いた SASS でのキーワード推薦法を示す。

定義: 検索 最終的な検索式に含まれるキーワード集合を「検索」と呼ぶ。ただし時点, 検索者による検索の違いが区別できるように, 各集合に識別子が付加されているものとする。

An Idea-Creating Searching Assistant System based on a Tendency of Keyword Choices

Mitsuru Oda and Toshiro Minami

Fujitsu Laboratories Ltd.

Institute of Systems & Information Technologies/KYUSHU

Kyushu University

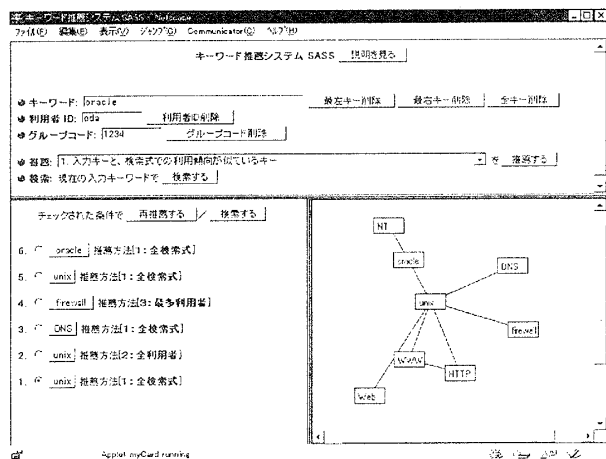


図 1: SASS システムの画面

定義：関連キーワード $o(r, a, b)$ は「検索 r 中にキーワード a, b が出現する」関係を表す。また、 $o'(a, b) = \exists x.o(x, a, b)$ とし、 o' の推移的閉包を o^* とする。 $o^*(a, k)$ を満たす k を「キーワード a に対する関連キーワード」と呼ぶ。

検索履歴の解析から得られた知見からキーワード間の関連性を推移的に拡張したが、逆に複数の検索を介して初めて関連性を持つキーワードと、仮定にあるように直接関連性を持つキーワードとの区別が付かない。そこで関連性の程度を表す評価値として、キーワード間の関連度を導入した [3]。基本的には、推移関連性を与え方が多いほど、全体としてのキーワード間関連度は高く、また、異なる与え方では、関与する直接関連性の個数が少ないほど、その評価値が高くなるように設定される。そして各キーワードに対して、キーワードを次元とし、キーワードに対する関連度を値とするベクトルを、その特徴ベクトルとして与える。それら特徴ベクトル間の内積相関値を、キーワード間の類似度として用い、その値のより高いキーワードをより優先的に推薦キーワードとして用いる。

SASS システムは、上記認識に基づきユーザの発想的キーワード探索を支援する。図 1 に SASS の画面例を示す。ユーザは、検索エンジンの場合と同様に、まず、自分の目的とする Web 文書の特徴づけるキーワードを指定する (画面上部)。システムは候補毎に、入力されたキーワードに対する他のキーワードの類似度を求め、入力キーワードと推薦キーワード間の類似度を距離に反映させた平面配置図を用い、提示する。(画面右下部分)。ここでの配置方法は多次元尺度法を用い求めている。このように推薦キーワードを平面配置することで、単純に類似度で 1 次元的にソートした提示方法に比べ、推薦キーワード間の関係、検索者の検索傾向が概観できその効果が高い。また過去の SASS の

利用履歴も表示され、いつでも元の状態に戻ることができる (画面左下部分)。

本稿においては、キーワード間の関連性として、検索に注目した関連性を中心に説明した。しかし SASS 自体では、検索履歴から特定の検索を実行した検索者、またその所属グループの情報が、抽出できる場合、その他に検索者、検索者の所属グループが利用したキーワード群に注目したキーワード推薦も行っている。

4 まとめ

SASS システムは、文書に出現する単語から推薦キーワードを選択する提供者側の意図に基づくキーワード推薦ではなく、利用者の検索履歴を基にし、利用者の興味を反映した発想的キーワード推薦、検索者らの検索意図の傾向を図的に概観させることにより、検索者の興味、検索意図の発想支援を行うことが、大きな特徴である。

この方向での文書選択の支援を更に進めるために次の課題の探求が重要である。(1) 検索履歴データより利用者にとり有効なキーワード群を抽出する技術が、本方式の有効性のキーとなる。効果の高い関連度抽出技術を更に追求する必要がある。(2) 本システムでは、検索者の興味、検索意図の発想を支援することに注目し、検索式におけるキーワードの出現関係を基に推薦キーワードを決定している。このため、検索者が利用したキーワードが出現する文書が存在しない場合に該当文書が無い可能性がある。検索の目的は、検索者にとって有益な文書の発見である。よりきめ細かな推薦機能をユーザに提示するために、検索エンジンとキーワード推薦機能の効果的な融合方法の研究が必要である。

参考文献

- [1] 國藤 進：発想支援システムの研究開発動向とその課題，人工知能学会誌 Vol.8 No.5, pp.552-559, 1993.
- [2] 織田充，南俊朗，有馬淳：検索履歴を用いたキーワード推薦エージェント，電子情報通信学会人工知能と知識処理研究会，AI98-58, pp.33-40, 12月, 1998.
- [3] 織田充，南俊朗，有馬淳：SASS:Web 文書の発想的検索支援システム，情報処理学会第 57 回全国大会 (3), pp.173-174, 10月, 1998.
- [4] 織田 充，南 俊朗：参照履歴を用いた Web 文書推薦方法の提案，情報処理学会第 56 回全国大会，pp.203, 3月, 1998.