

Web 検索ログに基づく関連度可視化システムによる情報ニーズの抽出

小澤 泰輔[†] 杉崎 正之[‡] 望月 崇由[‡] 徳永 幸生[†] 杉山 精^{*}

芝浦工業大学 工学部[†] NTT レゾナント株式会社[‡] 東京工芸大学^{*}

1. はじめに

Web 上の情報の多様化・多量化に伴い、インターネット利用者は大規模な情報の中から欲しい情報を得るために、Web 検索システムに検索語を入力し、試行錯誤しながら求める情報に近づいていく。この Web 検索システムの検索ログデータには、検索者の情報要求の生の声が潜んでいると考えられる。そこで、検索ログデータを分析することにより、検索語間の関連度を抽出し、検索語同士の背景に潜む構造や相互の関係から、情報取得の目的を探る研究がなされている^[1]。

柳らは、検索ログデータから時間間隔関連度と特徴ベクトルによる \cos 関連度を算出した。これら 2 種類の関連度を、様々な仮説の下で、相互に利用しながら試行錯誤し、検索者の情報ニーズを抽出する試みを報告している^[2]。しかし、それぞれの仮説ごとに 2 種類の関連度を計算し表示させていたため、その試行錯誤に要する時間は極めて長くなる。特徴的な情報ニーズの抽出を効率的に行うには、2 種類の関連語を相互に参照して利用できる関連度可視化システムを構築することが求められていた。

本稿では、人間の検索行動における法則性や情報ニーズをより効率的に抽出することを目的に、2 種類の関連語を組み合わせて利用できる関連度可視化システムを構築した。さらに、その有用性について検討したので、これを報告する。

2. 検索行動の分析と関連度の算出

2.1 利用者の検索行動の分析

図 1 は、Web 検索システムを利用した検索者の検索行動のモデルである^[1]。検索者は 3 つの STEP を踏み、求める情報を得られた、あるいは得られないと判断したとき、一連の検索行動を終了する。

STEP1 - STEP2 間の繰り返しは、検索語の再検討である。このような検索の試行錯誤が、比較的短い時間間隔で、頻りに繰り返されている^[1]。検索ログデータから検索の使用時間間隔を算出し、これを検索者の検索行動と照らし合わせる。これにより、検索語と検索語の間に STEP1 - STEP2 間の繰り返し、すなわち検索語の再検討による検索の試行錯誤が、どれだけ行われているかを分析する。

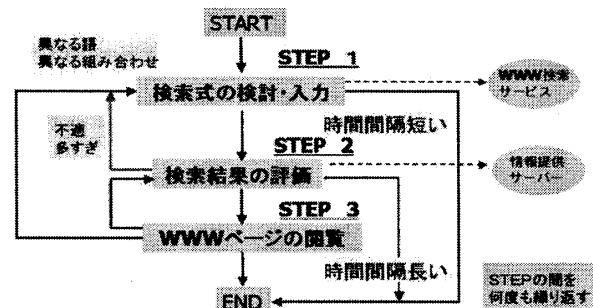


図 1. 検索者の検索行動のモデル

2.2 関連度の算出

検索の使用時間間隔の分布を基に定義した assoc 関数^[2] (図 2) を用いて、ある期間の検索ログデータを基に、任意の検索語のペアに対して 2 種類の関連度を算出する。

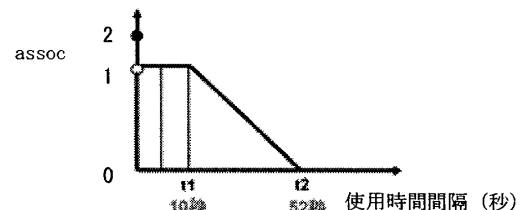


図 2. assoc 関数

① 時間間隔関連度

2 つの検索語間の使用時間間隔から、 assoc 関数を用いて関連度を数値化する。これを全検索者に対して行い、その総和を時間間隔関連度とする。検索者がある目的の情報を得るために、検索語のペアがどの程度一緒に検索されているかの指標となる。すなわち検索語の追加候補としての関連の強さである。

② 特徴ベクトルによる \cos 関連度

ある検索語と、他の全検索語との時間間隔関連度を要素とした特徴ベクトルを用いて、検索語間の距離を \cosine により算出する。これは検索のされ方がどの程度似ているかの指標となる。すなわち検索語の置換候補としての関連の強さである。

3. 関連度可視化システムと情報ニーズの抽出

3.1 関連度可視化システム

柳らが 2 種類の関連度を用いて情報ニーズを抽出した過程^[2]を追い、以下のように手順をまとめた。

- (1) ある検索語と \cos 関連度の高い単語を挙げる
- (2) それぞれの単語の時間間隔関連度を並べる
- (3) 単語ごとの傾向や特徴を比較する

柳らは、ある検索語と関連度の高い検索語を同心円状に並べる関連度可視化システムを用いて、情報ニーズを抽出した。(1)は、このシステムで \cos 関連度に

Extracting Information Needs
using Visualization System of Relation Levels
based on a WWW Search Log

[†] Taisuke OZAWA (t06024@shibaura-it.ac.jp)

[‡] Masayuki SUGIZAKI (sugizaki@nttr.co.jp)

[‡] Takayoshi MOCHIZUKI (mochizuki@nttr.co.jp)

[†] Yukio TOKUNAGA (tokunaga@shibaura-it.ac.jp)

^{*} Kiyoshi SUGIYAMA

[†] College of Engineering Shibaura Institute of Technology

[‡] NTT Resonant Inc.

^{*} Tokyo Polytechnic University^{*}

