

話題空間の構成に基づく探索的検索過程の可視化に関する研究*

枝 隼也(学籍番号 201021735)

研究指導教員:佐藤 哲司

1. はじめに

Webでの情報収集では、情報要求が曖昧な場合や調査対象が多岐に渡る場合も多い。このため、様々なページを閲覧し情報間の関係性等を考えながら検索と閲覧を繰り返す、探索的検索の重要性が高まっている。探索的検索では、ユーザが過去に閲覧したページの履歴を保存した閲覧履歴を解析することが、ユーザの探索行動を理解する事に重要な手段となっている。しかし、その保存方法はページタイトルとURL等のリストを時系列順に蓄積していくものがほとんどで、訪問したページに記述された内容や、探索過程における話題の遷移を知るためには新たな履歴の保存と解析手法が必要とされていた。また、調査範囲が広い場合や調査対象が複雑な場合は、複数人で分担して行う協調探索が有効とされており、探索済みの範囲を他者に伝える、あるいは他者の結果を的確に把握する手法も必要とされてきた。

2. 話題空間の構成に基づく探索的検索過程の可視化手法の提案

本研究では、探索範囲となる話題全体をキーワードの意味的關係を距離とした二次元空間に展開する「探索空間マップ」を作成し、そのマップ上にユーザが閲覧したページをマッピングしていく、可視化手法を提案する。提案法の流れを図1に示す。提案法は、その処理を大きく分けて2段階からなる。Iでは探索課題に対応する話題空間を構成する文書を解析し、SOMを用いてキーワードを意味的な關係性を距離で表した2次元の平面に表現する。IIでは作成した探索空間マップへユーザの閲覧履歴ページを探索空間マップのベクトルとキーワードベクトルのユークリッド距離を比較することで距離の近い場所へマッピングする。以下この手順に従って、各処理を説明する。



図1 提案手法の概要

2.1. 探索空間マップの作成法

探索空間マップとは探索課題内の話題に関するキーワード同士の関係をSOMの計算によって距離を関連度とする2次元のマップで表現したものである。そのため探索課題に対応する文書からキーワードの関連性を示したベクトルを作成する必要がある。そこで本手法では文書内を、適切な粒度で分割し、その部分ごとに出現するすべてのキーワードの *tfidf* 値を計算しベクトルで表す。これは、文書内で同ブロックに共起するキーワードは、関連性の高い概念を表していると考えられるためである。このように空間上に配置されたキーワードを本研究では、探索空間キーワードと呼ぶ。探索空間キーワードは、SOMによって計算された空間内の位置情報と、閲覧度の値を持つ。閲覧度とは、その探索空間キーワードに対してユーザが閲覧した度合いを示す指標である。ユーザが閲覧をしたページ内にその探索空間キーワードが存在するとき、空間内の距離と単語の出現頻度を用いた計算法によって得られるスコアを加算していく。

図2は、提案法を実装したシステムによる履歴提示の様子である。ここでは、探索空間マップを「平成22年度版環境・循環型社会・生物多様性白書」から作成し、環境問題について探索した履歴ページをマッピングしている。

2.2. 履歴ページのマッピング法

次に、作成した探索空間マップ上にユーザの探索的検

* "A study of exploratory search process visualization based on configuring topic space" by Junya EDA

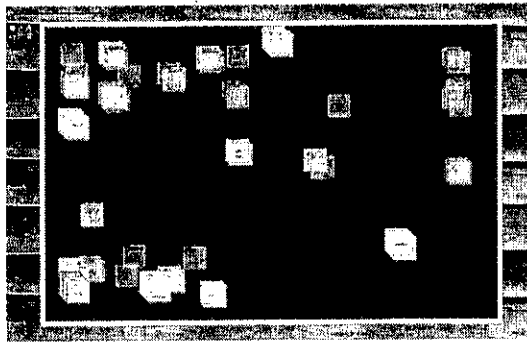


図2 提案システムにおける可視化結果の提示

索履歴をマッピングしていく。ここでは、ユーザの閲覧したページからキーワード出現数をもとにそのページのキーワードベクトルを計算し、探索空間マップのユニットそれぞれのベクトルと履歴文書のベクトルを比べ、距離の近いユニットをその文書の位置として、履歴のマッピングを行う。実際に提示する際には、各ページごとの話題がマップのどのあたりを閲覧していたのか時系列で示す必要があるため、時系列バーを操作し、その時間の閲覧済み話題空間を把握することができる。

閲覧ページ $P_i (i = 1, 2, \dots, l)$ に出現するキーワードを $w_{ij} (j = 1, 2, \dots, m)$ とし、ページ P_i が配置された SOM 空間上のクラスタを C_i とする。ここで SOM 空間は、閲覧ページを十分にマッピングだけの探索空間キーワード N 個 $w_k (k = 1, 2, \dots, N)$ で構成されている。したがって、閲覧ページに出現するキーワード w_{ij} は、SOM 空間を構成する複数のクラスタのいずれかに配置された探索空間キーワードと一致することが、高い確率で期待される。そこで、あるページ P_i を閲覧したことで得られるキーワード w_k の閲覧度 $\theta(i, k)$ を以下の式で算出する。

$$\theta(i, k) = \begin{cases} \frac{n_{ij}}{|C_i - C_k| + 1} & (w_k = w_{ij} \text{ のとき}) \\ 0 & (w_k \neq w_{ij} \text{ のとき}) \end{cases}$$

2.3. マップ充足率

提案法では、探索空間マップにユーザが閲覧したページとキーワードをマッピングしてゆくが、そのマッピングされたキーワードの全対数に対する割合を、マップの充足率と定義した。探索空間キーワードは閲覧度の値を持っているが、閲覧履歴ページのマッピングによって、探索空間キーワードの閲覧度値が0より大きくなったものすなわち一度でも出現したものをカウントし、それらのキーワードの探索空間マップキーワードにおける割合が充足率となる。

3. 評価実験

被験者 14 名の履歴データに提案法を適用し、ユーザごとの探索終了後のマップ充足率について分析を

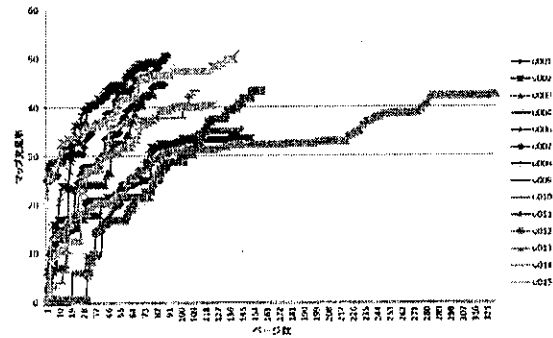


図3 マップ充足率のユーザごとの変化の過程

行った。ここで評価に用いた実際の履歴データについて説明する。CRES^[1]プロジェクトで実施した14名のデータであり、「特定の問題を想定し詳細な情報を収集することを想定し、様々な環境問題を紹介する連載記事の1号分を作成する」という課題。実験時間は45分。

3.1. 実験の結果

図3はそれぞれのユーザの探索空間マップの充足していく過程を示したものである。全ユーザで探索を行っていた時間は同じため、時間についてはより詳細なデータの解析が必要であるが、各ユーザ共に閲覧済みページ数の増加によって徐々にマップの充足率は増加していったことがわかる。ここで、グラフの傾きはどれだけ少ないページで、マップの充足率を上げることができたかを示している。傾きについてはユーザ間では、ある程度の差がみられる。マップ充足率の高かった、u007のユーザはここでも少ないページ数で、高いマップの充足率を得ている。このことは、他ユーザより多くのステップを経て、情報を見つけて行っているといえる。

4. おわりに

探索的検索課題において、そのテーマを示すような文書をもとに探索空間マップを作成し、マップ上へ履歴ページを配置することで探索的検索の支援を行うことができる手法を提案し実装した。この探索空間マップは、テーマに沿った内容が記されている文書群を用意することができれば、作成することができるため、Web探索を行う様々な場面において有効に機能すると考えられる。評価実験と分析により、提案法の各部分でWeb探索を行うことに有効な結果が得られた。

文献

- [1]CRES(Cognitive research for exploratory search). <http://cres.jp/>