

TETDMによる Exploratory Search の評価実験

The evaluation experiment which uses TETDM for exploratory search

徳永 秀和

Tokunaga Hidekazu

香川高等専門学校

Kagawa National College of Technology

Exploratory Search which attracts attention now has the following features. The search person does not have a clear search demand at the time of search. And while the search person has repeated search, he clarifies a search demand by increasing knowledge. TETDM is Challenge for Realizing Early Profits of The Japanese Society for Artificial Intelligence. This paper reports the experimental result of Exploratory Search using TETDM. TETDM can interlock two or more text data mining modules. So, it is effective in Exploratory Search.

1. はじめに

Web 上の Exploratory Search とは、検索者が検索開始時に明確な情報要求を持っておらず、Web 上の情報源の探索を行う過程で少しずつ情報要求を明確化していく探索行動である。したがって、明確な情報要求に基づいて、情報を絞り込んでいく Lookup 検索だけでは、Exploratory Search を行うことはできない。Exploratory Search を行うためには、学習 (Learn) と調査 (Investigate) が必要である [Marchionini 06]。

この、学習と調査を Web 上で行う行動の分析が三輪らにより調査されている [三輪 09]。それによると、初期のあいまいな探索要求を Wikipedia を利用して調査を行い、少し明確化した探索要求に対応したキーワードにより検索サイトで検索を行う。ここで、数個のページを閲覧し、さらに情報を絞り込む検索を行う場合と、探索要求が発散し、調査行動に移行する場合がある。現在の Web 上での Exploratory Search は、検索サイトを利用することを前提に考えることが自然であり、情報の絞込みと発散のためにクエリ拡張により学習と調査を行うことになる。

TETDM は、複数のテキストマイニングを連動されて使用できる統合環境である [砂山 13]。本論文では、Exploratory Search に TETDM を利用し、その有効性を検証する。

2. 実験

(1) 実験用モジュール

現在ダウンロードが可能な TETDM に含まれる処理モジュールとして、キーワード抽出と重要文抽出を行う展望台システム [砂山 02] と、文章の主題に関連する文をハイライトする光と影システム [西原 09] がある。展望台システムは、高頻出となる単語、またはユーザが指定する単語を文章の主題を表す観点語として、観点語と共起して文書の特徴付ける特徴語と、高頻出の背景語を抽出する。また、これらのキーワードを多く含む文を重要文として抽出する。光と影システムは、TETDM 内においては展望台システムと連動して、展望台システムで観点語とされた単語を文章の主題として、その主題を表す単語との関連度の高さに応じて各文をハイライトする。

図 1 に、これらのモジュールを利用した実験環境の画面を示

す。全部で 4 つのパネルから構成されており、左の 2 つに展望台システム、右の 2 つに光と影システムが処理モジュールとしてセットされている。組み合わせて用いられる可視化モジュールとしては、一番左のパネルが、重要文集合による要約文を表示する「テキストディスプレイ」、左から 2 番目のパネルが、キーワードの表示と観点語の入力を行える「キーワード(展望台)」が、一番右が、光と影システムの結果によりテキストをハイライト表示する「テキストディスプレイ(カラー)」が、右から 2 番目のパネルが、光と影の主題単語を選択するための「キーワード選択」となっている。



図1 TETDM の実験用モジュール

(2) 実験手順

これらのモジュールを利用して、齋藤らの実験 [齋藤 11] を参考に、環境問題をテーマとしたレポート作成のために、コンセプトマップを作成しながら情報を収集してもらう実験を行った。被験者は香川高専専攻科生 2 名とした。実験手順を以下に示す。

(1) 「環境問題」のスニペットを入力。

(a) 検索サイトで「環境問題」をキーワードに検索し、上位 100 のサイトのスニペットを 1 つの文書として TETDM に入力する。

(b) 表示される展望台の観点語と候補 (中央の縦) と関連語 (外側裾野) の中に検索キーワードの話題 (コンセプトマップの子概念) が存在すると解釈する。

(c) (b) の中またはキーワード選択の中から興味を持ったキーワードを観点語にする (展望台で横線の上に置か、キーワード選択でチェックボックスをクリックする)。展望台モジュールに表示される関連語を観点語の話題 (コンセプトマップの子概念) と解釈する。

(d) 文ハイライトのハイライトされた文を読み子概念の概要知識を得る。興味を持ったスニペットを記録する。

(2) ホームページの本文入力

連絡先: 徳永秀和, 香川高等専門学校 機械電子工学科,
〒761-8058 高松市勅使町355, tokunaga@t.kagawa-nct.ac.jp

特に興味を持った話題のスニペットのホームページを開き、本文を抽出し、TETDMに入力する。
 (2-1) (1)と同じ手順で知識を得る。
 (2-2) 以下の手順で要約文を利用する。
 (a) 展望台で興味を持った観点語を選ぶ。
 (b) 要約文より知識を得る。
 (3) 「環境問題」と興味を持った話題を AND 検索したスニペットを入力し、(1), (2)を行う。

3. 実験結果

被験者Aの検索過程を以下に示す。
 (a) 「環境問題」の検索スニペットの展望台より、「温暖」、「汚染」を観点語に選択。
 (b) 文ハイライトより、5つのスニペットに注目し「地球温暖化」、「オゾン層破壊」、「生物多様性減少」などの話題を抽出。
 (c) 汚染物質を含むスニペットのホームページより文章を取得し、TETDMに入力。
 (d) 展望台にて「温暖」、「オゾン」を観点語に選択。
 (e) 文ハイライトと要約文より、「紫外線による病気」、「オゾン層破壊の原因」を修得。
 (f) 次の話題として「生物多様性減少」の調査のため、「環境問題」、「生物多様性」での検索スニペットを取得し、TETDMに入力。
 (g) 展望台にて「生物」、「絶滅」を観点語に選択。
 (h) 文ハイライトより、5つのスニペットに注目し、各ページを開いた中で、興味を持ったページの文章を取得し、TETDMに入力した。
 (i) 展望台にて「生物」、「多様」、「絶滅」を観点語に選択。
 (j) 文ハイライトと要約文より、「人工増加による乱獲」、「水銀などの汚染物質」、「外来種による絶滅」、「食品、医薬品などへの影響」を修得。
 (k) (f)に戻り、展望台に「活動」、「日本」、「企業」を観点語に選択。
 (l) 文ハイライトより、「企業にとってどのような意味があるか」とあるホームページの文章を取得、TETDMに入力。
 (m) 展望台にて、「企業」、「動き」を観点語に選択。
 (n) 文ハイライトと要約文より、「企業価値を高め、収益獲得する」を修得。

被験者Bは、被験者Aに対して文ハイライトの使用が少なく、要約文を活用するなど TETDM の活用方法に差が見られた。
 2人の被験者の作成したコンセプトマップを図2, 図3に示す。

4. 結論

被験者は TETDM の環境を用いることで、インタラクティブに知識の修得と再分析を複数のパネルを連動させながら行うことができた。特に、観点となる単語の選択と知識の修得を繰り返す際に、観点語に関する Web ページの要約文と、観点語に関する文のハイライト表示を、インタラクティブに確認できたことが、効率的な知識修得に役立てられたと考えられる。また、2人の被験者は TETDM の使用方法に少し差があり、作成されたコンセプトマップには異なったキーワードが多く存在しており、ユーザの興味に応じた柔軟な情報収集を実現できていた。
 本実験で用いた2つの処理モジュールは、7年の時を隔て異なる研究者により独立に開発されており、TETDM でなければ両モジュールの連動はもちろん、両システムを組み合わせて利用すること自体も難しかった。

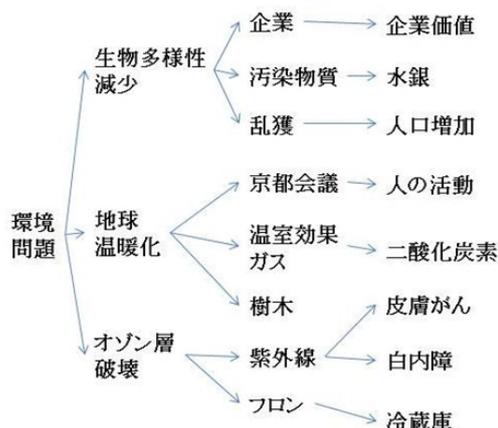


図2 被験者Aのコンセプトマップ

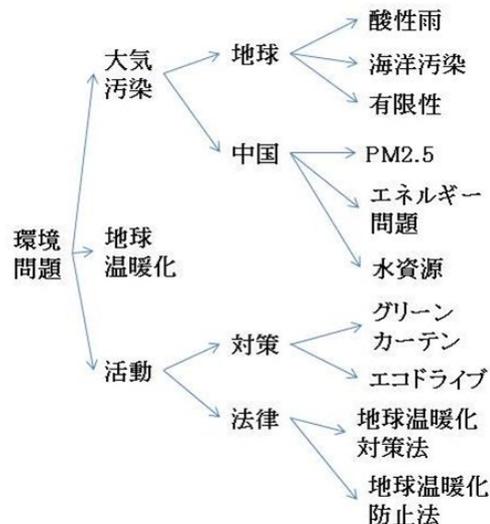


図3 被験者Bのコンセプトマップ

参考文献

[齋藤 11] 齋藤ひとみ, 中島諒, 江草由佳, 高久雅生, 寺井仁, 神門典子, 三輪眞木子: 情報検索の前後におけるユーザの知識構造の変化: コンセプトマップを使った分析, 情報知識学会誌, Vol.21, No.2, (2011).
 [砂山 13] 砂山渡, 高間康史, 西原陽子, 徳永秀和, 串間宗夫, 阿部秀尚, 梶並知記, ダヌシカ ボレガラ, 佐賀亮介: テキストデータマイニングのための統合環境 TETDM の活用と実践, 第27回人工知能学会全国大会, 3B3-NFC-01b-4, 2013.
 [砂山 02] 砂山渡, 谷内田正彦: 観点に基づいて重要文を抽出する展望台システムとそのサーチエンジンへの実装, 人工知能学会論文誌, Vol.17, No.1, 2002.
 [西原 09] 西原陽子, 佐藤圭太, 砂山渡: 光と影を用いたテキストのテーマ関連度の可視化, 人工知能学会論文誌, Vol.24, No.6, 2009.
 [Marchionini 06] Gary Marchionini, Exploratory search: Exploratory search: from finding to understanding, Communications of the ACM, Volume 49 Issue 4, 2006.
 [三輪 09] 三輪眞木子, 江草由佳, 齋藤ひとみ, 高久雅生, 寺井仁, 神門典子: 光 Web 上の exploratory search の特徴-発話プロトコルと事後インタビュー分析結果より-, 情報処理学会研究報告, vol.2009-FI-96 no2, 2009.