

平成 29 年 6 月 19 日現在

機関番号：62601

研究種目：若手研究(B)

研究期間：2013～2016

課題番号：25730193

研究課題名（和文）探索型検索前後の知識の変化をとらえる評価手法の開発

研究課題名（英文）Development of Analysis Methods for Knowledge Changes between before and after Exploratory Search

研究代表者

江草 由佳（Yuka, EGUSA）

国立教育政策研究所・研究企画開発部教育研究情報推進室・総括研究官

研究者番号：60413902

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：本研究では、検索前後の知識の変化を直接的にとらえる評価手法の開発を目指した。まずは、コンセプトマップを用いて検索前後の知識の変化を測れるかどうかについて検討した。検索群と検索と対照群のコンセプトマップの比較を行い、検索前後のコンセプトマップの変化が検索によるものである可能性を提示した。また、検索前後の知識の変化を計測するための基礎的なツールとして、ベースライン検索システムを開発し公開した。

研究成果の概要（英文）：We aim at developing the analysis methods for knowledge changes between before and after exploratory search. We examine how concept maps are usable as the analysis methods for knowledge changes. Through a user experiment, we compare two groups of concept maps either with or without exploratory search. The results from the experiment indicate that the changes of concept maps reflect process of searches. We have also developed a baseline retrieval system for experimental settings, and the baseline system is available as an open source software at Github.

研究分野：図書館情報学・人文社会情報学

キーワード：情報探索行動 情報検索 情報検索システムの評価 コンセプトマップ

1. 研究開始当初の背景

(1)探索型検索 (Exploratory Search) とは、探索者が学習しながら情報ニーズを明確化しつつ、問題解決を進めるプロセスである。現在、探索型検索を支援するシステムの開発が進められている。それとともに、探索型検索システムの評価方法の開発も必要になってきた。伝統的な情報検索システムであれば、クランフィールド実験にはじまった情報検索システム評価の仕組みが整いつつあるが、探索型検索システムの評価方法はいまだ模索が続いている。

(2)探索型検索システムは伝統的な情報検索システムとは異なり、検索結果の適合率や関連度順の評価ではなく、探索者がそのプロセスを通じて学習し、知識を変化させたかを評価することがキーとなる。

しかし、今まで提案されてきた探索型検索システムの評価方法は、直接的に知識の変化を図る方法ではなく、システムログや発話、インタビュー、アンケートに基づき、利用者がシステムを使ってうまく情報を検索できているか、また探索的な検索が来ているかによって評価する方法である。

2. 研究の目的

本研究では、検索前後の知識の変化を直接的にとらえる評価手法の開発を目指す。まずは、コンセプトマップ (図 1) を用いて検索前後の知識の変化を測れるかどうかについて検討した。

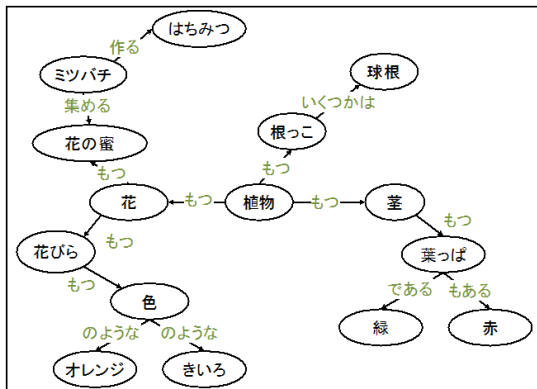


図 1 コンセプトマップの例

代表者らは、これまで探索のトピックや探索条件が明らかに異なる状況として、環境問題と旅行というトピック、収束と発散という探索条件を設定し、探索の前後のコンセプトマップの比較分析を行ってきた。その結果、条件やトピックによってマップの特徴が異なることが明らかになっており、コンセプトマップによって知識の変化を図ることができる可能性を示唆している。本研究ではこのコンセプトマップを用いた実験の分析をさらに進め、より精緻なコンセプトマップ分析を目指すとともに、追加で実験も行ない、評価手法としての考察を試みる。

3. 研究の方法

(1) 検索の有無によるコンセプトマップの変化

代表者らがこれまで行ってきた研究も含めて、情報検索におけるコンセプトマップに関する研究では、検索前後の変化のみを対象としてきた。そのため、実験で得られた利用者の検索前後のコンセプトマップの違いが、本当に検索による影響なのか、それとも検索をしなくても教示や時間を置くだけで変化したものなのかが明らかになっていない。そこで本研究では、先行研究で得られた結果が検索による変化なのかどうかを明らかにするために、検索ありと検索なしの 2 つの条件を設定して、課題の前後のコンセプトマップを比較した。

(2) ベースライン検索システムの開発

検索前後の被験者実験を行う場合には、検索の元となるデータとその検索システムを固定して、検索による違いなのか、検索可能なデータの違いなのかの区別をつけやすくする必要があるので、本研究では、ベースラインの検索システムとして、ウィキペディア日本語版のデータを元にした検索システムを開発した。

(3) コンセプトマップの内容的な変化

代表者らはこれまで、コンセプトマップのグラフ要素を量的に分析する手法を使ってきた。しかし、コンセプトマップの内容的な変化についての分析は十分でなく、その特徴は明らかとなっていない。そこで、コンセプトマップのノード間の内容を人手でタグ付けし、その内容分析を行った。

4. 研究成果

(1) 検索の有無によるコンセプトマップの変化

代表者らがこれまで行ってきた分析のひとつに、検索後のコンセプトマップに新たに表れたノードは New ノード、なくなってしまったノードを Lost ノード、検索前後のコンセプトマップに共通に表れるノードを Common ノードと名付け分析してきた。この分析を、検索していない群 (検索をせずにゲーム) 以降フィルター条件とよぶ) と、検索した群 (検索条件) とで比較した。図 2 は、収束タスクの際の Common, New, Lost ノードについてまとめたグラフである。Common ノード数は検索条件のほうがフィルター条件より少なく、逆に、New ノードと Lost ノードは検索条件のほうが多いことが示された。つまり、検索条件のほうがフィルター条件よりコンセプトマップがより変化した。また、中心ノードからの距離ごとの変化量については、検索条件の場合は、発散タスクは、距離が近いノードが事後に多くなっており、収束は距離が遠いノードが事後に多くなっていった。また、フィルター条件の場合は、発散の距離 2 のみ事後に多くなっていった。つまり、検索条件のほうがフィルター条

件よりタスクの影響をうけていたといえる。

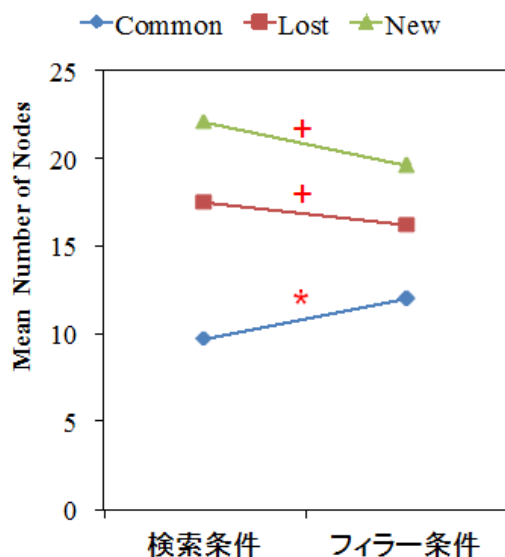


図 2: 収束タスクによる Common, Lost, New ノード

(2) ベースライン検索システムの開発
 ウィキペディア日本語版のデータを検索できるベースライン検索システムを開発した。開発したベースライン検索システムのソースコードは Github にて公開し、研究コミュニティで共有して利用できるようにした。また、本システムは、ウィキペディアのカテゴリ情報をもとに多様な検索結果を返す手法について実装し、NTCIR (情報検索システムの参加型評価ワークショップ) に、その検索結果を提出し、多様な検索結果についての評価も進めた。

(3) コンセプトマップの内容的な変化
 本研究で使用したノードとノードの関係を表すタグは、階層、因果、道具、状態・属性、場所・時間、反対、同義、その他の 8 種類である。この 8 種類のタグとタグ付ルールは、ボトムアップに作成した。まず、はじめに、著者ら 3 人によって、このコンセプトマップを見ながら仮のタグを作成した。仮のタグは階層、具体例、全体部分、効果、因果、SV、SVO、道具、時間・場所、反対、同義、その他の 12 種類であった。つぎに、タスク (発散、収束)、トピック (教育問題、環境問題)、条件 (検索、フィラー) の異なる組み合わせの 12 マップを抽出し、著者ら 3 人によって仮のタグを使ってタグ付けを行った。著者ら 3 人がタグ付けを行った結果を用いて、仮のタグ 12 種類から、タグ 8 種類を決定し、タグ付のルールも決定した。タグやタグ付けルールの決定の際には、タグ付けする人によって揺らぎがでにくくすること、ルールとして明示できるようなルールにすることを重視して決定した。
 コンセプトマップ中、最も数の多かったタグを「最頻タグ」と定義し、すべてのコンセプトマップの「最頻タグ」を導出した。例えば、

あるコンセプトマップ中に、因果タグが 13 つ、階層タグが 6 つ、場所・時間が 0 つ、状態・属性が 1 つ、同義が 0 つ、道具が 0 つ、反対が 0 つ、その他が 0 つ、あったとすると、因果タグが 13 つと最も多いので、因果タグが「最頻タグ」となる。最も多かったタグが複数あった場合は、それら全てを「最頻タグ」とした。この分析をすることにより、階層の関係が最も多いコンセプトマップである、因果の関係が最も多いコンセプトマップである、階層と因果が同じくらい多いコンセプトマップであるなどがわかる。
 分析の結果、最頻タグは、因果、階層、その他のマップがほとんどで、場所・時間、状態属性、同義、道具、反対のマップはあまりなかった。
 本研究では、描かれたコンセプトマップがどんな関係をより多く書いているコンセプトマップかを知る方法として、「最頻タグ」を提案し、例えば、階層関係についてもっともたくさん描いているコンセプトマップということを示すことができるようになった。

5. 主な発表論文等
 (研究代表者、研究分担者及び連携研究者には下線)

- [雑誌論文](計 2 件)
- Song Ximei, Yuka Egusa, Hitomi Saito, Masao Takaku; IRCE at the NTCIR-12 Imine-2 Task; In Proceeding of the Twelfth NTCIR conference (NTCIT-12); 2016 年; p.1-p.5 <http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings12/pdf/ntcir/IMINE/08-NTCIR12-IMINE-SongM.pdf> (査読あり)
 - Yuka Egusa, Masao Takaku, Hitomi Saito; How Concept Maps Change if a User Does Search or Not?; IliX '14 Proceedings of the 5th Information Interaction in Context Symposium, 2014 年, pp.68-75, DOI: 10.1145/2637002.2637012 (査読あり)

- [学会発表](計 4 件)
- Yuka Egusa, Masao Takaku, Hitomi Saito; Evaluating Complex Interactive Searches Using Concept Maps; Proceeding of the Second Workshop on Supporting Complex Search Tasks; 国際学会; 2017 年 3 月 11 日, Oslo, Norway. <http://ceur-ws.org/Vol-1798/>
 - 齋藤ひとみ, 江草由佳, 高久雅生; 検索の前後における利用者のコンセプトマップの分析: 検索の有無による比較; Web インテリジェンスとインタラクション研究会 第 3 回ステージ発表, 東京 (2015-11) (招待講演)
 - Yuka Egusa, Masao Takaku, Hitomi Saito; How to Evaluate Searching as Learning; Searching as Learning Workshop (IliX 2014 workshop); 国際会議; 2014 年 8 月 30 日 ~ 8 月 30 日, Regensburg, Germany

<http://www.diiigubc.ca/IIIXSAL/Papers/Egusaetal.pdf>

・齋藤ひとみ, 江草由佳, 高久雅生, 検索の前後における利用者のコンセプトマップの分析: 検索の有無による比較; 第5回 ARG-WI2 研究会; 2014年11月22日~11月23日; 東京

<http://www.sigwi2.org/wp-content/uploads/2016/02/WI2-2014-38.pdf>

〔その他〕

・IRCE ペディア

<https://github.com/cres-project/irce-wikipedia>

6. 研究組織

(1) 研究代表者

江草 由佳 (Yuka EGUSA)

国立教育政策研究所・研究企画開発部教育
研究情報推進センター・総括研究官

研究者番号: 60413902

(2) 連携研究者

高久 雅生 (Masao TAKAKU)

筑波大学・図書館情報メディア系・准教授

研究者番号: 00399271

齋藤 ひとみ (Hitomi Saito)

愛知教育大学・教育学部・准教授

研究者番号: 00378233