

Wikipedia からキーワードを援用し初出の知識の学習と 学習コンテンツの発掘を支援するサービスの研究

横山 誠[†] 村上 貴彦[‡] 中村 太戯留[†] 上林 憲行[†]
東京工科大学[†] 東京工科大学大学院[‡]

1. 研究背景

わが国では自律的・能動的な学習、つまり先生や教材が無い状況でも自ら行動し学び進めることを求めている[1]。学習者の所有する領域知識（学習対象に関する知識）は経験により異なるが、学習者にとって初出の知識、すなわち領域知識が非常に少ない状況で自律的な学習が必要な際、大量の情報へ容易にアクセス出来る Web 検索が重要な役割を担う。Web 検索の結果取得できるページは、入力したキーワードに依存する為、キーワードの入力を補助する機能が多くの検索エンジンで採用されている。

例えば、Google サジェスト[2]はキーワードの入力に伴い、自分や他人の検索履歴からキーワードを提案する。しかし、本研究における学習者は、所有する領域知識が非常に少ない為、提案されたキーワードの適合判断や、別のキーワードの考案は困難となる。同様の理由から、入力したキーワードの適合判断も困難な場合がある。また[2]は、検索履歴がない初回検索時はキーワードが提案出来ず、入力したキーワードが適切でなくても提案されてしまう為、不十分である。このように、検索が困難になると、検索結果から有用な学習コンテンツが顕在化せず、自律的・能動的な学習が困難となる。

中島ら[3]によると、領域知識のあるユーザーより知識のないユーザーの方が、問題文外の単語を使わずに検索している。初出の知識の学習者にとって、キーワードの考案や検索結果の読解はより困難になると予想された。

そこで、本研究では、学習目的で Web 検索を行う際、領域知識の理解を検索キーワードの提案を兼ねて支援する[図 1]ことで、学習者の検索と自律的・能動的な学習態度への変化を促すツールを開発し、本学演習内で効果を検証した。

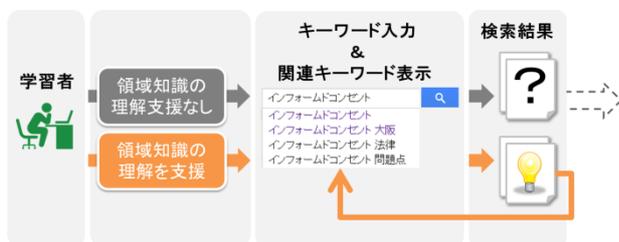


図 1：領域知識の理解支援ツールの位置づけ

“An empirical study of mushrooming search queries using Wikipedia for learning unexplored practical skills”
[†]MAKOTO YOKOYAMA, [‡]TAKAHIKO MURAKAMI,
[†]TAGIRU NAKAMURA, [†]NORIYUKI KAMIBAYASHI
[†]Tokyo University of Technology,
[‡]Tokyo University of Technology Graduate School

2. 方法

2.1.本ツールの要件

中島ら[3]により、領域知識が非常に少ないと検索が困難になる可能性が予想された。これを支援することで、検索が可能になり、ひいては学習者の自律的な学習態度を促せると考えた。そこで、次の 4 つを本ツールの要件とした。1) 入力したキーワードに関連するキーワードを取得できる、2) 関連するキーワードの情報源は学習者の利用履歴に依存しない、3) 関連性を視覚的に表現出来る、4) 必要に応じてキーワードを展開出来る。

2.2.アルゴリズム

堀ら[4]の関連単語抽出アルゴリズムが要件に近い為活用した。これは、入力したキーワードに対応する Wikipedia 記事内の文章を形態素解析して名詞を抽出し、得た関連キーワードの候補を「出現回数」「入力されたキーワードとの距離の近さ」を計算し関連度を重み付けする。本研究では、関連性を確実に取得する為、取得する単語は文章内のハイパーリンクされた単語に限定した。

2.3.プロトタイプ

サービスの利用画面を以下に示す[図 2]。画面左、放射状に表示されたグラフのうち、オレンジ色のノード（中央下）が入力したキーワード、周辺ノードは関連するキーワードを示す。周辺ノードには番号が振られており、若いほど関連が強い。初期画面で一度に表示する語は、島田ら[5]の研究を参考に上位 10 件とした。画面右側で関連するキーワードの選択と検索を行う。

選択中のキーワードは黄緑色に変わり、更に関連キーワードを抽出し表示する。表示方法は、中心ノードと周辺ノード間で関連度が強い順にした。例として、「インフォームドコンセント」と入力し、取得した「パターンナリズム」から関連キーワードを展開した。



図 2：サービス画面の例

2.4. 実験方法

28名の大学2年生に、実際に履修中の演習で求められる2つの技能を、Web上の動画を検索し学ぶという課題を課した。プロトタイプは演習用サイトにアップロードした後、入手方法を紹介し、誰でも利用できる状況にした。課題は1)「実験におけるインフォームドコンセントはどのようにして行いますか?」と2)「感情体験を観察する方法にはどのようなものがありますか?」の2つであった。実験前と終了後にアンケートを行い、各課題に対しどのぐらい領域知識を所有しているか、四択で回答してもらった。実験終了後、検索履歴と動画視聴履歴を収集した。プロトタイプの利用有無、所要時間のアンケート調査を行い、視聴履歴を基に課題の学習に有用だった動画がどれか尋ねた。なお、プロトタイプを利用するかどうかの判断は教育上の配慮から学生に選択させた。

3. 結果

プロトタイプ使用者(提案手法)は8名、未使用者(対照条件)は20名であった。実験前に行ったアンケートの結果、課題1及び2について、プロトタイプ使用者は、意味や方法は知っていても体験・実践した経験がある者はいなかった。プロトタイプ未使用者は、1名が課題1(インフォームドコンセント)の意味や方法を知っていたが、体験・実践した経験はないと答えた。残ったプロトタイプ未使用者は、各課題に対し体験・実践した経験はなく、意味は知っているが方法は理解していない、あるいは聞いたことはあるが意味や方法は理解していないと答えた。

プロトタイプの使用有無で単位時間あたりの平均獲得異なり検索キーワード数に有意な差が見られた[図3]($t(26)=0.0168, p<.05$, 片側検定)。これは、一人ひとりの検索履歴から重複するキーワードを省き、残ったキーワード数の合計を各実験参加者が回答した所要時間で割り、その平均をプロトタイプの利用有無で比較した。なお、提案手法では関連するキーワードを100個以上提案しており、対照条件より考案可能な検索キーワード数が減る可能性は、極めて小さいと推測されるため、片側検定を実施した。図表の数値は小数点第二位を四捨五入した。

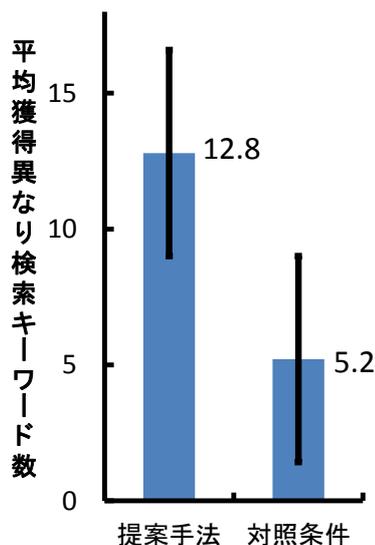


図3: プロトタイプ利用有無による平均獲得異なり検索キーワード数の違い

その他、平均検索回数、平均動画視聴本数、平均所要時間にも違いが見られた[表1]。実験参加者が獲得した有用動画数には有意な差は見られなかった[図4]。

表1: プロトタイプ使用有無による違い

	提案手法	対照条件	割合
平均検索回数	38.6	19.1	約2.0倍
平均動画視聴本数	33.5	20.0	約1.7倍
平均所要時間	2.3	3.0	約2割減

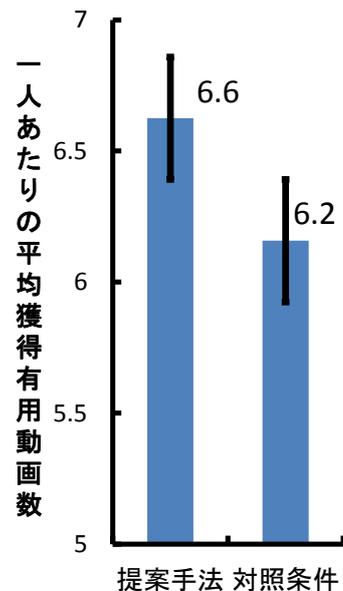


図4: プロトタイプ使用有無による一人あたりの平均獲得有用動画数の違い

4. 考察

単位時間あたりの平均獲得異なり検索キーワード数に有意な差が見られたことから、プロトタイプの利用が領域知識の学習とその効率化に加え、学習者の好奇心や興味に影響し、与えられた課題をこなす為の学習、すなわち他律的・受動的な学習態度から、自律的・能動的な学習態度を促進した可能性があることが示唆された。しかし、獲得した有用動画数に違いが見られなかった為、プロトタイプの利用が適切な学習コンテンツの発掘に寄与したかどうかは分からなかった。理由として、検索段階での領域知識の学習支援が不足していることが考えられる。今後は関連性の表現だけでなく、提案したキーワードに対する概説を表示する機能の実装等を検討していく。

参考文献

- [1] 文部科学省: 大学改革実行プラン～社会の変革のエンジンとなる大学づくり～, p3, (2012).
- [2] Google サジェスト, <https://www.google.co.jp/>
- [3] 中島 悠: 検索経験と領域知識の WWW 情報検索行動に与える影響, ヒューマンインタフェース学会, pp. 1-10, (2006).
- [4] 堀 憲太郎: Wikipedia からの拡張クエリ生成による Web 検索とその評価, 人工知能学会 pp1-8, (2008).
- [5] 島田 論: 次数制約を加えた共起語グラフに基づくキーワード間ナビゲーション, 情報処理学会, p8, (2009).