

LLMを用いた階層的トピック提示による 探索的論文検索支援

島津 秀弥¹ 岨野 太一¹ 高田 秀志¹

概要: 学術論文の探索では、初期段階に調査観点が定まらず、検索語の更新を繰り返しながら関連領域を把握する必要がある。しかし、キーワード入力中心の検索では切り口が固定化しやすく、関連語の発見や整理が追いつかない場合に探索が停滞する。本稿では、入力語から関連トピックを生成し、階層的に提示して可視化することで、探索の切り口更新を支援するシステムを提案する。利用者はトピックを展開・折りたたみで整理し、選択したトピック群を検索クエリへ接続して文献探索を進める。利用者実験により、提案システムが論文候補の発見しやすさや関連トピック把握に良い影響を与えることが確認された。

1. はじめに

学術論文の探索では、初期段階に調査目的や観点が定まらない場合が多く、試行錯誤しながら検索語を更新しつつ関連領域を把握する探索的探索が必要となる。しかし、キーワード入力を中心とする学術検索では、語彙選択が探索範囲を強く規定し、適切な検索語を思いつけない場合や関連語が不足する場合に探索が停滞しやすい。その結果、上位の検索結果を多数開いて比較する行動に偏り、探索の見通しや現在地を保ちにくくなる。

この背景には、(1) 初期クエリの表現に依存して探索の切り口が固定化しやすい点、(2) 関連語を広げても整理が追いつかず、探索の構造を把握しにくい点、の2点がある。

そこで、入力語から関連トピックをLLMを用いて生成し、階層的トピックとして提示する探索支援システムを提案する。提案システムは、提示トピックを展開・折りたたみにより整理しながら探索空間の構造を把握し、選択したトピック群を学術検索クエリへ接続して文献探索を進めることができるという特徴を持つ。

また、提案システムが関連トピックの把握と探索の進行を支援し、満足度や利便性の向上に寄与するかを検証する。

2. 関連研究

文献探索においては、入力したキーワードのみでは適切な切り口が得られず、関連語の発見が探索の進行を左右する場合がある。関連語提示の代表例として、大石らはWeb検索におけるキーワード関連語提案システムを示し、入力

語に対して周辺語を提示することで探索の広がりや支援する枠組みを報告している [1]。一方で、提示語の粒度や数が利用者の取捨選択に影響するため、提示結果を利用者が扱いやすい形で整理して提示する設計が重要と考えられる。

近年は、生成AIを文献調査に活用する実用的なツールが多数登場しており、森内は文献調査に利用できる生成AIツールとその使われ方を整理している [2]。ここでは、生成AIが関連文献の探索や見通しの獲得に利用され得ることを示す一方で、探索を支援するには、利用者が次の探索方針を決められるような切り口の提示が重要であることが示唆されている。

切り口提示の方法として、LLMを用いたサブトピック提示に着目した研究が進んでいる。堀川らは、多様なテーマを対象としてLLMを用いたサブトピック提示手法の有効性を評価し、サブトピック提示が情報探索における行動に影響し得ることを報告している [3]。これを踏まえ、本研究は、LLMにより生成した関連トピックを階層的サブトピックとして提示し、利用者が展開・折りたたみにより探索空間の構造を把握しながら探索方針を更新できる点が特徴である。

3. 提案システム

3.1 システムの全体像

図1に、提案システムの処理と操作の流れを示す。本システムでは、利用者が入力した探索語を起点として、LLMへのプロンプトにより関連トピック候補を得る。次に、候補に含まれる重複語や表記揺れ、汎用的で切り口になりにくい語を除去し、提示に適した語に整形する。整形後の候補は抽象度の違いが分かるように階層化し、トピックツ

¹ 立命館大学 情報理工学部

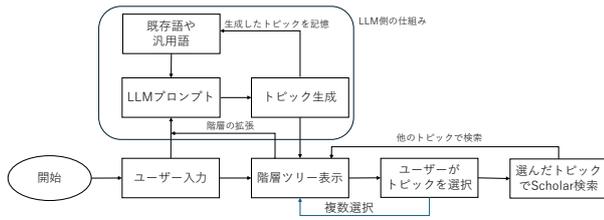


図 1 提案システムの処理フロー

リーとして表示する。

図 2 に画面例を示す。利用者はツリー上のノードをクリックして子トピックを展開し、必要に応じて折りたたんで表示を整理する。表示が煩雑になった場合は、画面上部の「Fit」ボタンで全体が画面内に収まるように調整し、「Reset」ボタンで表示を初期状態に戻す。

関心のある語はツリー上で選択し、検索語リストに追加する。検索語は複数語を保持でき、追加・削除により検索語リストを更新する。検索実行時には、検索語リストに登録された語を AND 連結したクエリとして Google Scholar の検索欄に渡して検索を行う。

3.2 トピック生成とフィルタリング

図 3 は、ノードをクリックして子トピックを展開した例である。LLM によるトピック生成では、入力語に対して探索の切り口となる候補語を複数生成し、提示に適した語を選択する。提示語は検索語として扱いやすいよう名詞句を基本とし、助詞を含む文や冗長な説明文にならないように制約する。

生成された候補には、表記揺れや同一内容の重複、および単独では切り口になりにくい汎用語が混在する。そこで提示前に、正規化とルールベースのフィルタリングを行う。正規化では、英数字の全角・半角、大小文字、記号（括弧やハイフン等）、空白の有無を統一し、同一語が同一表記になるよう整形する。フィルタリングでは、(i) 正規化後の完全一致により重複候補を除去し、(ii) ストップリスト（例：「研究」「手法」「システム」など分野横断で頻出し単独では探索の切り口になりにくい語）に該当する候補を除外する。さらに、検索語として扱いにくい候補を抑えるため、(iii) 極端に短い語や冗長な語（説明文に近い語句）を除外し、提示に適した語のみを残す。

さらに、このような候補語の選別によって提示数が不足する場合に備え、生成数を多めに確保した上で段階的に絞り込むことで、提示語数の安定化を図る。

3.3 階層的サブトピック提示

階層構造の構築では、最初に広義語（親）を提示し、利用者の操作に応じて狭義語（子）を追加生成する。これにより、探索の初期段階では広い切り口を得やすくし、関心が定まってきた段階では特定の切り口を掘り下げられるよ

うにする。提示語は階層ツリーとして可視化し、利用者がトピック間の関係を把握しながら探索を進められるようにする。

3.4 検索語の指定と検索エンジンへの接続

画面上では、提示されたトピックを展開して候補を増やしたり、折りたたんで見通しを回復したりできるようにする。また、表示領域の拡大縮小やスクロールなどの操作により、提示範囲を調整しながらトピックを確認できるようにする。利用者が選択したトピックは、画面上の検索語リストに追加される。検索語リストは複数の検索語を保持でき、検索語リストの追加・削除によりクエリを更新する。検索実行時には、検索語リストに登録された語を AND で連結した検索式を用い、Google Scholar で検索を実行する。

4. 評価実験

4.1 実験方法

提案システムが探索的文獻検索における満足度および利便性の向上に寄与するかを検証するため、利用者実験を実施した。参加者は 8 名とし、提案システム条件と既存の学術文獻検索サービス条件（Google Scholar のみ）の 2 条件で探索課題に取り組む。条件順序の影響を抑えるため、参加者間で条件順をカウンタバランスした。

探索課題は 2 種類（課題 A、課題 B）とした。課題 A は「プログラミング教育」、課題 B は「情報探索・検索支援」に関する文獻探索とし、参加者は指定テーマに対して関心のある観点を見つけながら文獻候補を収集する。各課題では、検索語を更新しながら探索を継続し、最終的に「興味がある」と判断した文獻候補を複数件挙げる。提案システム条件では、提示されたトピックを展開・折りたたみで整理し、選択したトピックを追加・削除しながら検索語リストを構成して文獻探索を進める。既存サービス条件では、参加者が検索語を直接入力して文獻探索を進める。

収集データは、(1) 主観評価（満足度、利便性、探索の進めやすさ等）、(2) 探索中の行動ログ（提案システム上のログ）、(3) 画面録画に基づく観察結果、とする。主観評価は各条件の課題終了後に質問紙で取得する。また、両条件の終了後に、2つの方法を直接比較する設問（3問）に回答を求める。さらに、提案システム条件については、システム固有の評価項目（3項目）を追加で取得する。

行動ログは提案システム上で記録し、トピック生成要求、展開・折りたたみ、検索語の追加・削除、検索実行といった主要操作に加え、文獻ページの閲覧回数、閲覧時間、探索語の更新回数を取得する。画面録画は探索行動の特徴を補足する目的で使い、必要に応じて特徴的な場面を抽出して比較する。なお、文獻閲覧数と興味あり候補数の集計では、参加者 1 名分の当該データが欠損していたため除外し、7 名分のデータとして分析した。それ以外の主観評価およ

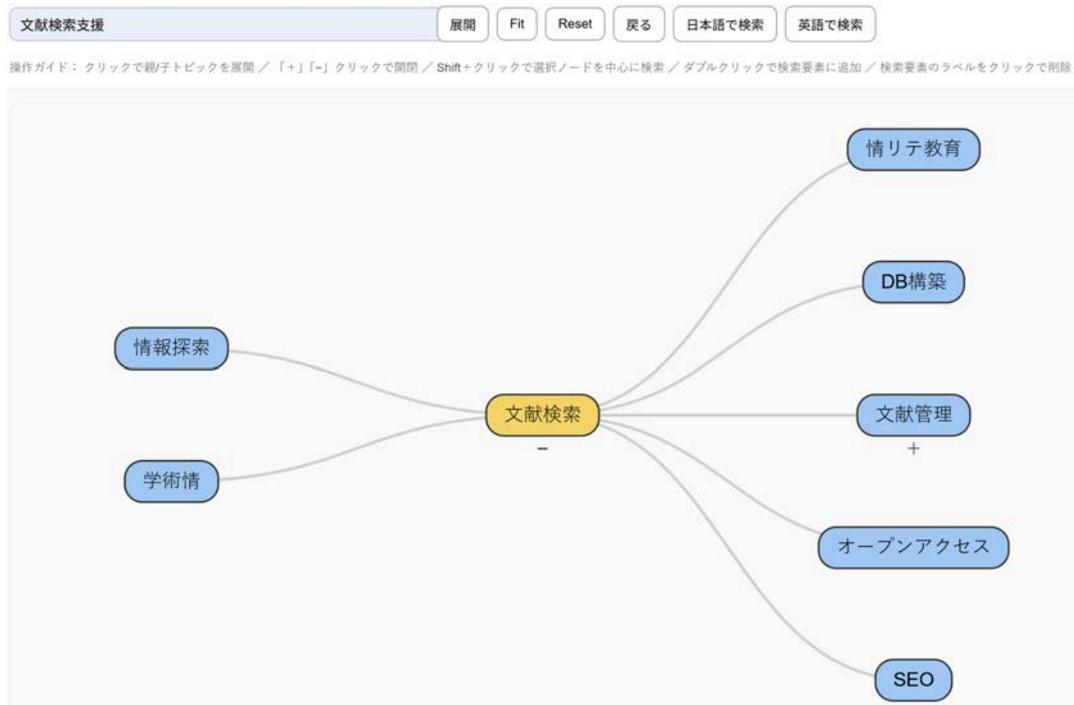


図 2 提案システムの画面例（操作ボタンとトピックツリー）

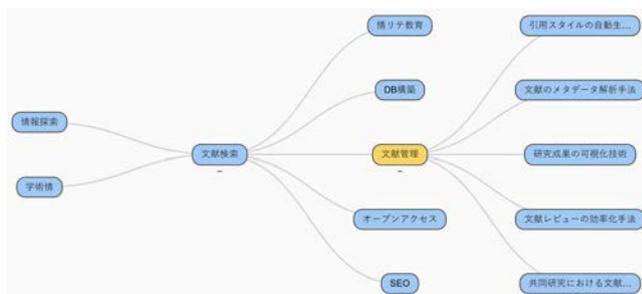


図 3 トピックツリーの展開例（子トピックの追加提示）

び直接比較の集計は全員分のデータを対象とした。

4.2 評価の観点

本研究では、提案システムが探索的文献探索に与える影響を、以下の観点から評価する。

RQ1（切り口の発見・関連領域の把握）：

関連トピックの把握や切り口の発見を支援できるか。

RQ2（探索の進めやすさ・負担感）：

探索の進めやすさと負担感に改善が見られるか。

RQ3（満足度・利便性）：

全体としての満足度や利便性を向上できるか。

4.3 実験結果

主観評価の結果を表 1 に示す。Wilcoxon 符号付き順位検定の結果、「興味のある論文候補を見つけやすかった」、および「関連しそうなキーワードやトピックの広がりをイ

メージしやすかった」で有意差が確認された ($p < 0.05$)。

文献閲覧数と興味あり候補数の集計を表 2 に示す。提案システム条件は、既存サービス条件と同程度の興味あり候補数に到達しつつ、閲覧数が少ない傾向を示した。

終了後の直接比較（3 問）の集計を表 3 に示す。総合的な使いやすさ、頼りになった度合い、今後の利用意向のいずれも、提案システム寄りの回答が多かった。

提案システムに関する 3 項目の結果を表 4 に示す。全体像把握ときっかけ提示に関する項目は高い評価を得た一方で、満足度は他項目に比べてやや低めであった。

自由記述では、良かった点として「関連トピック・キーワードの広がり」「視覚的な分かりやすさ」「検索語を思い浮かなくても辿れる点」などが挙げられた。一方、悪かった点として「操作の分かりにくさ/安定性」「探索が広がり過ぎてテーマから逸脱する懸念」などが挙げられた。

行動ログおよび画面録画の観察から、既存サービス条件では上位文献を短時間で比較する行動が多く、提案システム条件では提示トピックを確認しながら探索の切り口を更新した後文献候補へ到達する行動が見られた。これらは、主観評価および直接比較の結果とも整合する。

4.4 考察

RQ1（切り口の発見・関連領域の把握）について、主観評価では「関連しそうなキーワードやトピックの広がりをイメージしやすかった」が提案システム条件で有意に高かった。自由記述でも、「入力語だけでは思いつかなかった語に辿れた」「関連の見通しが立った」といった記述が見られ

表 1 主観評価の条件別平均と有意差検定結果

評価項目	既存サービス平均	システム平均	p 値
自分なりに興味を持ってそうな論文を、ある程度見つけることができた.	3.88	4.38	0.3125
タスクは全体として難しかった. ※ (逆転済み)	3.25	3.88	0.2812
「何をすればよいか分からなくなる (迷子になる)」と感じることが多かった. ※ (逆転済み)	2.50	3.63	0.1381
今回の方法の画面構成や操作方法は、直感的で分かりやすかった.	3.13	4.25	0.0625
今回の方法では、興味のある論文候補を見つけやすかった.	2.50	4.25	0.0312*
今回の方法では、関連しそうなキーワードやトピックの広がりをイメージしやすかった.	2.00	4.75	0.0078*
タスクを行っている間の精神的な負担・疲労感. ※ (逆転済み)	2.50	3.75	0.0781

* は有意差あり ($p < 0.05$) を表す.

表 2 文献閲覧数と興味あり候補数の集計

条件	閲覧数	興味あり	興味あり率
既存サービス	39	17	0.44
提案システム	33	17	0.52

興味あり率は (興味あり / 閲覧数) で算出.

表 3 終了後の直接比較 (3 問) の集計

設問	提案システム	既存サービス
使いやすかった方法	7	1
探索で頼りになった方法	8	0
今後使いたいと思う方法	7	1

表 4 提案システムに関する 3 項目の平均

設問	平均
関連トピックの全体像の把握に役立つ	4.5
興味に合った論文を見つけるきっかけを与える	4.5
システム全体に満足している	4.1

た. 以上より, 本システムで提供されるトピックツリーの階層を辿って候補を増やす操作は, 単一の検索語入力に比べて探索の切り口を増やしやすく, 関連領域を広げる足場として働いたと考えられる. 一方で, 探索が広がり過ぎてテーマから逸脱する懸念も挙げられており, 拡張した切り口を適切に絞り込む支援 (例: 候補の優先度づけや焦点化の導線) が課題である.

RQ2 (探索の進めやすさ・負担感) については, 主観評価では「迷子になりにくさ」や「負担感」に関する項目で提案システム条件が高い傾向を示したものの, 有意差は確認されなかった. ただし, 閲覧数と興味あり候補数の集計では, 提案システム条件は少ない閲覧数で同程度の候補に到達している. この結果は, 関連トピックの提示が「次に試す語」を決める手がかりとなり, 闇雲な閲覧を減らす方向に作用した可能性を示す. また, 行動ログ・画面録画の観察でも, ツリー上で切り口を確認して検索語を更新した後に文献候補へ到達する行動が見られ, 提示トピックが探索方針の更新に利用されていたことが示唆される. 一方で,

効果が参加者間でばらつく可能性もあるため, どのような利用パターンで負担軽減につながるのかを, ログ指標 (展開回数, 検索語更新回数など) と対応づけて検討する余地がある.

RQ3 (満足度・利便性) については, 直接比較において提案システムが「頼りになった方法」として選ばれる傾向が強く, 今後の利用意向も同様の傾向を示した. また, システム固有の評価では, 関連トピックの全体像把握や論文発見のきっかけに関する項目が高く評価された. 一方で, 自由記述には, 操作方法の分かりにくさに加えて, 操作が想定どおりに反映されないなどプロトタイプ実装に起因すると考えられる不安定さへの指摘が見られた. これらは体験の一貫性を損ね, 満足度を下げる要因になり得る. 今後は, (i) 展開・追加などの状態が一目で分かる表示やガイドにより操作意図と結果の対応を明確にすること, (ii) エラー発生時に元の状態へ戻れる導線 (例: 再実行, 状態復帰) の整備, (iii) 検索実行前にクエリを確認できる提示や候補プレビューの導入により, 検索への接続を支援することが改善点として挙げられる.

5. おわりに

本研究では, 探索的文献探索において切り口を得にくい課題に対し, 入力語から関連トピックを階層的に提示し, 選択語を用いて Google Scholar 検索へ接続する支援システムを提案した. 利用者はトピックの展開・折りたたみと選択により検索語を更新しながら探索を進める.

利用者実験の結果, 主観評価では「論文候補を見つけやすかった」および「関連キーワードやトピックの広がりをイメージしやすかった」で有意差が確認された. また, 直接比較でも提案システムが頼りになる方法として選ばれる傾向が見られた. さらに, 文献閲覧数の集計から, 提案システム条件では少ない閲覧数で同程度の候補数に到達する傾向が示された.

一方で, 操作の分かりにくさや, プロトタイプ実装に起因する不安定さ, 探索が広がり過ぎる懸念が課題として挙

げられた。今後は、状態提示とガイドによる操作理解の支援、エラー発生時の復帰導線の整備、提示語の一貫性向上、および検索前確認や候補プレビューによる接続支援を進める。

参考文献

- [1] 大石哲也, 長谷川隆三, 藤田博, 越村三幸: WEB 検索におけるキーワード関連語提案システム, 九州大学大学院システム情報科学紀要, Vol.9, No.1, pp.19-24, 2004.
- [2] 森内泰: 文献調査に利用できる生成 AI ツールと, その使われ方, 長崎県立大学論集 (経営学部・地域創造学部), 2023.
- [3] 堀川達平, 北山大輔: 多様なテーマに対する大規模言語モデルを用いたサブトピック提示手法の有効性評価, Web インテリジェンスとインタラクション研究会予稿集 (Web), 2024.