

学術情報を対象とした発展検索支援システム Evolving Search Support System for Academic Information

千葉 祐† ソウヤ† 李 保寧† 阿部 亨‡ 木下 哲男†
Yuu CHIBA SUOYA Baoning LI Toru ABE Tetsuo KINOSHITA

1. まえがき

情報探索の過程では、直前の検索で獲得された情報がユーザの知識に変化を与え新たな検索要求を生じさせる場合が多い。その結果、ユーザは、異なるクエリによる検索を次々と繰り返すことになる。クエリを変更しながら繰り返されるこのような検索は「発展検索 (evolving search)」と呼ばれる [1]。

文献調査など、学術情報を対象とした探索では、情報の検索・獲得によるユーザの知識の変化が活発であり、その過程で新たな検索要求を生じ易い。従って、学術情報の活用を図るためには、発展検索を支援する機構の実現が特に有効であると考えられる。しかし、従来の情報検索システムでは、探索の過程で得られる情報やユーザの知識の変化を直後の検索で十分に利用できないため、発展検索の効果的支援は困難であった。

これに対し、本稿では、マルチエージェントフレームワークに基づく新たな学術情報検索システムを提案する。提案システムでは、各学術情報をエージェントとして構造化することで、各々に自律性を与えている。これにより、直前の検索で獲得された学術情報が、自身のメタデータや検索のコンテキスト等に基づき、新たなクエリの生成とそれを用いた検索を能動的に行うことができるため、探索の過程で得られる多様な情報を効果的に利用した発展検索支援が可能となっている。

2. 学術情報の探索

まず、本稿で対象とする発展検索支援の範囲を説明し、次に、発展検索支援の観点から、従来の情報検索手法の問題を述べる。

2.1 本稿で対象とする発展検索支援の範囲

発展検索は、図1に示すように、直前の検索で得られた情報から新たな検索要求が生まれ、これを満たすために、ユーザが、クエリを変更し検索を繰り返す過程である。この過程で生じる検索要求は、(a) 直前の検索で得られる情報が主な要因であり直前の検索要求との関連が高いものと、(b) ユーザ自身の「ひらめき」が主な要因であり直前の検索要求との関連が低いものとに大別できる。

学術情報 (論文・図書・史料等) にはキーワードや参考文献等のメタデータが付与される場合が多く、これを有効利用できれば、(a) を推測し新たな検索をシステム側で代行することが可能となる。そこで、本稿では、直前の検索で得られる情報 (獲得された学術情報のメタデータ) を状況 (検索のコンテキスト等) に応じた適切な手法・タイミングで直後の検索へ利用する機構を提案し、(a) に基づく検索をシステム側で代行させることで発展検索の支援を図る。この種の発展検索は、メタデータを手掛りに他の学術情報を芋蔓式に辿る等の典型的・煩雑な作業に相当し、これをシステム側で代行させることの効果は高い。

また、(b) は予測が困難であり、これに基づく発展検索をシステム側で代行することは難しい。しかし、(a) に基づく発展検索を支援し、有用な学術情報をユーザへ効果的に提供できるならば、(b) に基づく発展検索の間接的な支援も期待できる。

2.2 従来の情報検索システム

現在、種々の分野において、情報検索を支援する多数の手法が提案されている。例えば、シソーラスやオントロジを用い、与えられたクエリをより適切なものへ修正する手法 [3]、プロフィールを検索履歴により随時更新し、そこから得られる統計量を用いユーザの興味 (一種の知識) の変化を検索処理へ導入する手法 [4]、クエリに対する検索結果をユーザが評価し、その評価結果をシステムにフィードバックする手法 [5]、情報間の類似度や参照関係を提示する手法 [6] 等がある。しかし、発展検索支援の観点からこれらの手法を考えた場合

- 新たな知識や観点が提案・発見されたとしても、シソーラスやオントロジに登録されない限り、検索処理へ利用できない
- ユーザの興味の変化が、プロフィールから得られる統計量の変化として現れるまでには遅延が生じる
- 探索の過程で得られる各種の情報の分析・利用はユーザ側に委ねられる

等の問題がある。このため、従来手法のみでは、探索の過程で得られる多様な情報を有効に利用できず、発展検索を効果的に支援するシステムの実現が困難であった。

3. 能動的情報資源による発展検索支援

本節では、ユーザによる発展検索を支援するために、まず、作業の一部をシステム側に代行させる手法を提案する。次に、能動的情報資源 (Active Information Resource: AIR) [2] の概念を用い、提案手法を効果的に実現するアプローチを述べる。

3.1 発展検索の代行サイクル

本稿では、図2に示すように、次の3つのステップからなるサイクルを繰り返し、発展検索の作業の一部をシステム側に代行させる手法を提案する。

- **基本検索 (Basic Search)**
各学術情報に付与されたメタデータとクエリとを比較し、クエリに適合する学術情報を検索する。
- **選択 (Selection)**
ユーザ側の検索履歴などに基づき、基本検索の結果を順位付け、有用と思われる学術情報を選択する。

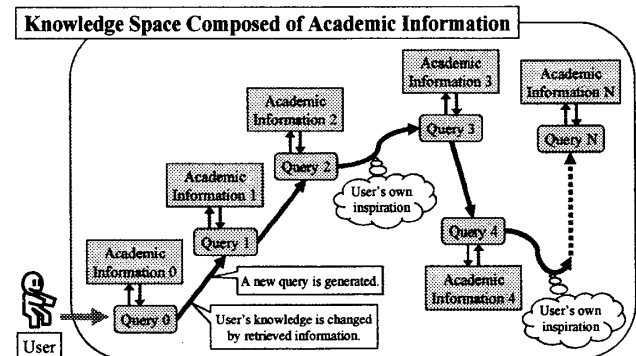


図1 学術情報を対象とした発展検索

† 東北大学大学院情報科学研究科

‡ 東北大学情報シナジーセンター

● クエリ生成 (Query Generation)

選択された学術情報のメタデータにより、直前の基本検索のクエリを変更し、新たなクエリを生成する。

これにより、直前のサイクルで得られたメタデータを、検索のコンテキストに応じた形で次のサイクルに利用し、学術情報を芋蔓式に辿る作業を実現する。

3.2 能動的情報資源 (AIR)

提案手法では、通常、1回のサイクルで複数の学術情報が選択されることになる。これらを次のサイクルで有効に利用するには、選択された学術情報毎に、その特徴に応じたクエリを生成し、次の基本検索を行う必要がある。このような処理の実現には、選択された学術情報が、各々自律的にクエリ生成と基本検索を行う機構が有効であると考えられる。そこで、提案手法の実現に際し、能動的情報資源 (AIR) の概念を学術情報に適用することで、各々が自律的に動作する機構の実現を図る。

AIR は、情報資源 (電子化された各種データ) の構造を強化し、従来ユーザが行っていた情報資源に対する操作の一部を情報資源自身に代行させるものである。具体的には、利用支援知識 (情報資源に関するメタデータや情報資源を操作するための知識) と利用支援機能 (情報資源を操作する機能) とを情報資源に付加したエージェントとして各 AIR が構成される。これにより、ユーザ (或いは他の AIR) からの要求メッセージを受信した AIR は、自身の利用支援知識・機能を用い、自身が管理する情報資源に対する操作を代行し、その結果を応答メッセージとしてユーザへ送信する。また、AIR は、他の AIR とのメッセージ交換により、分散環境上で相互に協調・連携し、より複雑・柔軟な操作を代行することも可能となる。

各学術情報を AIR として構造化すれば、基本検索・選択・クエリ生成を学術情報側に能動的に代行させることが可能となる。この結果、探索の過程で得られる種々の情報を、状況に応じた適切な手法・タイミングで各ステップに反映することができ、発展検索を効果的に代行するシステムの実現が期待できる。

4. 発展検索支援システム

ここでは、本稿で提案する学術情報検索システム (発展検索支援システム) の構成と動作について述べる。

4.1 発展検索支援システムの構成

提案システムは、図3に示すように、ユーザインターフェースである AIR (User-AIR) と、各々 AIR として構造化された学術情報 (AI-AIR) から構成される。

User-AIR は、ユーザ毎に用意され、各ユーザのプロファイルの項目 (キーワード、検索条件等) を利用支援知識として持つ (一部を表1に示す)。利用支援知識の coverage, subject にはキーワードが指定され、各キーワードにはユーザにより優先順位が与えられる。一方、各 AI-AIR は、電子化された学術情

表1 User-AIR, AI-AIR の利用支援知識 (一部)

● User-AIR	
知識の名称	知識の種類
(a) coverage	研究分野名
(b) subject	研究分野におけるキーワード
(c) max-a	1回の選択で選ばれる件数の上限

● AI-AIR	
知識の名称	知識の種類
(1) identifier	AIR の識別名
(2) title	学術情報のタイトル
(3) creator	学術情報の作成者
(4) coverage	研究分野名
(5) subject	学術情報のキーワード

報 (或いは実体へのリンク) を情報資源として管理し、自身が管理する学術情報に関するメタデータ等を利用支援知識として持つ (一部を表1に示す)。

4.2 発展検索支援システムの動作

提案システムでは、まず、ユーザの入力したキーワードが User-AIR により取得される。User-AIR は、入力されたキーワードに利用支援知識の項目を付加し、複数のキーワードと検索条件から成るクエリを生成する。クエリ内の各キーワードには優先順位が与えられ、ユーザが入力したキーワードには優先順位1が、他のキーワードには、プロファイルでの優先順位に従い、優先順位2以降が割り当てられる。生成されたクエリは、User-AIR により、全ての AI-AIR へメッセージとして送信され、発展検索代行のサイクルが開始される。

● 基本検索

各 AI-AIR は、クエリ内のキーワードと自身の利用支援知識の項目とを比較し、一致するキーワードの数により、自身の管理する学術情報がクエリに適合するか否かを判断する。

● 選択

1. 要求に適合すると判断した AI-AIR は、互いにメッセージを交換し合い、関連のある (表1中の (1) ~ (5) で一致する項目がある) AI-AIR 同士でグループを構成する。
2. 各グループ内で、各 AI-AIR の重要度 I を計算する。基本検索時のクエリに N 個のキーワードが含まれる場合、重要度は $I = \prod_{n=1}^N P_n$ で定義する。ここで、 P_n は、優先順位 n のキーワードが AI-AIR の利用支援知識の項目と一致しなかった場合のペナルティであり、 $P_1 = 0.80, P_2 = 0.85, P_3 = 0.90, P_{4 \sim N} = 0.95$ である (一致した場合は、優先順位にかかわらず $P_n = 1$)。
3. 選択された AI-AIR の内、重要度 I が上位 max-a 件 (表1中の (c)) の AI-AIR を選択する。

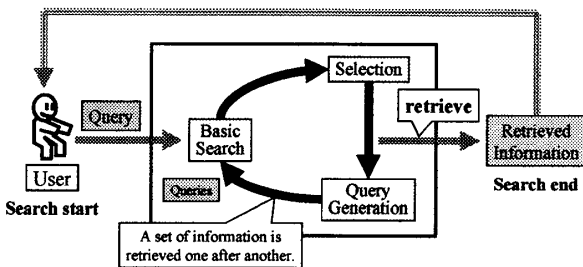


図2 発展検索の代行サイクル

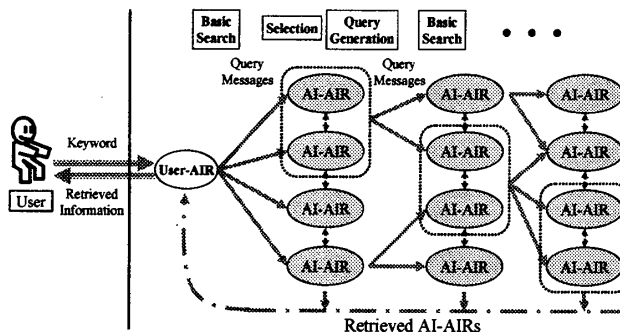


図3 発展検索支援システム

このステップで選択された AI-AIR は、User-AIR ヘメッセージを送信し、自身が選択されたことを伝える。

● クエリ生成

選択された各 AI-AIR は、以下の手順で元のクエリを修正し、新たなクエリを生成する。これを、他の AI-AIR ヘメッセージとして送信することで、次のサイクルを開始する。

1. まず、利用支援知識と一致しなかったクエリ内のキーワードで、最も優先順位の高いものを決定する。
2. 決定されたキーワードの内容を、自身の利用支援知識の内容で置き換える。
2. 置換されたキーワードの優先順位を1とする。

基本検索・選択・クエリ生成のステップから成るサイクルは、既定の回数に達するか、或いは、新たなクエリが生成されなくなるまで繰り返される。サイクルの繰り返しが終了した時点で、User-AIR は、検索結果である AI-AIR をユーザへ提示する。

5. 評価実験

提案システムの有効性を検証するために、試作システムを実装し、これを用いて発展検索を支援する実験を行った。

5.1 実験環境

提案システムを構成する User-AIR, AI-AIR は、エージェントフレームワーク DASH (Distributed Agent System on Hybrid architecture) [7] を用い実装した。これらの AIR は、ルール型の形式で記述された利用支援知識に基づき、Java プログラムとして実装された利用支援機能を起動し、各自に割当てられた処理や他の AIR との協調・連携を行う。実装環境は、PC (Celeron 2.0GHz, memory 768MB) 上の Linux (RedHat9.0) である。検索対象である AI-AIR は、情報検索および情報推薦に関する学術論文 150 件を構造化し作成した。各 AI-AIR の利用支援知識は、今回の実験で使用する項目 (表 1 中の (1) ~ (5)) のみを元の学術論文から抽出し付加している。実験に参加したユーザは、情報科学を専攻する大学生・大学院生 11 名である。

5.2 評価実験

まず、150 件の AI-AIR から興味のあるものを各ユーザに選択してもらい、ここで選択された AI-AIR を各ユーザが行う検索の正解とした。なお、各ユーザが選択した AI-AIR の平均件数は 48.0 件、選択された AI-AIR の和集合は 146 件、積集合は 0 件であった。

次に、各ユーザは、試作システムを用い、以下の 2 つの条件で検索を行った。User-AIR の利用支援知識は、ユーザ毎に事前に設定してもらい、2 つの実験で共通のものを用いた。

● 実験 1 (発展検索支援なし)

ユーザは、ontology, agent, web, network, knowledge の中から 1 つのキーワードを選択する。次に

1. 選択したキーワードをシステムへ入力
2. システムは、基本検索と選択のみを 1 回だけ実行し、選択結果をユーザへ提示
3. ユーザは、提示された AI-AIR の利用支援知識の中からキーワードとして 1 つを選択

1~3 を 5 回繰り返し、発展検索と同等の作業を行う。

● 実験 2 (発展検索支援あり)

1. 実験 1 と同じ初期キーワードをシステムへ入力
2. システムは、基本検索・選択・クエリ生成を 5 回 (或い

表 2 実験結果

	実験 1	実験 2
検索結果 (全件数)	発展検索支援なし 48.4 件	発展検索支援あり 36.4 件
(正解件数)	15.6 件	17.2 件
再現率	34.9%	38.9%
適合率	33.7%	43.8%
検索時間	853.0 秒	97.4 秒

は、新たなクエリが生成されなくなるまで) 繰り返し、検索結果をユーザへ提示

1, 2 により、発展検索をシステム側に代行させる。

2 つの実験の結果 (全ユーザの平均) を表 2 に示す。この結果から分かるように、提案システムを用い発展検索を支援した場合、再現率と適合率が向上できている。これは、発展検索の支援により、直前の検索で得られた情報が適切な手法・タイミングで直後の検索へ利用できたためであると考えられる。また、発展検索を支援した場合、検索結果からキーワードを選択しシステムへ入力する時間を除くことができるため、検索に要する時間が大幅に削減されている。以上の実験結果は、発展検索支援に対する提案システムの有効性を示すものである。

6. むすび

本稿では、ユーザによる学術情報の活用を図るために、学術情報の探索において重要な役割を果たす発展検索に着目し、この作業を効果的に支援する新たな学術情報検索システムを提案した。また、マルチエージェントフレームワーク上で試作システムを実装し、提案システムの有効性を検証する評価実験を行った。その結果から、提案システムを用いることで、発展検索における検索精度の向上と作業時間の短縮が実現されることが確認できた。今後は

- 獲得済みの学術情報からユーザの知識の変化を推測し、推定結果を、3 つのステップ (基本検索・選択・クエリ生成) へ効果的に反映する手法
- 学術情報をグループ化・階層化し、メッセージ交換の範囲をグループ間あるいはグループ内に限定することで、学術情報の大規模な集合を対象とした場合でも発展検索支援を効果的に実現する手法

について、さらに検討を進める予定である。

参考文献

- [1] M. J. Bates, "The design of browsing and berrypicking techniques for the online search interface," Online Review, vol.13, no.5, pp.407-424, 1989.
- [2] 木下 哲男, "分散情報資源活用の一手法 — 能動的情報資源の設計 —," 信学技報, AI99-54, pp.13-19, 1999.
- [3] K. M. Sim, et al., "Toward agency and ontology for web-based information retrieval," IEEE Trans. Syst., Man, Cybern. C, vol.34, no.3, pp.257-269, 2004.
- [4] F. Liu, et al., "Personalized web search for improving retrieval effectiveness," IEEE Trans. Knowledge Data Eng., vol.16, no.1, pp.28-40, 2004.
- [5] K. Chakrabarti, et al., "Evaluating refined queries in top-k retrieval systems," IEEE Trans. Knowledge Data Eng., vol.16, no.2, pp.256-270, 2004.
- [6] 宮寺 庸造 他, "学術論文関係情報のグラフ描画問題に基づく視覚化手法," 信学論 (D-I), vol.J87-D1, no.3, pp.398-415, 2004.
- [7] "DASH — Distributed agent system based on hybrid architecture," <http://www.agent-town.com/dash/>.