

F-030

能動的情報資源を用いた学術情報検索支援システムの設計 Design of Academic Information Retrieval Support System Using Active Information Resource

千葉 祐[†]
Yuu Chiba

阿部 亨[‡]
Toru Abe

木下 哲男[‡]
Tetsuo Kinoshita

1. はじめに

近年のインターネットの普及に伴ない、学術論文、貴重図書、史料など、各種学術情報の電子化が進んでいる。これら電子化された学術情報の活用を支援するために、キーワード型やブラウジング型の検索システムが数多く提案/開発され、電子図書館等で利用されている。しかし、現状では、それらの検索システムを用いたとしても、所望の学術情報を見付け出すまでには利用者が多くの作業を行う必要があり、学術情報を利用する際の利用者の負担は依然として大きい。

本稿では、電子化された学術情報を利用する際に生じる上記の問題の解決を図るために、まず、学術情報に対する検索形態の分類を行う。次に、各検索の効果的な支援を実現するために「能動的情報資源 (Active Information Resource: AIR)」[1] の概念を用いた検索システムを提案し、その具体的な設計/動作について検討する。

2. 学術情報の検索

2.1 従来の検索システム

現在、キーワードに基づく従来の検索に加え、Dublin Core [2] に代表されるメタデータ (書誌情報など) に基づく検索を利用したシステムが提供されている。また、概念の階層構造であるオントロジー [3] や、シソーラス [4] を用いた検索システムなどの研究も進められている。これらは、利用者からの明示的な入力に対する検索精度の向上を目指したものである。しかし、明示的な入力に対する検索精度を向上したとしても、検索結果から最終的に情報を選択する作業、あるいは、検索結果に基づいて新たな検索を行なう作業等は、利用者に委ねられたままになる。これに対し、利用者に提供する情報のフィルタリングや、情報間の引用関係を利用者に提示 [5] することで、利用者の検索作業の一部をシステム側で代行する研究も進められている。しかし、情報の取得後、利用者の知識構造は変化するはずであり、これに伴う要求の変化を検索に効果的に反映する手法については十分な検討が行われていない。

この問題は、従来の検索システムが、利用者による学術情報検索の流れを十分反映しておらず、検索の流れを効果的に支援する機構を実現していないことに起因する。

2.2 学術情報の検索形態

何らかの検索システムを用い利用者が学術情報を検索する場合、その処理の流れは、図 1 に示すようになる。berrypicking model[6] でも言及されているように、一度の検索で得られた情報により、利用者の持つ知識構造が変化し、新しいニーズが生まれる。このニーズを満たす

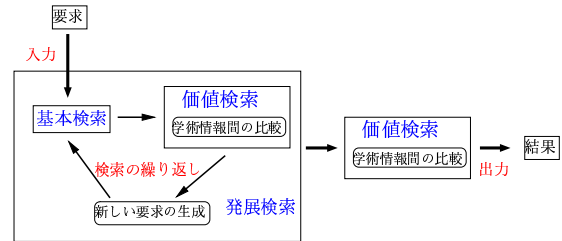


図 1: 学術情報検索の流れ

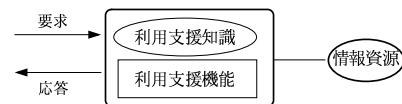


図 2: AIR の構成図

ため、新たな検索要求が生成され、検索が繰り返される。特に学術情報は、論文調査を始めとする研究活動に利用されることが多く、利用者の知識構造の変化も非常に活発であり、検索要求の再検討、再検索を行う場合が多いといえる。図 1 に示す学術情報検索の流れは、複数の検索プロセスから構成されており、各々以下の特徴を持つ。

基本検索 利用者の明示的な要求と、各学術情報が持つキーワードやメタデータとの比較により、要求に合致する学術情報の検索を行う。

価値検索 主に、利用者の状況、状態やシステムの利用履歴などを暗黙的な要求として用い、検索結果の順位付け/絞り込みを行う。

発展検索 検索結果から得られた情報 (選択された学術情報の参考文献やメタデータ等) に基づき、新たな要求を生成し再度検索を繰り返す。

これらの検索プロセスをシステム側で代行することで、利用者による検索作業を包括的に支援することが可能となる。発展検索を実現するためには、利用者の知識構造の変化をシステムの動作に反映させる必要がある。そこで、提案システムでは、利用者の知識構造の変化を、利用者が取得した学術情報に付加された知識に置き換える。また、検索要求を再検討する範囲を関連性の強い情報に限定することで、発展検索の一部の自動化を図る (検索要求の大幅な方向修正が必要な場合は利用者からの入力を求める)。書誌情報が明確な学術情報は、検索要求を自動生成するために必要な情報を取得しやすく、提案システムが特に有効であるといえる。

[†] 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

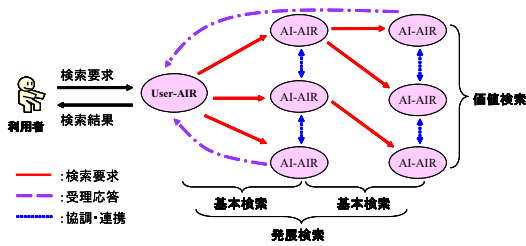


図 3: 提案システムの概念図

3. 能動的情報資源

前章で述べた発展検索では、検索結果から複数の学術情報が選択された場合、各学術情報毎に新たな要求を生成し再度基本検索を実行する必要がある。従って、システム側で発展検索を効果的に代行させるためには、各学術情報が能動的 / 自律的に動作する機構の導入が有効であると考えられる。本稿では、能動的情報資源 (AIR) の概念を学術情報に適用することで、この機構の実現を図り、学術情報検索の流れ全体を効果的に支援するシステムを提案する。

AIR は、ネットワーク上に分散した情報資源 (電子化された各種データ) の有効活用を図る手法である。この手法では、利用支援知識と利用支援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 (情報資源の AIR 化)。AIR の構成図を図 2 に示す。これにより、情報資源自体が能動性 / 自律性を持つことになり、今まで利用者へ全面的に委ねられていた情報資源に対する操作の一部を情報資源自身が代行できるようになる。

4. 学術情報検索支援システム

本稿で提案する学術情報検索支援システムは、利用者とのインターフェースとして振舞う User-AIR と、学術情報を各々 AIR 化した Academic Information-AIR (AI-AIR) から成る。提案システムの概念図を図 3 に示す。利用者からの要求は、まず、User-AIR を通じて検索対象である AI-AIR に送信される。これを受信した AI-AIR は、要求へ対応可能か否かの判断、自身が有する利用支援知識 / 学術情報の操作、他の AI-AIR との協調 / 連携を行うことで、利用者からの要求に基づく検索作業を代行し、その結果を利用者へ送信する。また、利用者の状況、状態をユーザプロファイルとし、利用者からの暗黙的な入力として用いる。

4.1 User-AIR の設計

User-AIR は、利用支援知識としてユーザプロファイルを持ち、また以下の利用支援機能を持つ。

- F1: ユーザプロファイルを作成 / 更新する機能
- F2: 利用者からの要求に基づき検索を開始する機能

各利用者は、自分が興味を有する研究分野毎に、表 1 に示すユーザプロファイルを持つ。これらのユーザプロファイルは、利用者の初期設定および活動状況に基づき、User-AIR が有する利用支援機能 F1 により作成 / 更新される。

表 1: User-AIR の利用支援知識

知識の種類	主たる記述属性の名称
K1: 研究分野名	research-area
K2: 研究分野におけるキーワード	keywords
K3: 研究分野に対する習熟度	level
K4: 学術情報にアクセスする権限	authority
K5: 一度の検索に費せる時間の上限	time
K6: 一度の検索で表示する情報の上限	quantity

表 2: AI-AIR の利用支援知識

知識の種類	主たる記述属性の名称
K7: AIR の名前, 作成者	title, creator
K8: AIR の属する研究分野名	research-area
K9: AIR の持つ特徴 (キーワード)	keywords
K10: 参照している AIR	reference
K11: 参照されている AIR	referenced-by
K12: 類似している AIR	similar
K13: 参照している AIR の優先度	priority
K14: 類似した AIR の類似度	similarity
K15: 他の AIR からのアクセス履歴	access
K16: AIR にアクセスする権限	authority

利用者からの検索要求は、User-AIR の F2 により、次の項目から成るメッセージに変換される。

- input-keyword: 利用者から入力されるキーワード
- importance: 要求の持つ重要性
- research-area: (ユーザプロファイルにより決定)
- time: (ユーザプロファイルにより決定)
- level: (ユーザプロファイルにより決定)
- authority: (ユーザプロファイルにより決定)
- keywords: (ユーザプロファイルにより決定)

メッセージ中、input-keyword は、利用者からの明示的な入力により決定され、importance は、初期値 1.00 を持つ、その他については、ユーザプロファイルから得られた情報を暗黙的に用いることで決定される。このメッセージが AI-AIR へ送信されることで検索が開始される。

4.2 AI-AIR の設計

AI-AIR は、表 2 に示す利用支援知識を持つ。また利用支援機能として

- F3: 自身の活動履歴を保存するために、自身の参照関係、アクセス履歴等の利用支援知識を更新する機能を有する。さらに、基本検索を実現するために
- F4: 検索要求と自身が持つ利用支援知識とを比較し、検索要求に回答すべきか否かを判断する機能を持ち、価値検索を実現するために
- F5: 関連ある AIR 同士でグループを形成する機能
- F6: 形成されたグループ内の学術情報に対し順位付け / 絞り込みを行う機能

を持つ。また、発展検索を実現するために

F7: 自身が受信した検索要求と自身が有する利用支援知識から、新たな検索要求を生成する機能

F8: 新たに生成した検索要求を他の AIR へ送信する機能を有する。機能 F7 は入力要求に含まれる重要度 (importance), 表 2 に示した K8 ~ K14 から以下のように新しい要求を生成する。

1 変化しない項目

- input-keyword, level, authority

2 変化する項目

- importance = importance(入力) × (K13 or K14) × research-area, keywords の比較結果 (1.00 or 0.90)
- time = time(入力) - 1
- research-area, keywords = K10 ~ K12 中で最多のもの

検索後、一度取得した AI-AIR は、利用者に特化した知識を付加し利用者毎に蓄積する。利用者の知識構造の変化を、各 AI-AIR に付加された知識に置き換え、これを新しい要求を生成する機能に反映させる。これにより、知識構造の変化を自動的に検索要求の変化に対応させることができる。

4.3 検索支援システムの動作例

提案する学術情報検索支援システムの具体的な動作例について、各検索プロセスでの動作を中心に説明する。

基本検索

Step1 User-AIR(あるいは、他の AI-AIR)からの検索要求を受信した AI-AIR は、利用支援機能 F4 を用い、自身がその要求に回答すべきか否かを判断する。

Step2 回答すべき(自身が要求に該当する)と判断した AI-AIR は、その結果を User-AIR へ送信する。

価値検索

Step1 基本検索の結果を受信した User-AIR は、回答を返したすべての AI-AIR に対し価値検索の開始を送信する。

Step2 価値検索の開始を受信した AI-AIR は、各自の利用支援知識を送受信し、F5 を用いて相互に協調/連携を行い、関連する AI-AIR 同士で自律的にグループを形成する。

Step3 各 AI-AIR が F6 を用いることで、ユーザプロフィールの情報および各自の利用支援知識に基づき、各グループ内で学術情報の順位付け/絞り込みを自律的に行う。

Step4 各グループのマネージャは検索処理の状態を判断し、発展検索実行時の場合は、順位付け/絞り込みを行ってから引き続き発展検索のステップへ移行する。発展検索終了時の場合は、User-AIR へ結果を送信し、利用者にその結果が送信される。

発展検索

Step1 User-AIR は、順位付け/絞り込みの結果に基づき、検索結果から中上位の AI-AIR を選択し、発展検索の開始を送信する。

Step2 発展検索の開始を受信した AI-AIR では、検索要求と自身の利用支援知識とに基づき、F7 により新たな検索要求を生成し、F8 を用いることで対象となる AI-AIR へ検索要求を送信する。

Step3 以降、importance の値が閾値以上、かつ、time の値が 0 より大である限り、基本検索、価値検索、発展検索の処理が繰り返される。

以上により、利用者の検索要求に対して関連性の強い情報の検索を自動化する。しかし、提案システムでは、検索要求の大幅な方向修正があった場合は、利用者の入力が必要とする。これは、利用者が、提示した結果から学術情報を選択し、その結果に対して新しい検索要求を入力することで再度発展検索を実行し実現する。

5. おわりに

本稿では、利用者の知識構造の変化による要求の変化を効果的に検索に反映するため、学術情報の検索の流れを、複数の検索プロセス(基本検索、価値検索、発展検索)が組み合わされたものと捉え、この流れ全体を効果的に支援するための学術情報検索支援システムを提案し、その設計/動作について検討を行った。提案システムでは、各学術情報を AIR として構成することにより、各検索プロセスを学術情報自身が能動的/自律的に代行させている。このため、学術情報の検索に際し、従来、利用者に課せられていた負担の大幅な軽減が期待できる。

現在、エージェントフレームワーク ADIPS / DASH [7] を用いて、エージェント型の AIR に基づく学術情報検索支援システムの実装を行っている。これをもとに、AI-AIR 間での協調プロトコルや利用支援知識を拡充し、新規情報資源の生成や流通を含めた学術情報の利用支援機構に関する検討を行う予定である。

参考文献

- [1] 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.
- [2] Dublin Core Metadata Initiative, “Dublin Core Metadata Initiative,” <http://dublincore.org>
- [3] P.Weinstein et al., “Agent-Based Digital Libraries : Decentralization and Coordination,” IEEE Communications Magazine, Vol.37, No.1, pp.110-115, Jan.1999.
- [4] 大山 敬三 他, “大規模学術情報データベースに適した情報検索システムの開発,” 電子情報通信学会論文誌, Vol.J84-D-I, No.6, pp.658-670, 2001.
- [5] 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,” 電子情報通信学会論文誌, Vol.J87-D-I, No.3, pp.398-415, 2004.
- [6] Marcia J.Bates, “The design of browsing and berrypicking techniques for the online search interface,” Online Review, Vol.13, No.5, pp.407-424, 1989.
- [7] DASH, “DASH - Distributed Agent System based on Hybrid architecture,” <http://www.agent-town.com/dash/index.html>