O-018

研究者ニーズに対応した多層式情報検索手法の提案
A Research Support System based on Multi-leveled Information Retrieval

松尾徳朗† 藤本貴之‡
Tokuro Matsuo and Takayuki Fujimoto

1 はじめに

近年，国内の研究者情報や研究内容に関する情報を公開・共有化を促進するため，国立研究機関を中心に様々なサービスが開始されている．例えば，独立行政法人科学技術振興機構による国内の研究機関・研究者情報・研究課題等の検索サービス“ ReaD（研究開発支援総合ディレクトリ）”[1] や，学術論文・図書雑誌，研究成果概要などの学術情報の統合的な検索サービスである国立情報学研究所による“ GeNii（Nii 学術コンテンツ・ポータル）”[2] などが挙げられる．

これら情報は学術情報に関して，きめ細かで適切な処理がなされ，閲覧や検索に十分に配慮したデザインとなっている．しかし，これらはあくまでも，科学技術振興機構や情報学研究所といった機関により，研究者・研究機関から自主的に提供される情報にのみでデータベースが構成されており，情報の適切さに対して，その網羅制や量的な充足度は決して十分とは言えない．研究者が学術情報を検索する時に必要となる要素は，これら公的機関が提供する「少ないが的確な情報」と，いわゆるサーチエンジンなどによってはじき出される「機械的だが大量な情報」という２つの要素を同時に満たすものでなければ十分とは言えない．そこで，本論文では，研究者がインターネットを用いて学術情報を検索する場合に必要となる２要素（質的情報，量的情報）を相補的に維持しつつ検索を可能とする手法およびシステムについて提案する．

提案手法を用いることで効果的な研究者育成および研究支援が可能となる．提案するシステムは，大学において研究を始めたばかりの学生を支援するだけではなく，ある研究分野の転向や新たな分野における研究を試みる研究者にとって効果的に利用されることも可能である．一般に，ある領域の研究に関する研究論文の推薦などに関する研究は多く存在しているが，一般に多くの研究論文は細分化および専門化されており，ある研究分野の概要や全体像を知るには適さない．さらに，既存の論文推薦システムの多くが，推薦機構における計算プロセスで，ある分野に関して多くの論文を発表している研究者のウェイトが極端に大きくなり，結果的に一つの研究に関する入力値（キーワード）に対して，狭い分野の研究論文が提示されることが多い．従って，既存のシステムを利用しようとするユーザにとっては，研究に対する狭い知識のみが与えられる．その結果，ある分野において研究を始めようとする研究者にとって研究の広がりの可能性が狭められる．一方，提案するシステムを用いることで，情報量に関する密度および濃度を保ったまま，数だけを減らすことができ，特に当該分野の研究に対して初心者であっても，効果的に研究情報／学術情報を得ることができる．

近年，情報学の分野において世界的に重要視され，指摘され続けている爆発的に増加する web 上の情報に対する「取捨選択」および情報や知識の「再利用・統合」を，本論文では具体的な手法において解決しており，情報の効果的な利用法に関する方法論に対しても貢献している．情報の効果的な取捨選択と再利用に関して，本論文で提案した多層的な情報検索手法は革新的な試みであり，研究者支援における利用は研究者だけではなく，科学の発展という意味でも提案するシステムは重要な役割を担っている．

† 宮城大学事業構想学部，‡ 園田学園女子大学未来デザイン学部

2 多層的な情報検索の提案

研究者が研究情報についてインターネットを用いて検索する場合に求められる「質的情報」と「量的情報」を相補的に充足させる手法について提案する．本論文において採用する「質的情報」とは“ ReaD（研究開発支援総合ディレクトリ）”や GeNii（Nii 学術コンテンツ・ポータル）”といった公的機関によって提供される一定のフォーマットを用いて適切な配置がなされた信頼出来る情報源を指す.「量的情報」とはロボット型検索エンジンによって行われる全文検索による情報を指す.

本研究で提案するシステムでは，特殊な機器や環境を必要としないことを特徴としている．既存に公開されているデータベースを母体として情報検索を行う．今回は「質的情報」源として ReaD を,「量的情報」源として Google[3] を用いた．本研究は既存の情報資源を活用することで，情報ソースの開発・備蓄にではなく，その格納された情報をいかに引き出し，提供するかという手段の開発に着目した．

本研究は「量的情報」と「質的情報」を相補的に検索・照会しつつ結果を提示させることで，複雑になりすぎる／あるいは簡単に成りすぎてしまう情報を，検索を行う研究者の目的にあった形で出力することが可能である．本システムのフローを図 1 に示す．検索手法を相補参照的に多段化し，情報掲示を多層化することで，ユーザはより目的の情報の取得を容易にするだけではなく，関心はあるものの，意識の外にあった「意外な情報」をも取得することができる．本システムではいわゆるキーワード検索だけではなく，より複雑な形式での検索が可能である．例えば，文書や新聞記事，論文の一説など，ある程度まとまった分量のテキストを検索にかけることが可能となる．入力されるテキスト情報は常にシステム内の形態素解析モジュールによって品詞分解され，必要名詞のみを抽出した形で，検索キーワード化をすることができる．

情報検索を多層化することで，ロボット検索による量的情報は，質的情報を照会した形で絞り込まれて掲示されるため，量的情報の曖昧さは大きく削減される．また，システム内で既にキーワードの絞り込みがなされた後に検索がかけられ，その結果も再び質的情報の参照対象としてその精度を高めることができる．インターネットによる情報検索は，無尽蔵に増殖するという欠陥を有するが，本システムを用いることで，情報が無尽蔵に増殖する情報を常に質的情報というガイドラインによって限定させることができる．そのため，ロボット検索などによる量的情報空間の広さを曖昧化させることなく，常に質的情報を検索するペースで利用することが可能となる．また，本システムでは，検索結果の重み付けを行い，その重要度がもっとも低いとされる情報も絞り込みの対象とする．最下位の Keyword を絞り込み対象として抽出する理由としては，複数の Keyword(s) が存在する場合，予想外な語句やコメント，門外漢な発言といったものが，概して発想のブレイクスルーに繋がるという点に着眼したためである [4]．本研究は離散的に増殖する量的情報の質的情報化への置換，また，隠蔽情報の顕在情報化を実現することで，研究者による情報検索をより簡易且つ的確なものにすることを射程とする．

3 システムの概要

本研究で提案するシステム概要について述べる．

まず，ユーザは自分が調べたい任意の Keyword(s) をシステムポータルから入力する（1）．入力される Keyword(s) は，

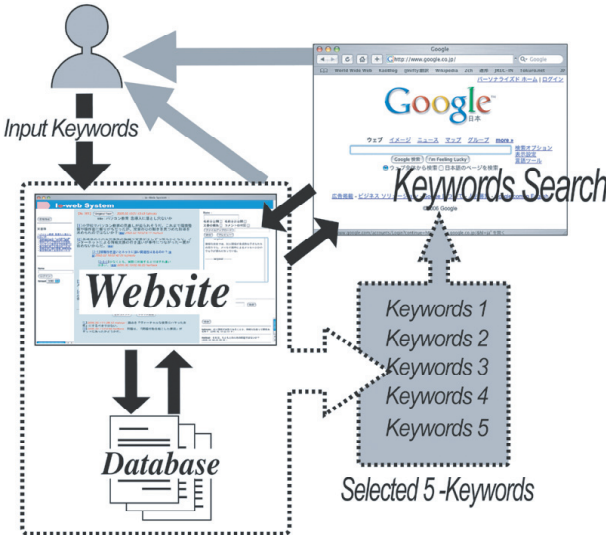

図 1: システムの概要

図 2: 実行例

単語であっても文章でも構わない．入力された，Keyword(s) が単語である場合は，そのまま単語のまま，文章である場合は，システム内の形態素解析モジュールによって品詞分解され，名詞のみを抽出する．なお，形態素解析には茶筌 (Chasen)[5] を用いる（2）．抽出された Keyword(s) はロボット型検索エンジンを介して検索が行われる．ロボット型検索エンジンには，15 億ページ以上を登録すると言われている Google を利用する（3）．Keyword(s) による検索結果をもとに，Keyword(s) の重み付けが行われる．重み付けは純粋に検索ヒット数による．重み付けされ，抽出される Keyword(s) は 5 -words に絞られる．Keywords が 5 -words 以下の場合はそのまま，6-words 以上の場合には，上位 4-words と最下位の 1-word が抽出される（4）．これら抽出された Keyword(s) を検索語として用いて，再度のポータルを通して再度の検索をかけるが，この時の検索は Google と ReaD の 2 つの検索サービス両方に対して相補的に行われる（5）．Google から検出される膨大な結果は，常に ReaD のデータベース・検索結果へと照会される．すなわち，システムポータルに示される検索結果は三層構造となる．本システムの検出結果を図 2 に示す．まず，第一層目として，抽出された Keyword(s) が並び，その下階層に，ReaD による検索結果が示される．ReaD による検索結果それぞれが Google の検索結果につながっている．また，出力される Google の結果を再び検索対象としてシステムにフィードバックさせることも可能である．

4　関連研究と提案手法の有効性

学術情報に関する関連研究を示す．単に情報を備蓄し，検索対象とするといったものではない研究者・学習者を対象としたユニークなポータルに関する研究としては，札幌医科大学附属図書館による“ PIRKA ”システムなどが挙げられる [6]．これは，米国国立医学図書館“ IAIMS ”[7] をモデルに提案・構築された異種データベース間連携システムとして注目される．ここでは，格納される情報にではなく，既存のデータベースからいかにユーザの目的に適した情報を的確に提示するかという手法に着目されている．しかし，本論文の目的は情報やデータの整理ではなく，「量的情報」を「質的情報」源にフィードバックをかけること，その量的増殖が抱える情報の離散化を抑制するという側面からのアプローチを行った．

研究者間の関係をグラフとして表現し，グラフ上のアークにおいて研究者間の結びつきの性質を分析することが出来るシステムとして松尾らの研究がある [6]．しかし，そこではある特定の分野における研究者の関係が示されるのみで，具体的な研究支援や研究に関するキーワードに基づいた情報の提示は行われない．

5　まとめ

研究者がインターネットを用いて学術情報を検索する場合に必要となる 2 要素（質的情報，量的情報）を相補的に維持しつつ検索を可能とする手法およびシステムについて提案した．情報検索を多層化することで，ロボット検索による量的情報は，質的情報を照会した形で絞り込まれて掲示される．従って，量的情報の曖昧さは大きく削減される．システム内で既にキーワードの絞り込みがなされた後に検索がかけられ，その結果も再び質的情報の参照対象としてその精度を高めることができる．このことにより，ある領域の研究に関して，狭い分野に関する検索結果の提示ではなく，情報量の密度および濃度を保ったまま，数だけを減らすことができ，効果的に研究情報／学術情報を取得することができる．本システムは研究室における学生の教育や，適当な共同研究者を探す目的にも利用可能である．今後，主として研究者育成に注目したシステムの改良を行う．

参考文献

[1] “ ReaD（研究開発支援総合ディレクトリ） ” http://read.jst.go.jp/

[2] “ GeNii（Nii 学術コンテンツ・ポータル） ” http://ge.nii.ac.jp/genii/jsp/

[3] http://www.google.co.jp/

[4] 西本一志, 安部伸治, 宮里勉, 岸野文郎, 発散的思考支援を目的とする関連性と異質性を併せ持つ情報の抽出手法の検討, 人工知能学会論文誌,Vol.11,No.6,pp896-904,Nov.1996

[5] 茶筌　http://chasen.naist.jp/hiki/ChaSen/

[6] 医学図書館と統合型学術情報システム IAIMS, 医学図書館, Vol.34, No3, p.190-200, 1987

[7] 松尾 豊、篠田 孝佑、中島 秀之：中心性に着目した合理エージェントのネットワーク形成、人工知能学会誌、Vol.21, No.1, pp.122-132, 2006