

知的な情報検索行為における思考過程モデル

5 Y-6

○亀田 弘之[†]川浪 章欣[†]藤崎 博也[‡]東京工科大学工学部[†]東京理科大学基礎工学部[‡]

1. はじめに

現在、インターネットをはじめとする情報ネットワーク環境における情報検索が注目を集めているが、必要な情報を漏れなく迅速かつ的確に検索することは、場合によっては極めて困難であり、実際、情報検索の専門家としてサーチャと呼ばれる有資格者がいるほどである。このような状況に鑑み筆者らは、知的な情報検索システムの研究に従事している[1]。

本稿ではそのうち、検索を思考過程の一つと捕え、情報検索における思考過程モデルについて提案する。なお以下の議論では、情報検索の対象をインターネット上のホームページ（以下、HPとも記す）に限定するものとする。

2. インターネットにおける従来の情報検索方式

インターネットの World Wide Web（以下、WWW）上における情報検索には、以下に示す2つの検索方式が通常利用されており、それぞれ下記のような長所・短所がある[2]。

(1) **ディレクトリー階層方式**：この方式では、リンク集作成者が、HPの内容を調べ各カテゴリー別に分類しているため正確であるが、世界中に散在する情報すべてを漏れなく、また偏りなく記録することが実際上できないため、リンク情報が必ずしも多くなく、また、網羅的ではない。さらには、日々更新されるHPもあるため、リンク情報の新鮮さを保つことが困難である。

(2) **キーワード入力方式**：この方式では、最近、自動的かつ定期的にHPのテキスト情報を収集しているため、情報量は極めて豊富であり、HPの内容が更新されても定期的に修正が加えられるといった利点があるが、その反面、入力されたキーワード（以下、KWとも記す）の文字列だけで情報収集・検索するため、同表記異義語で記述された検索結果が出力されてしまうといったことや、異表記同義語を利用して検索することができないため検索漏れが生じるといった問題がある。

3. 知的な情報検索に必要な諸機能

先に示した問題点を考慮し、知的情報検索には文理解、未知語処理、意図理解、キー概念検索、知識獲得の諸機能が必要である。

(1) **文理解機能**：検索者からの検索要求の入力（日本語文）を対話的に処理する対話文理解処理機能と、テキスト情報（日本語文）からなるHPを処理するための文書理解機能。言語表現からキー概念（以下、KCとも記す）やKWを抽出したり、検索者の個人情報抽出したりする。

(2) **未知語処理機能**：検索者とシステムとの対話文やHPに記載されている文には、システム辞書にはない表現（単語）が出現する可能性がある。先の文理解において、システム辞書に未登録な単語（未知語）が入力された場合、それを処理（検出・意味推定等）し、文理解を円滑に進める機能。

(3) **意図理解機能**：検索者が検索対象を陽に言語表現化できないとき、検索者の言わんとすること（意図）を、談話情報等から推定する機能。

(4) **キー概念検索機能**：KWの担う概念（意味）を抽出し、概念を介して検索する機能。キーワード入力方式の検索システムの場合のように、KWの文字列一致を主とする検索とは異なり、異表記同義語や、同表記異義語の存在によるノイズや検索漏れを回避することができる[2]。

(5) **知識獲得機能**：検索要求入力文における未知語や検索者の語用論的な癖、意図理解に使われる対話履歴、検索経験により得られる、WWW上のさまざまなHPに関するメタ的情報・知識、検索に有効なKW・KCの組合せに関する知識等を獲得する機能。この機能を充実させることで、処理経験を検索システムが自律的かつ知的に有効利用することができる。

4. 知的な情報検索における思考過程モデル

前記の知見に基づき知的な情報検索における思考過程モデルを図1に示す。このモデルでは、リンク集を検索する検索エンジン（IFA; Interface Agent とIRA; Information Retrieval Agent とから構成されている）と、リンク集にデータを蓄積する蓄積エンジン（ICA; Information Collect Agent により構成されている）とにより、知的情報検索システムが構成されている。

A Model of Thought Process in Retrieving Information Intelligently
† Hiroyuki KAMEDA, † Akiyoshi KAWANAMI, and ‡ Hiroya FUJISAKI

† Tokyo Engineering University 1404-1 Katakura, Hachioji, Tokyo 192-8584.

‡ Science University of Tokyo 2641 Yamazaki, Noda, Chiba, 278

まず検索エンジンでは、対話インタフェイスを介して、検索者から検索要求を日本語文で受け取り、それを対話文理解に引き渡して処理する。対話文理解内部には、未知語処理機能を備え持つ、形態素解析・統語解析・意味解析・意図解析があり、入力文を受け取ると、検索対象に関するKWとKCを抽出する。抽出されたKWとKCとの妥当性を確認するために、対話インタフェイスを介して検索者にこれらを呈示する。検索者がこれらが妥当であると判断するれば、検索を実行し、そうでなければ繰り返し対話を行い、検索要求の内容を明確に確定する。

次に、上述の妥当なKWとKCとをKW生成に引き渡し、実際の検索に利用するKW候補を作成し、それらからなる検索式を作成し、実際にリンク集の検索を行う。

その結果、リンクデータが選出されるが、これらは、選定処理に引き渡され、対話インタフェイスを介して、HPアドレスの重複や要求資料内容の確認を検索者に行ってもらい、検索結果が多いときは、呈示順序を考慮しながら検索者に検索結果を呈示し、必要なものを選定してもらい、この際、検索結果がすべて不適切な場合は、再び検索者との対話からやり直すが、次回の処理時により有効な処理が行えるように、対話内容の履歴や要求資料の内容等を個人情報としてシステムに登録する。

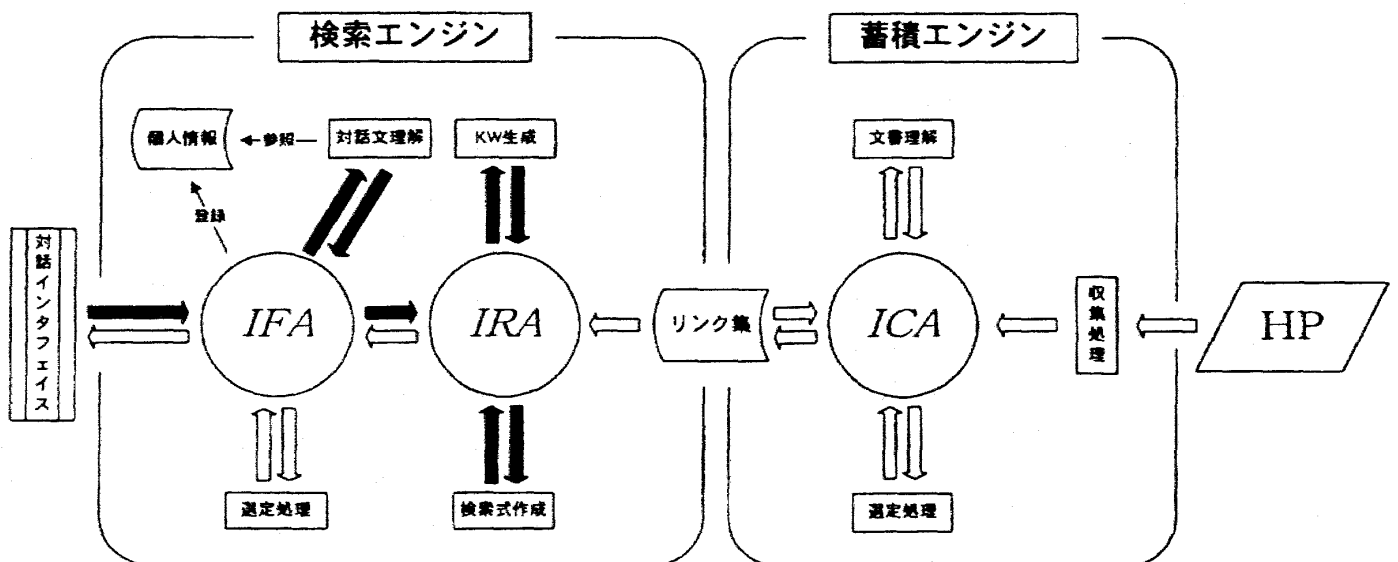
一方、蓄積エンジンでは、収集ロボットにより収集処理を自動的に処理させることで、世界中に散在するHPのテキスト情報を収集し、その内容を文書理解に引き渡し、テキスト情報からKWとKCとを抽出し、リンク集に蓄積する。

5. おわりに

情報検索における思考過程モデルについて述べた。なお本研究の一部は、文部省科学研究費補助基盤研究(B)(1)「情報ネットワーク環境における知的情報検索システムの構築」と日本学術振興会未来開拓学術振興研究推進事業研究プロジェクト「音声言語による人間-機械対話システムの研究」との援助による。

<<参考文献>>

- [1] 亀田・川浪・藤崎：“情報検索における思考過程モデルとそれに基づく知的情報検索プロトタイプシステム的设计,” 電子情報通信学会技術研究報告 TL97-11, pp.1-8(1997).
- [2] 亀田・藤崎：“テーマ・キー概念・キーワード間の階層構造を利用する新聞記事情報の分類・検索システム,” 情報処理学会論文誌, Vol.28, No.11, pp1103-1111(1987).



(注) **⇒** : 要求の流れ KW : キーワード IFA : Interface Agent ICA : Information Collect Agent
⇔ : 情報の流れ HP : ホームページ IRA : Information Retrieval Agent

図1 知的な情報検索における思考過程モデル