

卷頭言

関根智明*

昨年の11月号がソフトウェアを主題とする特集号であったのに対して、本年は情報検索（IR）に関する特集号を出すことになった。

情報検索という日本語はインフォメイション・リトリーバルと片仮名でも書かれるが、もともとInformation Storage and Retrieval すなわち“情報の貯蔵と検索”的である。当然、貯蔵されたものがあつてはじめて検索できることは誰しも異論あるまい。

インフォメイション・リトリーバルに対してもうひとつドキュメンテーションという言葉がある。ドキュメンテーションとは「資料の提供、再生、配布、収集蓄積、主題分析、組織と検索を通じて記録された専門知識の利用を容易にする技術である」といわれている。したがって、IRはドキュメンテーションの一部分ということになる。しかし、実際には、伝統的なドキュメンテーションとよばれるものが、特に電子計算機の高速な処理能力、高度の論理判断とに影響されて、より能動的なものに変化したもののが現在IRとよばれているものなのではないかと思われる。

情報検索という狭い意味での作業が機械化されるためには、そのための情報の貯蔵方法が必要となるし、検索したものを配布するための方法も機械化される必要が生れてくる。これらの一貫した体系がIRシステムとよばれるものであり、またそれを目標として個々の分野の研究が進められているのが現状である。

たとえば伝統的な情報センターと近代的な情報センターの特色を次表のようにまとめると、この間の事情

	伝統的なセンター	近代的なセンター
目的	貯蔵本位 非営利的 保管本位	利用本位 営利的 内容本位
価値能率効率	保管年数 費用節約 管理	利用度 利益増大 情報網羅
	受動的	能動的

* Tomsharu Sekine (Dept. of Administration Engrg Faculty of Engrg., Keio Gijuku Univ.)
慶應義塾大学工学部管理工学科

はより明確になるだろう。

このように、本号の目的は近代的な意味での機能的なIRシステムを解析することにある。もちろん種々様々なものが含まれるし、その一つ一つがとても簡単に扱えるしろものではないので、1冊の特集号でこの主題を完全に論じ尽すことなどはどうていできることではない。だから本号では、とくに重要と思われる基本的な面を取り上げ、しかも現時点で特に問題になっている点に重点を置いて、一つの展望を試みることを執筆方針とした。したがって“IRのすべて”を本号に期待されることは困るが、IRを考えるための方向づけには十分役立つことと期待してよからう。

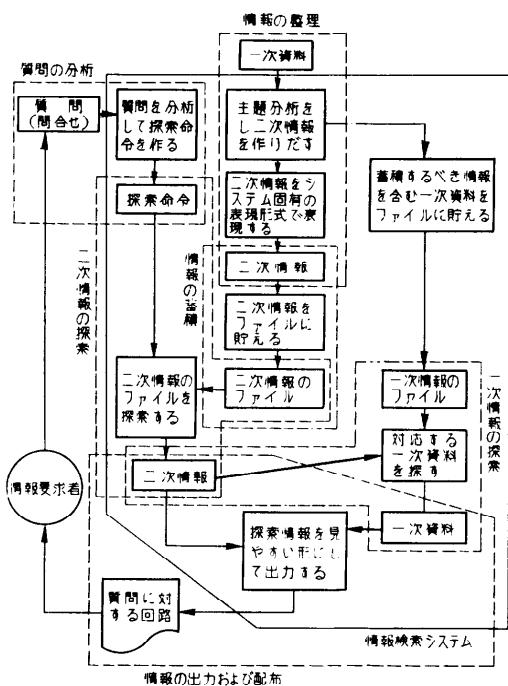
ここで、特に次のことを注意しておいた方がよいようと思われる。すなわち、ここでいう情報(information)といわゆる諜報(intelligence)とを混同してはならないということである。とりたててこのような注意を行なうのは、情報検索が現在主力をおいているのはどちらかといえば文法論的(syntactic)な面で、決して意味論的(semantic)な面ではないからである。意味論的な情報検索が実現されるためには“機械翻訳”的の完成を持たなければならない。この特集号でもそのような話題を用意したが、都合で次号に延期されることになった。もちろん現在までの“機械翻訳”的の成果が無視されているわけではない。たとえばSMARTシステムをその例としてあげができるだろう。

このような点からまた次のことを注意しておこう。現在ではIRを次の二つに大別するのが普通である。

1. ドキュメント・リトリーバル
2. ファクト・リトリーバル

すなわち、前者は所望の情報を含む資料を検索することをいい、後者は質問に対する内容的な回答そのものを検索することをいい。上で述べたように、現在時点での中心課題はドキュメント・リトリーバルであつてファクト・リトリーバルではない。もちろん自然言語であたえられた初等算術問題を解くシステムやS. Cooperのシステムはこのような試みの例である。本特集号ではドキュメント・リトリーバルに限って解説することになっている。

さて IR システム全般について一とおり話をしておくことにしよう（図参照）。



まず原始資料を集め、蓄積し、後刻の検索の便を考えて適当な見出しを作り、システムに適当した形で保管する。一次資料の保管、二次資料の保管、それぞれ問題があるが、このとき“情報内容の処理”が問題となる。ここでは資料の見出しを作るための技術を、“内容分析と主題分析”という名のもとに抄録法、主題索引法、分類法などについて、さらにこれらのコード化についてシソーラス、分析合成関係、分類法に及ぶ説明がある。最後にこれらの実用的な効果を測定するための参考として検索効率測定の実験が紹介されている。

次に、主として計算機を中心として情報の蓄積、すなわち“ファイルの蓄積と検索”が取りあげられる。ここでは各種の検索方式に関する説明、ある種の構造を表現するためのリスト構造などの説明、最後に特に化学化合物の IR で主役を演ずるグラフの照合法を論

づる。

第3には、現在実用されている“IR システム”について説明する。KWIC, KWOC システムはすでによく知られているとおりのものであり、国産機でかなり用意されている。利用者と IR システムとの交流をはかるもの一つの例として資料の配布まで考えた SDI システムがある。ある意味では IR システムの各種のテクニックのテスターとも考えられる SMART システム、人事情報の検索システムである SIR システム、化学情報検索システムとして CAS、医学情報検索システムとしてすでにかなり実用の段階に入っている MEDLARS システムなど各種各様のシステムが存在するが、いわゆる実時間の IR システムは現在のところ実験段階を出ていない。

第4の“情報検索用機械”は、今まで試作された専用機の紹介と、IR システムの最後の段階とで必然的に要求される印刷技術および印刷物の蓄積のための機械類の説明である。この面でのマイクロフィルムの役割はかなり大きい。また機械的処理した情報を印刷物にする行程の機械化（たとえば索引誌の作成）にも注目しなければならない。

第5に一応の“資料”をあげてある。第2次世界大戦前の年間論文数は15年に2倍の速度で増加していたという。戦後はこの増加率が8年に2倍といわれている。これら的情報を処理するために情報センターが拡大されてきているのである。その一端が見られれば幸いである。

最後は、“日本における情報検索の現状と将来”を語る座談会である。情報検索を専門に研究する研究者のほとんど存在しない日本で、今まで実施に苦労してきた苦労話といった方がよいかも知れないが、これらの経験にもとづいたシステムがさらに発展することを願わざるをえない。

本号の編集にあたって、その基礎となったのは昭和39年度電子計算機ソフトウェアの開発研究補助事実の一部として行なわれた“インフォメーション・リトリーバル・システムの研究”的報告書“情報検索と電子計算機利用（技術的問題を中心として）”である。また大駒誠一、昆野誠司の両君に大いに協力してもらった。ここに記して謝意を表したい。