

連載報告

情報学とその課題



1. 情報学の概要とその現状†

尾 関 雅 則¹⁾ 田 畑 孝 一²⁾ 根 岸 正 光³⁾
 藤 原 讓⁴⁾ 大 塚 擁 雄⁵⁾ 開 原 成 允⁶⁾
 杉 田 繁 治⁷⁾ 長 尾 真⁸⁾ 早 川 武 夫⁹⁾
 仲 本 秀 四 朗¹⁰⁾

1. はじめに

人類の発明した言語、文字、紙、活字などに加え、コンピュータと通信は、まさに来るべき、情報化社会の基礎をなす技術である。情報の管理、処理および伝送を取り扱う通信・情報工学が、いわば情報の通路と容れ物を対象とする学問であるのに対し、中身である情報そのものの特性や構造を究明する情報学の進歩に対する期待は、今後、複雑化、高度化の一途を辿るであろう情報化社会の正しい理解と方向づけにとってきわめて大きい。ここに、情報学の確立と発展が望まれる基本的な意義が存在する。

本連載報告では、いまだ十分に確立されるに至っていない情報学の内容と領域を考究し、現時点で情報学の中心的な課題と考えられているものを示し、さらに将来の情報化社会において、大きな貢献があると考えられる情報技術および、それに携わるプロフェッショナルの在り方など、新しい

社会規範の必要性とその中心課題についても触れた。この連載(1)では「情報学の概要とその現状」、連載(2)では「情報化社会の規範」、連載(3)では「わが国における情報学の課題」について述べる。

なお、本報告は第15期日本学術会議情報学研究連絡委員会を取りまとめた報告書「情報学研究連絡委員会報告：—情報学とその課題—、平成6年6月27日、日本学術会議、情報学研究連絡委員会、要約および本文」を編集したものである。報告書の作成にあたった情報学研究連絡委員会委員を文末にあげる。また報告書の作成にあたって協力いただいた委員外の方々も同様に文末にあげ、ここに感謝する次第である。

2. 情報学の概要

情報学とは、「情報の本質に関する理論や知識を体系化すること、さらにその応用として思考活動の質と効率の向上を図ること」つまり、「情報とは何か」、「情報をどう使うか」を追求することであるとする。情報を有効に活用するためには情報や知識がどういうものであるかを知る必要がある。それを基に情報の特性と理論の体系化、情報に関する技術、手法の開発を行い、それを具体的に各分野の情報、思考活動へと応用していく。

情報学でまず問題になるのは、情報、知識、データなどの定義である。情報とは最も広義では「認知とか思考の対象となる実体についての認識内容」であり、普通の意味でいわれる情報はすべてこれに入る。次に知識とは一般的には情報と同じに使われることもあるし、情報処理、特に人工知能の分野では一定の形式化された知識を指すの

† Information Science: What has to be done? (1) Survey of Information Science by Masanori OZEKI (a member of 15th term, Science Council of Japan and also Railway Technical Research Institute), Koichi TABATA (University of Library and Information Science), Masamitsu NEGISHI (National Center for Science Information Systems), Yuzuru FUJIWARA (Institute of Electronics and Information Science, University of Tsukuba), Yasuo OTSUKA (Faculty of Agriculture, Niigata University), Shigekoto KAIHARA (Faculty of Medicine, University of Tokyo), Shigeharu SUGITA (National Museum of Ethnology), Makoto NAGAO (Faculty of Engineering, Kyoto University), Takeo HAYAKAWA (Professor Emeritus, Kobe University) and Hideshiro NAKAMOTO (Integrated Research for Information Science).

1) 第15期日本学術会議会員(第4部)、(財)鉄道総合技術研究所、2) 図書館情報大学、3) 学術情報センター、4) 筑波大学電子情報工学系、5) 新潟大学農学部、6) 東京大学医学部、7) 国立民族学博物館、8) 京都大学工学部、9) 神戸大学名誉教授、10) IRIS 情報学研究所

で、具体的にはプロダクションルールとか1階述語論理で表現されたもの、またはその延長上にあるものということになる。ここでは最広義と最狭義との中間になる、「体系化された情報」という意味で使うことにする。次に情報はいろいろな形で記述され、表現されるがその「最小単位」をデータという。またその「集合」もデータという。以下はこれらの定義に従うことにする。

情報の基本的特性をあげると、a.媒体依存性、b.記述、表現の多様性、c.様相性 (Modality)、d.非加算性、e.階層性 (入れ子構造)、f.相対性、双対などがある。このような情報の特徴とそれに関連する課題についてこれまで多くの研究がなされているが、未解決の問題も山積して、学問としてはまだ緒についたばかりである。

情報学は全体としてどのような専門分野から構成されているかを自然科学の側から整理してみると、理論的、実験的、および応用的な側面があり、次のようになる。

情報学

(1) 原 論

対象領域 (大量情報の特性、資源化、操作に関する理論と基本課題)、背景、歴史、基本概念の定義

(2) 情報解析

特性解析 (属性、媒体、動態)、意味解析 (物理関係、概念関係、論理関係、従属関係など)

(3) 情報構造

グラフ、ハイパグラフ、木構造、網構造、拡張ハイパグラフ、双対、入れ子、部分共有、動的構造など情報空間モデル化

(4) 完全性制約

空値問題、実在制約、識別、同定

(5) 媒 体

物理媒体、論理媒体、表現媒体、記録媒体、表示媒体、通信媒体、および媒体変換

(6) 情報記述、情報表現、情報表示

情報記述 (属性空間、記述項目、差分記述、様相性)、情報表現 (媒体依存性、多様性、多義性、lattice)、情報表示 (多元媒体)

(7) 情報構造化

意味関係構造化、自己組織化、学習、概念構造 (シソーラス、構文解析辞書、定義辞書)、論理構造 (述語論理、様相論理、ファジー論理)、物理

構造 (アドレス、索引、所在、ファイル)

(8) 構造操作

記憶構造、直接アクセス、構造経由アクセス、同型性、準同型性、関係型操作、抽象化 (汎化、集約化)

(9) 意味処理、類似性処理

意味処理 (内容検索、演繹推論)、類似性処理 (共有概念、帰納、類推、仮説推論、連想、発想など)

(10) 応用システムおよび展望

自己組織型情報ベースシステム、人工頭脳など

3. 情報学研究の現状と展望

(1) 言語情報

情報を担っているものは言語だけでなく、図面、写真、映像、その他種々のものがある。それらは表現しようとする情報内容によって使い分けられる。また情報はいかに客観的に相手に伝えられるかという立場からその媒体を考えることもできる。機械の設計図面などは世界中で共通的に理解されうるものである。それでは言語はどのような情報の表現に適しているのだろうか。あるいは情報学における言語の位置付けはどうであろうか。それは次のように考えられる。

(i) 言葉は思想を表現するための最適の媒体である。

(ii) 言葉はだれにでも理解でき、人による理解の相違を最も小さくし、的確な情報を伝達することのできる媒体である。

(iii) 歴史的に見て、人類の知的財産が言語によって最も多く表現され、また蓄積され今日に伝えられている。

(iv) 今日の情報技術においては、言語が最も安価に、最も容易に扱える媒体である。

このような理由から、言語技術は情報学の中で重要な位置を占めていることが分かる。

言語技術は字面処理、文解析、文生成、翻訳、テキスト情報の圧縮と分類などの段階があり、課題としては、テキストデータの蓄積、辞書、解析などのソフトウェアなどの基盤整備、言語理解のための知識辞書の作成から電子図書館の実現までがある。

(2) 情報標準化

情報に関する標準化には、科学技術会議が総理

大臣に答申した「科学技術情報流通に関する基本政策」に従った情報流通技術の標準化 (SIST) と、従来展開されてきた日本工業規格 (JIS) での情報処理技術の標準化の2つの流れがある。いずれもそれぞれの目的に沿って、情報の流通・処理における技術の整合性を高めることに主眼を置いている。両者の活動の経緯には、科学技術振興における支援活動として早くから進められてきた前者と、機器の進展に比して産業化の遅れた情報活動に対する後者との認識の差が認められる。経緯はともあれ、両者とも国際標準への整合を掲げており、その点で共通の基盤を持っているということができる。しいていえば、後者が工業技術の延長上で「情報」を眺めているのに対し、前者は科学技術が内包する「情報」に視点を置いているところに相違がある。また、前者が20年の実績を持っているのに対し、後者は国際標準化機構への国家対応の公式規格であることに、それぞれの特徴を有している。制度的には、前者が科学技術庁科学技術振興局の発行文書であるのに対し、後者は工業標準化法に基づく法律の認知を受けている。内容的には前者が書誌情報および情報生産を対象としているのに対し、後者は用語・略語・記号・符号などが法律にその対象として例示されている。両者の境界については微妙なところがないでもないが、SISTには抄録作成・参照文献記述・レコードフォーマット形式など書誌情報に関するものと、学術雑誌構成・学術論文構成・科学技術レポート様式・予稿集様式などの情報生産に関するものの2種となる。ISO/TC 46 (情報ドキュメンテーション) の制定した国際規格の中で、JISとして国名コード、ISBN、ISSN、ドキュメンテーション用語などが制定された。

(3) 全文データベース

いわゆるデータベース・サービスは、文献の書誌的データと要旨を収録した二次文献データベースを中心に発達してきた。これには、従来の抄録誌の編集作業が電算化され、そこで得られる電子化ファイルが、当初は副産物的位置付けにあったが、次第に主製品としてのデータベースに転化してきたという背景がある。一方、数値情報系のデータベースは、株価の即時配信システムという、データベース的ではないオンラインシステムから始まったが、その後、データの蓄積機能を取り

入れてデータベースを構築するようになった。また統計情報関係では、統計データのデータベース的整備と、これに対する分析プログラムの組合せでサービスするシステムが早期に事業化されている。全文データベースは、上述の二次文献データベースの進化の延長上にあるものである。

全文データベースの発達は、画像・音声マルチメディア型データに関する情報技術のみならず文字検索にしてもシグナチャーファイルなど全文対応の技術を必要としている。もっとも、高速・大容量・低価格のインターネットの普及にともなって、オンライン系でも、画像・音声を含むハイパertextの普及の兆しを見せている。また、既存出版物の各頁を画像として蓄積・配信することも、インターネットの普及によって現実的になっており、このような全文データベースの集積と配信を総合したシステムを「電子図書館」と称して、その実用化にむけた開発計画が米国を始めとしていくつも試みられている。

(4) マルチメディア

我々の日常はマルチメディアの世界である。文字、画像、音響など様々な情報メディアに囲まれて生活している。ところが1980年代に入ってマルチメディアという言葉が新聞や雑誌広告などにしばしば見られるようになった。これはコンピュータや通信の世界において単に文字や数値だけではなく、静止画のみならず、動画や音響も扱えるようになりつつあることを反映して特に強調されて出てきた表現である。

画像や音響データはそれをデジタル化すると膨大なデータ量になる。1枚の写真も文庫本10冊分くらいの量になる。それを数千枚数万枚を扱うとなると大変な量である。従来のコンピュータではシステムが大きくなりすぎた。ところが大量のデータを蓄積することができる媒体、光ディスクの出現によって事情が一変した。数百メガバイトのデータを蓄積することができる光磁気ディスクやCD-ROMをコンピュータに接続することによって従来の計算機のメディアの世界が飛躍的に広がって、我々の日常的なメディアの世界に近づいてきたのである。

入力用スキャナ、データ圧縮技術、光通信、衛星通信などの技術の進歩に支えられて応用が広がりつつあるが、データ間の関係付け、言語による

記述法などの課題も多い。

(5) 情報自己組織化

高度な思考機能に対応する情報処理には情報の意味処理が必要である。ところが意味の関わる問題の多くは未解決である。たとえばデータベースや知識ベースでは個別実体 (Distinct Entities) の集合として、外延的にデータや知識を対象としている。つまり考えている対象領域では、ある概念の表現と別の概念の表現との間に重なりがなく、別々のものであるというのが基本的な考えである。実際に使われる情報では特許や法律、あるいは化合物でも、概念には非常に多くの重なりがあり、それを考慮しないで処理することは無理である。また類似性というのも、もともと重なりがある概念の関係であるから同様である。あるべき情報が欠落している空値の取扱いではさらに意味処理が困難である。これらが情報の管理や、知識の獲得の困難さの原因となっている。

表現の多様性も情報の記述、目的、内容に応じて大きく変化する。分類や表現の多様性は情報の表現の本質的な性質であるので意味の処理が困難になるのである。このような問題の解決策として脳における機能と同じように情報の意味的關係を自動的に構造化する自己組織化の研究が注目されている。

脳における学習に対応して、概念や情報間の意味的關係を抽出して情報の組織化を行う。概念に内在する関係は概念構造としてシソーラスを構築する。論理関係は、原因-結果、理由-結果、要因-結果を主としてタクソノミーの形で組織化する。原情報の所在、書誌情報、アクセス情報などは物理構造として構築する。メディア変換、意味関係抽出などによりマルチメディア型原情報の概念構造、論理構造、物理構造などの自己組織化を行い演繹推論、帰納推論、類推などの可能な人工頭脳型システムを設計する。研究に必要な情報の動的構造の記述操作のためには新しい型の情報構造を持つモデル SSR (Structured Semantic Relationship) の開発がなされている。そのためデータベース、知識ベースおよびハイパメディアなどの要素技術を大幅に拡張し、新しいモデルで統合的に記述表現し、操作し、思考支援できる機能を持つ情報ベースシステムの設計と、基礎となる理論の研究が行われている。

(6) 国際全文データベースサービス

学術文献情報に関するデータベースサービスは長くいわゆる二次情報を提供するサービスであった。利用者が指定したキーワードを基にデータベースを検索し、論文名やその書誌的事項、あるいはその抄録を提供するというものである。データベースの構築者は、学術文献のためまぬ収集と蓄積に努め、それに基づく利用者へのサービス提供をビジネスとして維持してきている。

最近になって、学術文献の一次情報、すなわち論文の本文そのものを提供する全文データベースシステム構築の国際的な実験プロジェクトが欧米を中心に開始されている。このシステムの特徴は次のようである。(1)論文・記事の本文を提供、(2)国際学術ネットワークを介し、多くのサイトが自由に参入して相互にデータベースを提供、(3)全文検索、ハイパテキストなどの新検索手法の導入、(4)記述の国際標準化、(5)文字情報のみならず画像、音響情報を提供、(6)国際学術ネットワークを介し、どの国のだれでも利用可能。

これらのプロジェクトの例として、WAISとWWWの2つがある。これらはいずれもインターネットの上に構築されていて、インターネットに接続したワークステーションからだれでも利用できる。

4. 専門領域における情報学の現状と展望

専門領域として、ここでは医療情報、判例データベース、農業情報、博物館情報を取り上げる。

日本の医療情報の分野は、実用面では一部で高い水準の保健医療情報システムが稼働しているが、全体的なポリシーが欠けているために、システムが相互に連携しない形で作られ、多くの無駄や非効率な点が見られる。また、データに基づき客観的な意思決定をする習慣がないため、情報システムが医療機関の運用効率の追及に使われ、情報を収集・分析して意思決定に使うという認識が少ない。

一方、研究面では、日本で生まれたイノベティブな医療情報関連技術は少ない。医療関連の新しい技術は、ソフトウェアの形で生まれることが多いが、日本ではソフトウェアの価値に対する認識が少ないことも上記の理由の1つであろう。

今後重要な政策的課題として、ここでは次の7

つをあげる。(1)医療情報政策を討議する場の創設,(2)ソフトウェアの価値の認識,(3)医療情報関連技術の標準化の促進,(4)情報を利用することに対する経済基盤の確立,(5)情報技術応用の障害となる法令の見直し,(6)一般人への医療情報の提供の原則の確立,(7)情報技術を用いた国際医療協力。

判例検索は乾草の山の中から1本の針を探すのにも似た大変な作業である。判例法を主要法源とする米国においては、年々連邦や諸州その他の法域の最高裁以下の諸裁判所から流出されるおびただしい数の判例がデータベース化されている。判例データベースは、当初は判例への手がかりを集大成したものであったが、現在では判例の全文が入力されている。検索方法は、全文を対象としたKWIC(Key Words In Combination)方式が採用されている。キーワードは全文の中の前置詞、接続詞、冠詞、人称代名詞などを除いたすべての語である。

農業情報について考察する。農業研究の分野は、生物の改良・機能の解明、生産の場としての物理的・生物的環境の解明・制御、環境に人為的に投入される資材の開発・改良・管理、農業基盤整備、機械化、施設化・農作業の効率化、農業・農村計画、農家経営の改善、農業生産動向の解析・予測、食品の安全性・高品質化など多様であり、研究方法も自然科学から人文科学まで広範囲でしかも扱われる情報も多種・多様である。さらに、最近では農村の活性化、環境保全、国土保全機能など農業の持つ多面的機能の向上が重要な研究課題となっている。

農業情報の特徴と関連して農業分野における情報研究の課題をいくつかあげる。

(a)巨大情報の管理・処理手法、(b)システム科学的手法、(c)データベースの開発、(d)農業・農村の高度技術化、高度情報化。

現在、農業における情報研究は情報の計測・収集手法、管理・蓄積手法、情報システム化手法など広範囲に実施されているが、必ずしもうまく進んでいるとは言えない。それは情報学専門の研究者がいなかったこと、組織的な研究体制がなかったことおよび情報研究の成果を正当に評価する認識に欠けていたことによるものと考えられる。

博物館情報について考察する。普通博物館とい

う名称は考古資料や生活資料を陳列している場所として狭い意味に使われている、しかし美術館、水族館、動物園、植物園、科学館、なども博物館の範疇である。そこでは対象の違いはあっても、様々なモノを集め、それをある種の体系にしたがって展示をし、そのモノの世界をよりよく理解させるようにする施設である。

普通どこの博物館でもその館が対象にしているモノに対して情報カード、あるいは資料カードが用意されている。そこにはいろいろな項目がかかっている。名称などの基本情報はカードに記入されているが、その他の情報となるとほとんど記入されていないのが現状である。ましてやコンピュータ化がなされている場合は少ない。情報カードの作成ができていない原因の1つは、どのように分類するかで行き詰まってしまうからである。分類体系の確立は困難であるから、コンピュータを利用した別の検索体系を模索する必要がある。たとえば、情報カードに含まれている文字列すべてが検索の対象にできるから、断片的なデータであっても活用可能なのである。

大小様々な博物館で情報化が進んでいない要因として標準化ということにこだわっているところがある。図書の分類体系と同様の標準的な項目をすべての博物館で共通に利用できるのを待っているというケースがある。標準化はそれができに越したことはないが、なかなか足並みが揃うものではない。それを待っていたのでは資料がどんどんたまる一方である。各館独自のデータフォーマットでも検索できる体制にあれば、他の博物館とのネットワークによって相互利用ができるのである。記述データだけではなく博物館では立体的な画像を含めた画像データが重要になる。

5. おわりに

情報学の現状と展望について本報告で取り上げた専門領域は4つであるが、これは単に例示で、本研究連絡委員会の委員構成を反映したにすぎない。ありとあらゆる専門分野で同様の、あるいはもっと進んだ考察がなされているに違いない。ある専門分野で有効な手法が他の専門分野にとってそのまま有効であったり、あるいは参考になることが期待できるので、専門分野を越えて互いに意志疎通をはかることが肝要である。

付記 第15期日本学術会議情報学研究連絡委員会：尾関雅則（委員長，第15期日本学術会議第4部会員），田畑孝一（幹事），根岸正光（幹事），藤原 讓（幹事），大塚擁雄，開原成允，杉田繁治，長尾 真，早川武夫，米田幸夫。

本報告作成に協力した方々：神田利彦，黒川恒雄，仲本秀四朗，名和小太郎，横井俊夫。

（平成7年2月16日受付）



尾関 雅則（名誉会員）

1946年東京帝国大学第1工学部電気工学科卒業。同年運輸技官。1972年国鉄電気局長。1975年常務理事。1979年日立製作所入社。1982年OA事業部長。1983年常務取締役。1987年より(財)鉄道総合技術研究所理事長。現在に至る。国鉄時代，マイクロ波通信の実用化，電話の自動交換網の推進・普及，貨物操作場の自動化システムの指導などに従事。また，みどりの窓口などの旅客総合予約システムを開発。この功績により紫綬褒章と電子通信学会業績賞を受賞。現在は主に超電導磁気浮上式鉄道の開発プロジェクトの指導・推進に従事。運輸技術審議会委員，情報処理国際連合(IFIP)理事など。

