JICST ファクトデータベース (1)

その概要とオンラインサービス

黒澤 慎治*

JICST Factual Database (1)

The outline and online services

KUROSAWA Shinji

[著者抄録] JICST が昭和 63 年 1 月よりサービスを開始した、JICST ファクトデータベース(JOIS-F)について、その全体の概要とオンラインサービスを概説した。初めに、昭和 53 年の構築開始より現在までの経緯を述べ、「科学技術振興調整費プロジェクト」との関わりについても触れた。次に、昭和 63 年度から 64 年度にかけてサービスを開始する予定で現在システムを開発中の、DNA、金属材料強度、結晶構造、化学物質法規などのデータベースについて述べ、最後にオンラインによるサービスに簡単に触れた。

[著者付与キーワード] ファクトデータベース, JOIS-F オンラインネットワークシステム, 質量スペクトル, 熱物性, 化合物辞書, DNA, 金属材料強度, 結晶構造, 化学物質法規

[Author Abstract] The outline of JICST factual database (JOIS-F), which JICST has started from January, 1988, and its online service are described in this paper. First, the author mentions the circumstances from 1973, when its planning was started, to the present, and its relation to "Project by Special Coordination Funds for Promoting Science and Technology". Secondly, databases, which are now under development aiming to start its services from fiscal 1988 or fiscal 1989, of DNA, metallic material intensity, crystal structure, chemical substance regulations, and so forth, are described. Lastly, its online service is briefly explained.

[Keywords] factual database, JOIS-F Online Network System, mass spectrum, thermophysical property, chemical compound dictionary, DNA, metallic material intensity, crystal structure, chemical substance regulations

* JICST 技術管理室 (〒100 千代田区永田町 2-5-2) Tel. 03(581)6411
JICST, Technical Coordination and Development Office
(5-2, Nagatacho 2-chome, Chiyoda-ku, Tokyo 100)

1. ファクトデータベースの構築の経緯

この1月からJOIS-FとしてJICSTファクトデータベースのサービスを開始することとなった。JICSTが初めてファクトデータベースの構築を手がけたのは昭和53年からであり、本格的サービスに至るまでに9年の歳月を要している。

このファクトデータベースの構築はわが国の科学技術情報政策として出された「科学技術情報の全国的流通システム」、通称NIST構想の一環として具体化されたものである。

53年5月に科学技術庁に設置された「科学技術情報活動推進懇談会」において、「科学技術情報活動推進の目標と施策について」が提言され、この中でデータベースの拡充の一環として数値データや画像データなどのデータベースの作成の重要性を指摘している。

一方、物性データの整備についての調査検討を目的とした物性データ活動検討会が科学技術庁内に設置され、物性データの集積・流通のための機能の整備についての報告書が53年7月に出された。この当時、ファクトデータベースはまだわが国では少なく、しかも作成されたデータベースもその流通はインハウスに限られ、一般への流通は相変わらず冊子体を中心であった。

59年8月には、科学技術会議政策委員会科学技術情報小委員会において、「科学技術情報の流通に関して当面する課題への対応について」の報告書が提出された。この中でファクトデータベースの拡充の重要性を指摘し、データの生産・蓄積、データの評価、相互利用について勧告した。

また、59年11月に出された科学技術会議の第11号答申である「新たな情報変化に対応し、長期的展望に立った科学技術振興の総合的基本方策について」では、特に遅れているファクトデータを中心に科学技術情報の生産および流通の促進を図ることを提言している。

かかることは、わが国の研究開発を促進せしめる上でも、また海外からの科学技術情報に依存する体制からの脱却を図る上でも、わが国独自の科

学技術に関するファクトデータベースを数多く構築し、流通せしめる必要があることを物語っている。

かかる状況のもとに、55年度からJICSTにおいても本格的なファクトデータベースの構築および流通を図ることとなった。そこで、化学物質の同定などに利用されることが多い「質量スペクトル」と物質の特性の研究のために必要な「熱物性」についてデータベースを構築することとなった。

JICSTにおいては、文献二次情報の経験は豊富ではあるが、ファクトデータベースは初めてのことであり、しかも専門性が非常に高いため、その分野の専門家の協力を必要としていた。そこで、質量スペクトルについては日本質量分析学会の協力のもとに、また熱物性については勸産開発科学研究所と(株)化学工学協会の協力のもとにデータベースの構築に取りかかった。

当時、科学技術庁では科学技術振興調整費（振調費）によるネットワークとデータベースに関する共同研究を計画していた。すなわち、「ネットワーク共用による化合物情報等の利用高度化に関する研究」であり、56年度から61年度にかけて十数機関の協力のもとに研究が行われた。JICSTもこの共同研究に参加することとなり、55年度からデータの蓄積を開始している熱物性については検索システムの開発を、化合物については辞書の作成と化学構造の検索システムの開発を担当することとなった。

質量スペクトルについては、53年度から55年度にかけて英国のMSDCのデータをもとに検索の実験システムを開発した。その後、この実験結果をもとに機能の改良や充実を図り、62年3月より米国のNBSのデータとともにTSSによる試験サービスを開始した。

現在、質量スペクトルと熱物性の他に、60年度より金属材料強度と結晶構造、62年度から化学物質法規についてデータの蓄積を行っている。また、検索システムの開発としては、金属材料強度と結晶構造の他に、DNAについても60年度から検索システムの開発に取りかかっている。

2. 科学技術振興調整費プロジェクトとしての技術開発

昭和 56 年度より技術開発を行っている「ネットワーク共用による化合物情報等の利用高度化に関する研究」は、オンラインネットワークシステムの形成に関する研究と化合物総合データベースシステムの形成に関する研究に分かれている。

前者のオンラインネットワークシステムの形成に関する研究は、農林水産技術会議、工業技術院 RIPS センターなど、7 機関のメーカの異なる計算機を接続し、それぞれの機関が所有しているデータベースを自由に利用できることを目的としたネットワークシステムである。

例えば、RIPS センターの研究者がアスコルビン酸の分析を行い、RIPS センターにある赤外スペクトルのデータベースを検索して、アスコルビ

ン酸であることを確認する。次いで、この物質の化学構造を知るために、JICST の日本語化合物辞書データベースにアクセスし、この物質の構造を検索する。さらには、この食品添加物としての特性を求めるために、農林水産技術会議の計算センターのバイオケミカルデータベースを検索して、アスコルビン酸の添加物としての特性を知る、というようなわけである。

この研究では、共通のプロトコルと通信ソフトウェアの共通設計により異機種 of 計算機結合に成功したことに共に、各機関のデータベースを相互に利用できたことが大きな成果である。このために、標準的サービス仕様として、情報の表現形式、検索言語等について統一を図った。特に、いくつものデータベースを検索するための「わたり検索」の機能が大きな役割を演じた。この機能は JOIS-F にも取り込んでおり、質量スペクトルから熱物

図 1 振調費によるオンラインネットワークシステム構成の概念図

性に、さらには化合物辞書にとデータベース間を検索していくことができる。

化合物総合データベースシステムの形成に関する研究では、工業技術院化学技術研究所、環境庁公害研究所など10機関が化合物に関するデータベースを構築し、検索システムを開発して、その相互利用を図ることを目的としている。特に、スペクトル、バイオロジカル、安全性、化合物環境、熱物性、バイオケミカル、農薬、医薬品、急性中毒など10種類のデータベースを日本語化合物辞書番号により関係付けて、データベースの相互利用を可能にしている。

この日本語化合物辞書システムは、計算機処理により化合物の日本語の体系名から立体構造を識別できる原子結合表を自動作成するシステムであ

り、さらに慣用名、分子式、分子量などを加え、各データベースにある化合物を登録し、また辞書システムにある情報をデータベースに利用できるようにしている。そのため、各データベースは対象とする物質、データ項目、表現形式を統一している。

JOIS-Fでは、この日本語化合物辞書システムを提供しているが、辞書自体はJICSTにおいて作成されたものである。また、当面単独の低分子有機化合物が対象であるが、さらにマーカッシュ型構造や混合物など当初開発の仕様から外した化合物についても現在開発を進めている。

熱物性については、データ自体は55年度から前記学協会の協力により蓄積を行っているが、検索システムは振調費により開発している。このシス

表1 振調費により構築したデータベースの概要

データベース名称	データの概要	実施機関
①化合物辞書	各化合物データベースが取り扱う物質を登録し、立体構造まで識別可能な日本語化合物辞書	日本科学技術情報センター 化学情報協会
②スペクトル	純物質を対象とした赤外、 ^1H -NMR、 ^{13}C -NMR等のスペクトルデータ	工業技術院化学技術研究所
③バイオロジカル	元素、無機化合物、有機化合物、天然物等の広範な化学物質を対象とした発癌性、変異原性および催奇形性についての試験データ	国立衛生試験所
④安全性	化学物質の安全および衛生に関する国内の法令、規制に含まれる化学物質を対象とした情報	日本化学物質安全・情報センター
⑤化合物環境	環境中に存在する有害な化学物質を対象とした物性、化学構造、環境中での測定値、分析法等の情報	国立公害研究所
⑥熱物性	元素、無機化合物、低分子有機化合物の一、二、三成分系の化学物質を対象とした熱力学、熱化学、相平衡、燃焼、爆発等の物性データ	日本科学技術情報センター
⑦バイオケミカル	肥料、飼料、食品添加物等の物質を対象とした物理化学的性質、生化学的性質、生産量規制等の情報	食品総合研究所 農業環境技術研究所
⑧農薬	我が国で登録されている農薬の名称、有効成分、使用方法、分類、統計等に関する情報	農薬工業会
⑨医薬品	国内に流通している医薬品の商品名、有効成分名、含量、製造・発売者名、薬剤の形態・コード、薬効分類番号等の情報	日本医薬情報センター
⑩急性中毒	急性中毒を引き起す可能性のある商品を対象とした有効成分、症状、処置例等の情報	同上

テムの大きな特徴は多面的に検索できるように項目数を多くしていることと、数値検索である。特に、必要とするデータがその範囲にない場合、自動的に範囲を拡大できるようにしている。

熱物性については、利用面の仕様変更を別にすれば、この検索システムは JOIS-F に取り込んでいる。

3. システム開発中のファクトデータベース

化合物辞書については、マーカッシュ型構造や混合物などの化合物について、その入力システムおよび検索システムを開発しており、これにより対象とする化合物は広くなり、より一層利用しやすくなるであろう。

60 年度から遺伝子情報に関するデータとして DNA を扱うこととなり、その検索システムを開発している。対象とするデータベースは EMBL と GenBank である。この検索システムの特徴は塩基配列を調べるホモロジー検索であり、研究者が調べたデータと該当する DNA との照合が重要なテーマとなっている。

同じく 60 年度から科学技術庁の金属材料技術研究所および無機質材料研究所とのデータベースの構築・流通に関する共同研究を行うこととなった。その一つは金属材料強度データベースであり、もう一つは結晶構造である。

金属材料技術研究所との共同研究である金属材料強度データベースは鉄鋼などの構造材料に関するクリープおよび疲労についてのデータであり、同研究所が永年にわたって試験したデータを集めたものである。この検索システムの特徴は数多くのデータを解析するシステム、それをグラフ化するシステムにある。

無機質材料研究所との共同研究は結晶構造の入力およびその検索システムの開発であり、データは各種文献から採択している。検索システムの特徴は結晶構造の種々の特徴からの検索と三次元構造の投影図の表示にある。

振調費における共同研究機関である勸業化学

物質安全・情報センターの協力により、化学物質法規データベースを JOIS-F に搭載するために 62 年度より蓄積を開始している。

このデータベースは化学物質の審査および製造に関する法律（化審法）とか、労働安全衛生法といった化学物質の安全および衛生に関する国内の法令や規制を集めたものである。これまでの科学データを中心としたデータベースとはやや色彩が異なり、一種の法令情報に当たるデータベースである。

4. オンラインデータベースサービス

この 1 月からサービスを開始した JOIS-F は質量スペクトル、熱物性および化合物辞書の 3 種のデータベースを搭載している。

質量スペクトルについては、62 年 3 月から試験的にサービスしているシステムを、オンラインリアルタイムに改造しただけであり、基本的な検索部分には変更はない。今回、300 bps の他に 1200 bps の通信速度が利用できるのも、より便利になっている。

熱物性は、61 年度末まで振調費によって行われた検索システムについては変更はないが、課金関係とか障害対策といった商用サービスに必要なプログラムについては新たに追加している。化合物辞書についても熱物性と同様に課金関係や障害対策などについて新たに追加している。

現在開発中の DNA、金属材料強度、結晶構造、さらには化学物質法規などのデータベースは 63 年度から 64 年度にかけてサービスしていく予定である。これにより、物性系として熱物性や質量スペクトル、結晶構造、材料系では金属材料強度、ライフサイエンス系では DNA、その他として化学物質法規と日本語化合物辞書と、JOIS-F は多彩なファクトデータベースを持つこととなる。

また利用面でも、グラフィック系では化合物と結晶構造、数値系では質量スペクトル、熱物性および金属材料、文字・記号情報系では化学物質法規と DNA などがあり、ファクトデータベースの

処理に必要な技術はほぼ出揃っている。

現在の JOIS-II では、システム的にかかるファクトデータベースを扱うことは困難なため、JOIS-F としてシステムも計算機も別の体系のもとにサービスすることとなった。

ファクトデータベースの構築では、専門性とその労力や費用については問題は大きいですが、技術的にはそれほど問題はない。しかるに、検索システムはデータベースそれぞれに個性があることから、データベースに合わせて検索システムを開発

する必要がある。

たとえ材料系であっても、その利用の仕方により検索システムは非常に異なり、データだけを利用する、図表化が必要、計算式を組み込む、自己のデータとの関連を求めるためにアップローディングが必要など、利用形態により検索システムは複雑かつ高度化する。しかも、検索システムの開発は一度で満足することは希であり、しばしば改良を必要とするので、システムの改善が今後の重要な課題である。