数理解析研究所講究録
第433巻 1981年 20-31

20

数学分野の情報検索について

京都大学数理解析研究所　一松　信

§1. 情報検索とは？

本稿は、研究会での報告ではなく、解説のために新規に書き下ろしたものである。

近年学術情報検索という言葉がよく使われる。これは一口にいえば、計算機の記憶装置中に多量の学術上有用な情報を蓄積し、その中から自分の知りたい知識を得る手段である。

どんな分野でも、学術研究に対して、情報は不可欠である。ここで情報とは、広義に解して、研究課題、そのためのデータ、これまでに既知の事実などから、その問題の研究者といった対象まで含まれる。

そういう情報を各人が自分用に集めて整理するのは、いやしくも研究者たるものにとって、当然の前提である。そしてこれまで、特に数学のような比較的「小さい」集団においては、その種の個人や小グループの努力で間にあっていた。また日本では各大学の数学教室が文献の蒐集に熱心であり、優秀な司書に支えられて、何とかすんできたようである。

計算機による情報検索は，もちろん計算機並びにその関連技術の発展に負うものであるが，論文数が多くて人力で処理が困難になった化学（年間論文数約40万）や，人命にかかわる場合があって費用をかけることが許される医学などの分野から，まず実用化された．それとともに天文学，地球物理学などでは膨大な観測データの整理・保管が必要であるし，病院の空きベッドについては，座席予約のようなシステムが考えられた．いずれも人力によるのに比べて，速度・正確さ・柔軟性を増すことを主眼としている．

そのような「先端的」諸分野に比べると，数学分野は「遅れて」いる．しかしそれには多くの理由がある．本冊の最初の二論文に論ぜられている通り，日本でもかなり以前から小規模な研究が進められており，経験も蓄積されている．問題は技術よりその必要性を許えることと，限りある人才をどの才面に投入するかといった多分に「政策的」な点にある．

計算機の進展はものすごく早い．十年前の知識を金科玉条と心得えていれば，笑いものになるだけである．したがって現在困難ないし不可能と思われる課題に対しても，何が障害であるのか，それが除かれる可能性があるのか，といった対応さしておく必要がある．けっして安易な夢にうかれてはいけないが，自己の現状に満足するのでなく，将来のことをまく

考えておかないと，後世の研究者から批難されることになるであろう．

§2. 数学における情報検索の問題点

数学分野における二次情報（論文リスト・抄録など）誌は意外に歴史が古い．Jahrbuch über die Fortschritte der Mathematik（現在廃刊），Zentralblatt für Mathematik und ihre Grenzgebiete，Mathematical Reviews は，それぞれ1869，1931，1940年に創刊されている．これらは周知のように，非常によく整備されているが，刊行までに時間的ずれが生ずるのが難点である．最初の雑誌が廃刊になったのも，時間的遅れが大きくなりすぎて，役に立ちにくくなったからである．

もともと数学では，「だいたいよさそうだ」は意味が薄い．情報検索に対しても，他分野とくらべて，非常に適確な質の高い情報でないと満足しない傾向がある．この「完全主義」が障害になる場合が多い．どこまで妥協するかが課題だろう．

数学においては，正しい定理の寿命は永久である．もっと簡単な証明ができたり，拡張されたりして，不必要になることはあっても，「実験事実」そのものが新しい論文がでるために変ることはない．その意味では正しい定理集があれば，原理的には十分なはずである．—— もちろんこれは夢にすぎな

い，「数学辞典」がけっしてその種の目的のハンドブックとして十分でないことからも想像される．量的な制約だけでなく，内容を智的に判断し，適確な情報を検索することは，人工知能の一環として研究に値する課題であるが，早急な実用化は望めそうもない．さらに数学の定理は，単に羅列するだけでは無意味であって，体系化することが本質的である．しかもその方法は，必ずしも一通りでない．そして，本をめくっているうちに，とんでもない(?)標題の下にある定理が，自分の当面の課題に密接に関連しているのに気づく，といった体験をおもちの研究者が，少くないと思う．

というわけで，定理検索は，当分の間は安易に計算機に期待するよりも，ハンドブック，教科書，総合報告類の整備と，物知りの学者の養成というほうが先決のように思われる．

これに対する二三の「傍証」を示そう．かってある定理が既知か否かを土がすため，人海戦術で Math. Reviews を十数年分くってみて，ついに発見した例がある．しかし Math. Reviews のデータベースができていたとして，この作業を計算機で実行して成功したか否か，疑問である．さがした人々が高度の専門家達であり，その専門知識が十二分に活用されたからである．

また数学関係の論文に引用される文献で，近年最も多く現

れる著者が、Erdélyi（公式集の編集者）とBourbakiであるという統計がある。生物科学などで多用される引用回数類の統計をとると、数学の場合上位を示めるのは、大半論文ではなくて、古典的な教科書や公式集になるといわれている。

これは他方では、数学の論文の寿命が永く、過去への遡及検索が不可欠であることをも物語っている。

もっとも公式の検索は、すでに数式処理体系のあるもので実用化されている。数学を道具として利用したい多勢の人々の需要を考えると、一つの重要な課題である。しかしそれには数式の表現法とその標準化といった難題がある。

数学分野ではthesaurusの類が余り有用でない。それは国により時代によって術語の変化が多いほかに、数学者各自の発想法にも多きな差があり、一つの事項に辿りつくまでの道が多様なことに起因するらしい。さしあたっては、個々の事項よりも、それの載っていそうな文献とその所在検索を中心にせざるをえない。しかしこれさえも意外と難しい。

その大きな原因は、数学のチュウショウ性である（抽象と中小にかけたシャレ！）。現在全世界で刊行されている学術雑誌が3万種ほどであり、Math. Reviewsに引用されているのが1200誌ほどである（比率4%）— ただし周辺分野を除けば800誌程度。ところが発表される論文数のほうは

400万中の4万程度であって、1%そこそこである。じっさい数学の専門誌の多くは季刊前後である。月刊は「大雑誌」であり、週刊誌は皆無に近い。また一つ一つに掲載される論文数が少く、一つ一つの論文が長い傾向がある。これらの統計は、いずれも他分野と著るしい差異を示している。

数学のそれぞれの分野には、専門誌があるが、それらが目につくようになったのは1960年以降である。いまでも「総合雑誌」の比重が高い。そして小国や小大学の紀要の類に、稀ではあるが極めて質の高い論文が載ることがあり、小雑誌を無視することができない状況である。

こういう状態は、big scienceの専門家から「数学の後進性」と批難されてきたが、私は必ずしもそうは思わない。個人のアイディアが中心である数学の研究には、大規模集中化はかえって有害であり、小過りのきく小グループ多数という形態が、むしろ研究上にも望ましい姿なのではないかという気さえする。しかしそのことは、逆に機械検索の面からは、大きな障害になる。

さしあたっては、著者の側にも、適切な題をつけ、適切な雑誌に発表し、必要ならkey wordを付するよう努力してほしい、というしかない。

しかし一方、数学のチュウショウ性のおかげで、単著の論

文が多く，著者名自体が重要な key word であること，多くの場合，自分の専門に近い core journal や重要な参考書は個人ファイルが作れる程度であることなど，やる気になれば利点となる面も少くない．不完全な検索体系から，自分の必要な情報を立がし出すのは，ナゾナゾや頓智問答に似た智的パズルとしての面白味がある，というマニアもないではない．（実用上それでは困るけれども！）．

情報検索のような実用を主目的とする課題にあっては，多少とも利用者側に立，「よいものができたら使おう」ではなく，建設的批判をくりかえしつつ育ててゆこうという気持が必要と思われる．そういう雰囲気をもり上げることが，特に日本において第一の課題かもしれない．

§3．現在何ができるか？（利用上の注意）

情報検索のためには，まず検索されるべき情報を蒐集して計算機に入力しなければならない．そのような集積された情報を（いろいろの名があるが）データベース（Data Base）という．次に検索のためのプログラム（Data Base Management System，略して DBMS；いろいろの名を冠する既製品がある）がいる．しかしこれらはデータベース作成者側の作業であり，利用者が直接関与することではない．

数学分野で最大の悩みは，MOSがつぶれたあと，世界的なデータベースがなく，外国からよいデータベースを買って来て日本の計算機に載せればすむ，というわけにゆかないことである．もっともZentralblattやMath Reviewsのテープ化は近く実現しそうであり（サンプルはある），Isi社のデータベースComp-mathも近くサービスを始める予定である．日本でも本冊に見られるように，「手作り」のデータベースが少しずつではあるが整備されてきている．それらの多くは大型計算機センターで公開されており，所定の手続きをして所定の料金を支払えば利用できる．

情報検索は，やはり端末機の前に座り，計算機ときめ細かい「会話」を進めながら使うのでなければ，よい結果は得難い（本冊の後の論文に実例がある）．しかしそのために端末機の使用法を習得することは必ずしも必要ないかもしれない．（これからの研究者の必須の技術の一つという気がするが）当分は優秀な図書室司書にたのみ，端末の操作員と協同してあれこれと工夫してみることで済むかもしれない．そのような人手のない所では，少しの努力で各自操作法と検索のコツを学ぶべきである．

手作りのデータベースでは，説明書が不備で，どれだけのことがどうすればわかるか，理解しにくい場合が多い．しか

しこの種の実用技法は，理論があってその応用，という発想よりも，「手を動かしつつ身体で覚える」ほうが効果的である．すなわち最低の操作法だけを教わり，あとは智的パズル（とあらさがし）精神で，実際に計算機と対話しながら使用法を覚えるほうがよいらしい．――ただしそのためには，十分の暇と，心の余裕と，多少の「授業料」が必要であって，これらが全部十分に恵れている環境が少いようなのが問題らしいが――

前述のような状況であることを踏えて，少々見当違いな情報や誤った情報が現れたときには，どうしてそうなったかという吟味や建設的批判をお願いしたい．またこういうことはできないか，という希望も，すぐには実現困難かもしれないが積極的に寄せてほしい．そうした利用者の協力が，よいシステムを育ててゆくための不可欠の条件なのである．

個々のシステムによって差があるが，現在のたいていの検索プログラムは，著者名，分野名（或はコード），標題やその他の重要語，その組合せ，といった項目で検索できる．国際会議録類は，場所，日時，通称といった不完全な断片的情報から現物に辿りつけるように相当に工夫をしてある．数学分野の利用者に特に注意したいのは，イニシアルが同一の別人がしばしば同一人物に化けたり，逆にローマ字の綴り方で日本人，中国人，ロシア人などの場合，一人が何人にも化け

ている場合が多いことである。これに対処するには最終的には完全な「人名辞典」を作るしかないらしい。—— 検索なら笑い話ですむが、注文した本がイニシアルの同一な別人の所に届いて、請求書だけが本人の所に来た、となると大問題である。

術語についても、field, lattice, regular, distribution といった、分野によって全く別の意味に使われる用語は、要注意である。少くとも今後新しい概念に名をつける場合には、それなりの工夫を望みたい。略字も P.D.E. はかなり定着して後方一致の key word として使えるが、Q.C., L.P., I.R. などは危険である。

数式は、数学においては重要な key word であるが、現在の計算機検索では余りあてにならない。この面でも数式の表現の標準化が望まれる。

Key word の時代による変遷は、物理学において問題になりかけているが、寿命の長い数学の論文については、特に重要であろう。その意味では、「数学辞典」の類では、機械的な術語の統一をはかるよりも、慣用の用例を広く集めて分類整理する「辞典」的な編集方針を望みたいものである。

生物科学や化学方面でよく使われる相互引用の回数などでは、数学の文献は孤立化、ないしグループ独立の形態を示し、評価の適切な規準になり難い。Isi では引用文献の関

連から，clusterの自動設定を試みている．現在までの試験例では，はなはだしく大小不揃いで漫然とclusterができている印象である．しかしこれは伝統的な数学の分類が先入観になっているせいなのか，それともともともと物質や生物の名といった具体名に有効であった手法を，抽象概念が中心の数学に機械的に適用したせいなのか，(判定には)もう少し実験を重ねてみる必要がある．数学(論文)の分類は，reviewerにとっても大変な苦労であり，MOS codeが必ずしもぴったりでない例が多いだけに，機械的に有用なcluster設定ができれば，検索にも有用であろう．

最後のほうは，利用者に対するよりも，データベースや検索体系の作成者・研究者に対する注文や希望であるが，建設的批判を賜るために一言した．

§4. むすび

以上はなはだ表面的な記述であるが，数学分野の情報検索とその現状について，若干解説した．

1979年より，情報センター設立のための「特定研究」に数学分野の代表として参加し，他分野との関連を知る機会を得た．筆者について特に驚くべきだったのは，文献及び情報検索についていえば，数学は実験系自然科学とはまったく異

質であり，むしろ社会科学の諸分野との類似が大きい点であった。――寿命が長いこと，特定のグループには重要な情報が少し外れた分野には無用の長物に近い場合が多いこと，かなり膨大な基礎教養的知識が不可欠であり，それらが陽に明示されない場合が多いこと，などである。これらは単なる形式的な類似ではないらしい。学問の性格上の類似があるのではなかろうか？ ――(数学は理学部所属でよいか？)

数学分野の特殊性を過大に強調するつもりはないが，在来の基準評価から，数学関係の文献が(全体として)過小評価されることがないように望みたい。それには数学の専門家がもっと関心をもつことが必要であろう。さらにまた早急に商業ベースに乗り難い分野を「冗費節減」として軽々しく切り捨てることがないよう，為政者に望みたい。

研究会では，文部省特定研究委員会の報告をしたが，近くこの委員会の正式の報告書も刊行される予定なので(数学分野を筆者が執筆)，その分はこの講究録からは割愛した。代りに本解説を執筆した。後掲の検索実例を参照しつつ読んで下されば幸いである。