数理解析研究所講究録
448巻 1982年 12-19

12

日本語データベース作成について
ー数理解析研究所講究録を対象にー

京大 数理研　中司里美

1　はじめに

日本語の計算機処理はハードウェアソフトウェアの両面の発達により普及し，一般化しつつある。京都大学大型計算機センターの情報検索システムFAIRSのデータベースにも漢字データベースCHINA 1が1981年5月公開された。

このような状況のもとで数理解析研究所では共同研究の成果である数理解析研究所講究録の文献データベースを作成することを計画した。検索システムはFAIRS，使用計算機は京都大学大型計算機FACOM M200，プログラム言語はPLIを使用する予定である。

データベースは計画段階にあるが，以下データベース化の対象である講究録の性格，データベース化することの目的、意義，講究録データベースの概要の順に述べ，中間報告とする。

2　講究録の特色

2．1　書誌上の位置づけ

データベース化は現実世界のモデル化の手続きを必要とする。文献の書誌記述については文献についての平面的羅列的な説明というものから，書誌記述の最小のデータ要素への分解と，文献の特性に応じてレコードを構成するデータ要素の選択と構築，という構造的な表現に変わってきている。このことはフォーマットと記述の標準化の動きと共に記録の機械可読化を前提にした，書誌情報交換，オンライン情報検索などの情報システム形成上必要な思想と技術として一般的になってきた。

たとえば UNISIST Reference Manual では，出版物のカテゴリーを，Serials Books Reports Thesis and Dissertations Patent Documents Conference Publications 等に分け，それぞれのカテゴリーに特有の記述すべきデータ要素を定めており，具体的な適用に際しては記録の対象が，物理的にみて Analytic Monographic Collective のどのレベルに属するかによってデータ要素のコードを定めている。

講究録は学術文献のタイプとしては一連番号を持つ不定期の逐次刊行物であり，おおむね集会の記録であり多数著者の論文からなる。すなわち Serials と Conference Proceedings

の性格を持つ。

なお講究録は数理解析研究所要覧によれば「非公式の発表機関であるが書下ろしのものであるべきこと，有益なレクチャーノートの場合もありうること，英文和文は問わないこと，および（共同研究の記録の場合は）編集責任は研究代表者にあること」とされている。

2.2 出版物としての完成度

講究録は出版物としてみると多少不安定な面がある。講究録は執筆者の提出原稿（タイプもしくは手書き）に目次，ページづけを行なったものである。欧米では標題紙に出版物についての情報を集中する習慣があるが，講究録では表紙と目次に情報が分かれており，研究集会についての情報は目次の部分に記載がある。また欧文標題紙が追加されている場合もある。この場合講究録の欧文誌名として RIMS Kokyuroku の記載がある。さらに，書誌記述の原則として情報は資料中に求めるが，和文原稿の場合日本人著者名の読みが不明であり欧文の場合は日本語表記がわからない。

均質の検索を可能にするためには，上に述べたことについて，データを加工する際に注意する必要がある。

3 講究録データベースの目的・意義

講究録の索引は1～100号について著者目録を手作業で編集，タイプ印刷により作成配布した。101号からは，著者カードの形では準備しているが，編集印刷にまでは至っていない。著者目録の作成はひきつづいて行なうことが利用者（研究者，図書館）から望まれている。

当図書室において古いものについては著者がわかっていたりまたかって読んだことがあるがそれが講究録の何番であったかがわからない，ということがよくあり，新しいものについては研究集会の情報はわかっているがそれが何番の講究録となって発行されたかわからないという質問を受けることが多い。前者の場合は著者からの検索，後者の場合は集会の情報または主題からの検索ができればよい。

現在の時点での講究録索引の作成を考えると，機械可読にしておく方が，将来にわたって有効であることは明白である。

3.1 学術情報の組織化との関わり

通常学術情報の種類をレベルで分け，1次情報（原著論文）2次情報（Bibliography Index 等）に分けるが，1次情報への到達手段であり，次の1次情報生産につながる，2次情報のデータベース化とオンライン情報検索は，学術情報の組

織化の必須条件である。国際的なレベルで主題別にデータベースの作成と検索サービスは行なわれているが，一研究機関において発生した学術情報を加工しデータベースとすることは，主題別の巨大データベースの存在意義とは別に必要なことであると思われる。

3.2 数理解析研究所の他のデータベース

当研究所はRIMS，PICMSの2つのデータベースを構築保守し，京都大学大型計算機センターの情報検索システムFAIRSのデータベースとして公開されている。

RIMSは当研究所の他全国10大学の数学教室の入手したレクチャーノート，プレプリントの書誌事項を収録した文献データベース，PICMSは当研究所の数学を中心とした会議録について同じく書誌事項を収録した文献データベースである。主題分析は行なっていないが，所蔵情報はある。

RIMSの収録対象は1次情報以前の情報が多く，既成データベース，2次資料には収録されず，資料の収集もしにくいものであり，数学関係研究機関相互の資源の共有を前提とした情報の共有の考えのもとに従来は手作業でリスト作成が行なわれていたものである。大学間大型計算機センターネットワークの普及により，より使いやすいものになると思われ

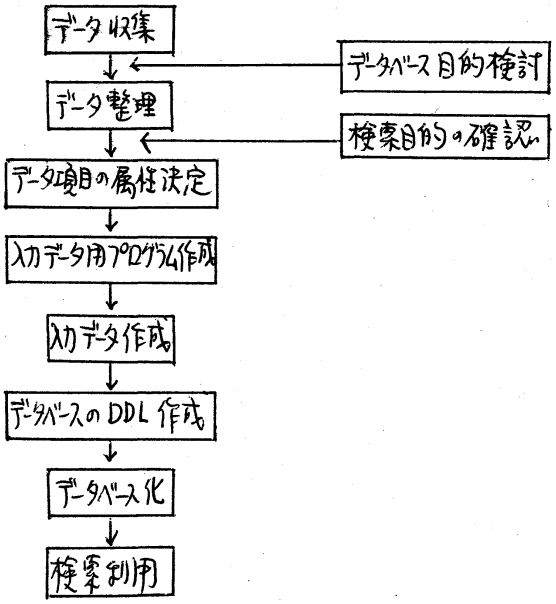
る。

PICMSは会議録の多角的な検索にはデータベース以外にない，という前提で試作されたもので，RIMSに比べ，項目は多く構造が複雑であるが，データの形式についてのモデルを考える上で役に立った。収録対象は通常の単行本も多いので，データ入力については将来既成の文献データベースたとえばMARCなどの利用を考えた方が作業が節約でき，質の良いデータベースになると思われる。

4 講究録データベースの概要

4.1 データベース作成手続き

データベース作成の過程は下図のようになる。

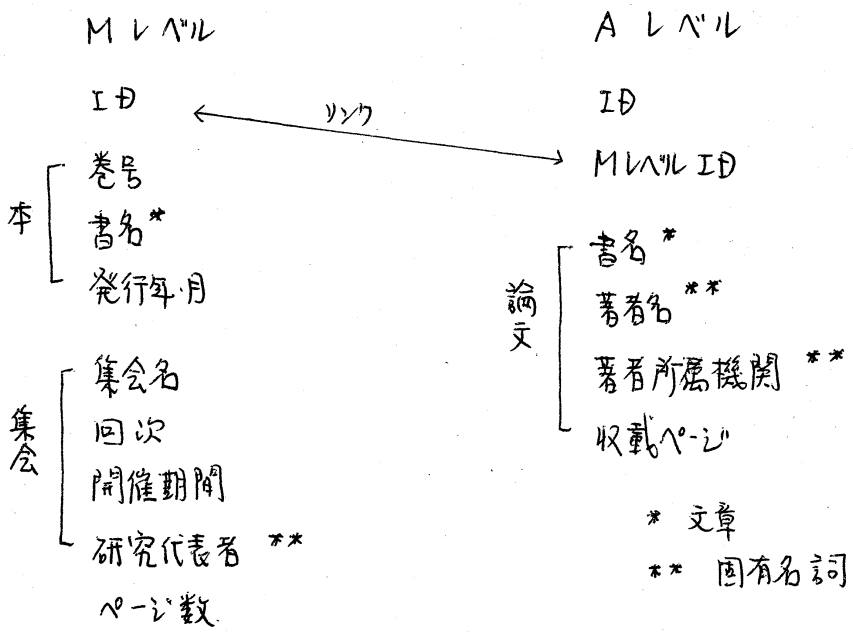


4.2 データ整理上の問題点

入力用データ整備の段階での問題点は，検索を想定して，必要な工夫をどの程度行なうかである。主題分析を行なうこと，和文，欧文の両方のテキストがあるので検索用に表記を統一した項目を著者，書名について用意すること，という方針でデータを準備している。その段階で著者の抱いた困難を感じている事情は既に述べた。

4.3 データ項目・属性の定義について

講究録データベースは2.1に述べたレコードのレベルに分けるとそれ自体がSレベルにあり本単位のレコード（Mレベル）論文単位のレコード（Aレベル）からなり，レベル間のリンク用項目を必要とする。大体項目は次のようになる。



＊は文節，＊＊は固有名詞の部分であり，検索用に工夫の必要な部分である。

5　おわりに

データベース作成はまだ準備中であり，項目の定義・属性の詳細，検索の実例等が紹介できないのは残念である。公開に至った段階で，それらの部分を，この報告の後半部として発表したいと思う。