

公文書型目録データベースの構築

沖縄県公文書館 大城 博光

はじめに

目録データベースは主に、「情報資源の検索を容易にする道具」として用いられます。公文書館で扱う目録データベースは、検索用途の他に「情報資源を継続して管理する道具」として必要になります。これは、公文書館が他の情報管理機関と比較すると、その情報資源を保管する期間が圧倒的に長いためだと思います。

公文書館の情報資源、すなわち公文書等の所蔵資料は、時とともにその管理者が変わります。公文書館を継続的に運営していくには、所蔵資料の背景となる情報、受入先との契約事項、記録媒体の管理方法等も所蔵資料と共に次世代の管理者へ継承していかなければなりません。その役割を適切に果たすために、個々の書類を分散して管理するのではなく、目録データベースという「情報資源を継続して管理する道具」に集約して引き継いでいく必要があります。

沖縄県公文書館は、平成7年の開館時に構築した目録データベースを、約5年間の運用を経て平成12年にリレーショナル型のデータベースとして再構築しました。本稿では、沖縄県公文書館の目録データベースの構造及び特徴的な機能等を設計方法とともに紹介します。

1 記述要素と記述単位

目録データベースの構造は、行・列で構成される表で表現されます（ここでは、行を記述単位、列を記述要素と称します）。目録データベースの入力情報となる記述要素を決めるには、公文書館業務で日常的に使用している検索用目録、受入台帳、契約書類、修復カルテ、その他文書・帳票類等の紙ベースの出力情報を洗い出すことから始めます。この出力情報を個々の情報に分解したものと、目録データベースの役割を定義した上で、新たに追加すべき情報を見直して、目録データベースの記述要素を決めます。

次に、従属性の強いものごとにグループを再編成し、記述要素が唯一になる

グループの記述単位を決めます。この設計作業は、記述単位の異なる記述要素を同一の目録データベースに収める時に、無駄な繰り返しの記述を省くためです。たとえば、媒体情報と受入情報を1つのデータベースに同じ記録媒体位単位で記述すると、受け入れた媒体の数だけ、受入情報を繰り返し記述しなくてはなりません。目録データベースからこのような冗長性を排除したグループ編成例を表1に示します。

表1 記述要素のグループ編成

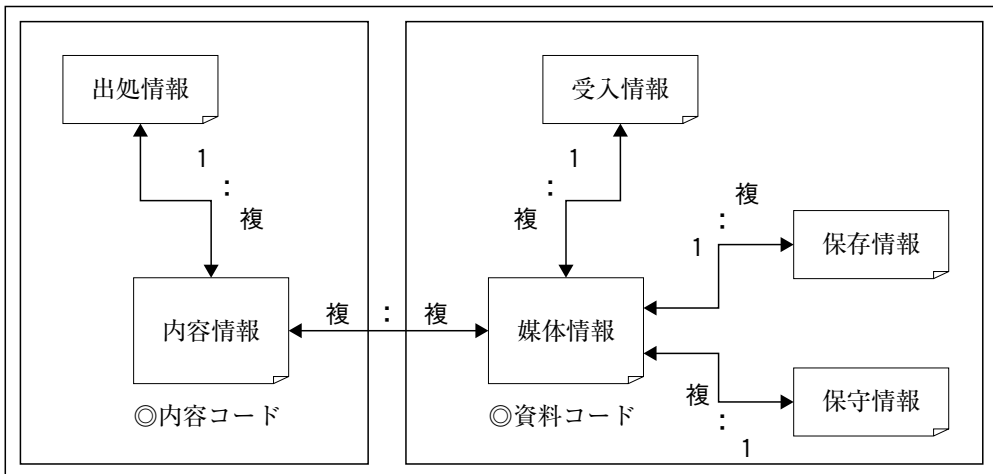
グループ	記述要素	記述単位
出处情報	資料の主管部署の業務、組織体制等	資料の主管部署
内容情報	資料の主題、作成日、作成者、内容等	資料の内容
受入情報	受入経緯や契約事項及び権利関係等	資料の受入
媒体情報	媒体の種類、形状、所在、公開等	記録媒体
保存情報	媒体の保存状態、修復措置の履歴等	記録媒体の状態
保守情報	媒体の保守周期、方法等	媒体の種類

* 公文書館の所蔵資料は、数件の文書を取めている簿冊、フォルダーあるいは写真アルバム、マイクロフィルム、電子記録など、形態が多様なため、内容情報の記述単位は、形態毎に具体的に決める必要があります。

2 データベースの管理構造

表1のグループ編成に基づき、その情報の格納庫となるデータベースをグループ毎に作成します。

図1 データベース間の相互関係



* 図中の1：複や複：1等は記述単位の相互関係を示します。内容情報と媒体情報における複：複の相互関係を、リレーショナル型データベースに実装するには、制御用のデータベースが必要になります。

リレーショナルデータベースは、複数のデータベースに分散された記述要素を、利用目的に合わせて自由に組み合わせ、更に、相互関係に基づいて必要な数だけ繰り返して、出力情報を表現できます。また、記述要素の唯一性が保たれていることは、記述要素を変更した場合には、利用している全ての出力情報に反映され、より正確な情報の管理をもたらします。

この管理構造による特徴的な情報管理機能の一例をご紹介します。公文書館の所蔵資料には、原資料の他に、マイクロフィルムや電子記録のように、複数の資料を1つの媒体に集約した複製物を保存している場合があります。このような記録媒体は、媒体情報が"1"に対し、内容情報が"複"の相互関係になります。媒体情報は、媒体を一意に識別する媒体コードで管理され、内容情報は、図書に使用されるISBNコードのように、資料の内容を一意に識別する内容コード（同じ内容の資料には同じコード）で管理されます。この管理コードによって、資料タイトル等の内容情報を特定した場合、その内容情報に関連付けられている複数の媒体情報（所在等）を取得することができます。一方、記録媒体を特定すると、その媒体情報と、媒体に記録されている複数資料の内容情報（資料タイトル等）を取得することができます。この管理構造は、内容情報を記録媒体の記述単位に制約されることなく記述することができ、閲覧用や保存用などの用途別の複製物も適切に管理することができます。

3 検索用目録の表示配列

目録データベースから情報を出力する場合、その情報量が多い帳票類には、紙目録、コンピュータディスプレイのどちらにおいても、出力表現に規則的な配列が必要になります。例えば電話帳は、利用者が目的の電話番号を調べる際に、該当者の氏名と市町村を把握していることを前提に、都道府県、市町村毎の階層的分類を設けて検索の範囲を限定し、その範囲内で氏名を50音順で配列しています。では、公文書の検索用目録を出力する場合には、どのような配列が適しているのでしょうか。公文書は、行政組織がその業務を遂行する上で作成・収受した文書等です。公文書の検索者は、資料タイトル等を明確に把握している事は少なく、「目的の事案を所管していた部署はどこか」という視点で検索範囲を絞り込む方法が効果的に思われます。このような検索手段には、行政組織の体制や所管業務を体系的に示した出处情報を分類として提供する必用があります。さらに、出力情報の記述要素は、検索者の意図する情報であるか

否かの判断に最も適用しやすい資料タイトルを昇順に、さらに同一資料タイトルの場合には、作成日の古い順に並べて配列します。

4 順次検索と横断検索

大量の情報の中から必要な情報を検索する場合に、順次検索と横断検索を組み合わせる方法が画一された検索手法として多く使われています。

たとえば現代用語辞典などは、主題を階層的に細分化した目次（順次検索）と用語を50音順に並べた索引（横断検索）の両面から目的の用語を探せるように工夫されています。また、コンピュータを用いた情報検索でも、インターネットポータルサイトのように、階層的なカテゴリ（順次検索）とキーワード（横断検索）の複合条件による検索手法が使われます。

目録データベースから所蔵資料を検索する場合においても、横断検索は、検索者の想定する検索語が記述要素に含まれる情報を瞬時に抽出することができますが、目的の情報の記述要素の中に利用者の意図する検索語が含まれない場合や、全く目的外の情報が大量に抽出される場合もあります。よって、前項で紹介した分類による検索範囲の限定や、規則的な並びに従って人間が漏れなく照合できる順次検索と、検索語の機械的な高速抽出による横断検索を組み合わせることにより、より精度の高い検索結果を提供することができます。

5 標準様式への対応

ここまで、目録データベースを構築する時のポイントとなる記述要素、記述単位、グループ編成及び配列等について述べてきましたが、それらは、目録データベースを保有する公文書館の方針や機能によって異なってきます。一方、情報の共有、相互運用、操作性の統一化等の目的で公文書館型の目録データベースが標準化されることは、意義のある事だと思えます。

図2 検索用目録の分類と並び

