

大谷大学所蔵チベット語図書目録データ
ベースの構築について

那須 和樹

目 次

| | |
|---------------------------------|-----------|
| 1 はじめに | 1 |
| 1 制作物について | 1 |
| 2 必要としている人々 | 1 |
| 3 制作に向けて | 2 |
| 2 制作の流れ | 3 |
| 1 データベースの定義 | 3 |
| 2 テキストファイルをデータベースにする | 4 |
| 3 PHP について | 6 |
| 4 言語の問題 | 13 |
| 3 まとめ | 14 |
| 1 データベース検索システムを作成してみた | 14 |
| 2 今後の課題 | 18 |
| 3 終わりに | 20 |

1 はじめに

(1) 制作物について

この論文の題目にある大谷大学所蔵チベット語図書目録データベースは、ゼミのテーマ「人の役に立つものを作る」に基づいて制作されたものだ。これは、大谷大学が所蔵している書籍のうち、チベット語文献だけを集めたデータベースを作ろうというものだ。なぜ、このようなものを制作するのかというと、現状では、大谷大学が所蔵しているチベット語文献をうまく探す事ができないのが理由である。

そのため、チベット語図書目録データベースを作る事で、大学が所蔵しているチベット語文献を探しやすくする事が今回の目的となる。このデータベースはどこからでもアクセスでき、誰でも利用できるものになりたいと思う。チベット語文献をうまく探す事ができない理由については後に述べる事にする。

(2) 必要としている人々

チベット語図書目録データベースを作るにあたって、このデータベースを利用する人がいなければ、せっかく作成しても意味が無い。どのような人が対象になるのか、それを明らかにしないことにはチベット語図書目録データベースの存在意義が問われてしまう。このデータベースは、大谷大学の図書館が所蔵するチベット語文献を利用する人々を対象としている。

チベット文学や仏教学の研究者はもちろんの事、それ以外にも学生やチベット語に興味がある人など、チベット語文献を必要としている人はたくさんいるはずだ。そのような人々が、チベット語文献をうまく探す事ができるようにするために、このデータベースは制作されるわけで

ある。

(3) 制作に向けて

まず、チベット語図書目録データベースを制作する前に、現在、図書館でチベット語文献を探すにはどうしているのかを、述べていこうと思う。図書館にチベット語文献のデータベースがないのかといえば、全くない、ということではない。図書館には、すでに蔵書検索システムがあるのである。そしてそれは、図書館のデータベースにある本を全ての項目を対象にしてキーワード検索が可能になっている。

もちろん、言語がチベット語の文献だけを条件にして検索する事もできる。しかし、いざチベット語文献を検索してみても、うまく思い通りの成果が得られない。出てくるデータの数が足りない。図書館にある蔵書検索システムは、所蔵する全ての文献のデータが入っており、全体的な項目で分類されたデータベースが使われている。そのため、チベット語文献のデータはあちこちに散乱してしまっている。

チベット語文献だけで分類されているわけではないので、全体的な分類によって、チベット語文献はそれぞれの内容に即した分類の所に散らばめられてしまっている。様々な文献の分類のため、仕方の無い事なのだが、データの方もチベット語文献はあちこちに散乱しているので、チベット語文献だけを検索する事がやりにくくなっている。

言語をチベット語で検索しても、うまく検索に掛からないものもある。このままだとせつかく所蔵されている本が見つからず、必要としている人の作業の妨げになったり、検索されなかった目的の本が有効利用されないまま終わってしまう。そこで、チベット語文献だけのデータベースを作り、その検索システムを実装することとなった。

2 制作の流れ

(1) データベースの定義

チベット語文献のデータベースを作るに当たって、まず、あちこちに散乱しているチベット語文献のデータを集める必要があった。そこで、図書館に依頼し、チベット語文献だけのデータをテキストデータで受け取ることにした。最初に貰ったデータには次のような項目があった。所蔵_所蔵 ID・書誌_書誌番号・出版地・出版社等表示・所蔵_表示用請求記号・所蔵_備考・所蔵_配架場所 C である。このデータを元に、SQL 文に変換してデータベースを作ろうということだ。

しかし、最初にもらったデータでは、項目が少なく、データベースとしては物足りないものであった。そこで、足りないものは何かを探し出し、再度図書館にデータの提供を求めた。このとき足りないと思われた項目は、形態・著者名・注記・その他のタイトルである。次にもらったデータは、著者のデータはなかったものの、求めた項目のうち他の項目はもちろん、さらに別の項目も入っていたので、これでデータベースを作ることに決定した。

データベースで定義したテーブルにはタイトルテーブルと所蔵テーブルの二つがある。二つテーブルを定義した理由については次のセクションで述べる事にする。次に、定義したフィールド（項目）を挙げる。これらについても、次のセクションで説明しているものがある。タイトルテーブルには ID・書誌_書名・著者・巻書名・版・出版地・出版社等表示・出版年月等・数量・挿図・大きさ・書誌_叢書名・親書誌番号等・注記・所蔵_備考がある。所蔵テーブルには ID・1 巻の ID・所蔵_所蔵 ID・所蔵_表示用請求記号・配架場所がある。これらのテーブルとフィールドでデータベースを定義するファイル「database.sql」を作成した。

(2) テキストファイルをデータベースにする

しかし、図書館からもらったデータにもまだ問題が残っていた。まず、複数巻からなる図書のデータが全て別々のデータとして入っており、特に巻数の多い図書については、ID だけが違う同じデータが大量に並ぶようになっていた。さらに、これらのデータが一カ所に固まっていないものもあり、あちこちに散乱しているものもあった。

そこで、まず、散乱しているものを整理するために、Excel を使って、所蔵_所蔵 ID を基準にしてデータを並び替えた。その際、SQL ファイルに変換しやすくするために、タブ区切りのファイルにした。同じデータが並ぶのを防ぐために、SQL ファイルはタイトルと巻書名が同じ文献を一つにまとめたタイトルテーブルとタイトルの重複に関係なく、全てのデータで異なる項目をまとめた所蔵テーブルに分けて作成することにした。

タイトルテーブルは、複数巻からなるものを一つだけ取り出せるように、タイトルと巻書名をセットにしたものが同じであるデータを一つと考え、順番にデータを入れていく過程でこのセットが異なるデータが出るたびに連番を一つずつ増やしていき、これを ID とした。所蔵テーブルは、複数巻からなるデータを、タイトルテーブルに割り当てられた ID と照合することで、リレーションを組むことにした。所蔵テーブルでタイトルテーブルの ID を 1 巻の ID、つまりその種類の本の ID としてフィールドを作った。これとは別に、所蔵には連番の ID も存在する。この二つのテーブルの項目については資料 1 「database.sql」を参照してほしい。

データベースの項目が決定したので、次に、図書館からもらったテキストデータを実際に SQL 文に変換しなくてはならない。その際に、タ

タイトルテーブルと所蔵テーブルに分けて変換する必要がある。Python というスクリプト言語を使って、変換のためのプログラムを作る事にした。まずは、テキストデータをタイトルのデータと所蔵のデータに分けるために sort.py というプログラムを使った。sort.py は

もし書名と巻を文字列足し算したものが、前のものと違っていたら、タイトルテーブルの連番を一つ増やす。→タイトル ID
タイトルテーブルを書き出す。

書名と巻を足したものを、一つ前のものとして変数に保存。

ID を新設して、所蔵 ID と請求記号と配架場所を所蔵情報テーブルに書き出す。

そのとき、タイトル ID フィールドにタイトル ID を書き込む。という方法で、タイトルテーブルと所蔵テーブルに分けるプログラムである。このようにして、テキストデータをタイトルデータと所蔵データに分ける事ができた。(資料 2 「sort.py」参照)

次に、タイトルテーブルと所蔵テーブルの二つを SQL 文に変換する必要がある。変換に使うプログラムは、それぞれ別に用意する事にする。これらのプログラムはタブ区切りの項目を insert 文に当てはめる方法で SQL 文に変換している。一行ずつ読み込み、改行文字を取り払い、タブで区切って項目を順番に入れていった。これは、タイトルテーブルも所蔵テーブルも同じ方法である。項目の種類が違うので、別々に変換プログラムを作る事になった。こうして出来たのが、「title_table.sql」と「syozo.sql」である。(資料 3 「title_insert.py」、4 「syozo_insert.py」参照)

(3) PHP について

(i) PHP でページを作成

データベースが出来上がったので PHP ファイルで Web ページを作っていくことにした。各ページのデザインは CSS で一括して作成した。この PHP ファイルで作ったページには次のような機能がある。まず、蔵書検索・一覧表示・詳細表示の機能があり、それぞれ、データベース上の文献のタイトルを単語で検索、検索なしで全てのデータの一覧を表示、巻書名や注記などの詳細データを表示するという機能である

また、データベースを修正するための機能も備えている。データの編集機能と、データの削除機能がこれにあたる。編集機能は、詳細データを見て、足りないデータや間違いを修正するために使う機能である。分類検索のための分類の項目は今回追加したもので、既定値が入っていないので、分類の項目を入力するのもこの編集機能を使うことになる。編集した後、直前に表示していた一覧をすぐに表示できるようにするために、詳細ページを開くときに新規ウィンドウで開くようにした。この機能は後に、三宅先生の意見を反映して導入したものである。(資料5「編集画面」参照)

削除機能は、一覧や検索結果を表示したときに、タイトル名の左側にあるチェックボックスにチェックした状態で、削除のボタンを押すことによって、ページに表示されている項目を好きな数だけ一度に削除できる。一度に最高 100 件 (20、50、100 から選択) まで表示できるため、削除するときにチェックする文献を間違えているかもしれないので、削除するデータを確認できるページも作った。このページでもチェックボックスを使って、取捨選択できるようになっている。

削除したデータを元に戻す機能も備えている。削除したかどうかは

データベースの項目の一つに削除したデータかどうかを判断する項目を入れたものなので、削除したデータだけを表示する削除一覧のページがある。このページでは検索はできず、一覧の表示のみで、削除したデータを復帰する方法は、データを削除する方法と同じで、チェックボックスを使って選択して実行する形式になっている。また、削除一覧からでも詳細データを見ることができる。

作成したページをオフラインで実際に動かすために、XAMPP という環境を用意した。これは、ローカルをサーバーの代わりとして扱うことのできる環境で、PHP プログラムを扱うことができる。PHP プログラムはこのように、専用の環境を用意しないと動かせないのである。

この XAMPP という環境は Windows 用のものも存在しており、Windows のパソコンでも PHP プログラムを扱うことができるが、私の家でやってみたところ、アルファベットと数字以外（おそらく漢字）がうまく表示できなかった。

(ii) 表示上の問題

実際に、PHP で作成したページを動かして、一覧表示をしたり、検索をしたり、詳細データを見ていくことによって、データベースに関する様々な問題が見つかった。

まず、タイトルや注記で通常と形状の異なるアルファベットの文字が「_」に文字化けしており、正常に表記されていない問題が見つかった。これは、エクセルでタブ区切りのファイルを作る際、ファイルの種類を「タブ区切り」にしたときに起こったようである。これは、エクセルでタブ区切りのファイルを作るときに、ファイルの種類を「タブ区切り」ではなく、「utf-16」の形式で保存する事により、文字化けを防ぐ事ができた。

次に、英文タイトルの書名の中にある「'」が引用記号と取られ、エラーを起こす問題が見つかった。この問題は、タイトル中の「'」を「\'」に置換して、引用記号と取られないようにすることで解決する事ができた。置換の方法は、タブ区切りのデータの時点で、「'」の場所を検索し、「\'」に全て置換で行った。

さらに、注記などに「”」が必要以上に残るという問題も発生した。これは、タブの周りにある「”」が影響していたようで、「\t”」を「\t」に、「”\t」を「\t」に置換する事によって、解決する事ができた。これも、タブ区切りのファイルの時点で、「\t”」や「”\t」を検索して、全て置換で「\t」に置換した。

また、責任表示（著者名）が無いという問題もあった。これでは一人の著者の文献をまとめて検索することができず、タイトルと著者で絞り込んでの検索も出来ない。しかし、これはデータの中身の問題であり、私がどうこう出来る問題ではないので、このまま保留する事になった。

こうして、「title_table.sql」と「syozo.sql」を作り直すことになった。ちなみに、SQL ファイルが複数あるので、MySQL に読み込ませる場合は、データベースの定義である「database.sql」を読み込ませ、その後、「title_table.sql」、「syozo.sql」の順に読み込ませる。これまでに作った PHP ファイルとデータベースは大学の Web サーバー上に置き、どこからでもアクセスできるようにした。こうして、ファイルを作成したパソコンのローカルでログインしなくてもチベット語文献のデータベースを見る事ができ、データの編集ができるので便利になった。

大学のサーバーにファイルを置く方法はまず、ターミナルで大学のサーバーにファイルを置くためのフォルダを作る。そして、「Cyber Duck」という FTP ソフトを使って大学のサーバーの指定のフォルダに

ファイルを送る事ができた。サーバーに置いてあるファイルもその場で編集できるので、何か間違いを見つけたときにはすぐに修正ができる。

どこからでも作成中のデータベースにアクセスできるようになったので、Windows のパソコンでアクセスしてみたところ、「Safari」と「Internet Explore」の表示の仕方の違いが判明した。検索結果の表示のところで、「○件見つかりました。」というのが、「Safari」ではきちんとテーブルの上に表示されるが、「Internet Explore」では下に表示されるのである。「Internet Explore」は不具合が多いようなので、「Safari」に合わせてこれはそのままにしておいた。

(iii) 機能の追加とそれによる問題の解決

これで、ひとまずデータベースを構築することが出来たので、チベット文学の研究をしておられる三宅先生にデータベースの検索システムを見てもらうことにした。そこで、出された意見に次のようなものがある。

- ・ 書名検索を 2 語の組み合わせの AND 検索を可能にする
- ・ 書名だけでなく、著者、出版社、注記でも検索できるようにする
- ・ 検索結果画面の書名の表記のところを、検索文字列と一致した箇所の色付けをする

- ・ 詳細情報のリンクを書名の部分に貼る

AND 検索機能と検索項目を増やす事については、まず、検索するページに項目を増やし、分類検索も同じページに入力するようにした。(資料 6「検索画面」参照)そして、検索結果を表示するページでの処理は、そのページの PHP プログラムで、各項目ひとつずつ、何か入力されているかどうかで分岐し、何か入力されていたら、その都度「and フィールド like 検索内容」という SQL 文を追加していき、最後にデー

データベースを読み込ませる SQL 文の後ろにつなぐやり方で実装した。(資料 7 「kensaku.php」 参照)

検索結果の書名の検索文字列と一致した箇所の色を変えるところは、最初は、`str_replace()` を使って処理をしたが、このやり方だと大文字小文字が完全に一致した所しか色が変わらなかった。検索結果には大文字小文字を区別せずに検索されるので、これでは都合が悪い。そこで、別の方法をとることにした。`str_replace("/検索文字列/", ~)` となっていた所を、`preg_replace("/検索文字列/i", ~)` とすることによって解決する事ができた。(資料 8 「検索結果」 参照)

この書名の所にはリンクを貼っており、一度クリックした所は色が変わるので、正常に表示されるかが心配されたが、実際に動かしてみた所、問題なく表示された。

しかし、今度は、検索文字列が半角 1 文字だと検索結果がおかしくなるという問題も発生した。例えば、「a」という文字一文字で検索したときに、漢字のタイトルの文献が検索で引っかかるのである。これは全角 2 バイトで 1 文字として処理しているため起こった問題である。そこで、検索文字列が半角 1 文字の項目があるときにエラーメッセージを返し、検索できないようにした。

さらに、検索文字列を何も入力しなかった場合、全件表示するという状態でもあったので、検索語句が全てからの場合もエラーメッセージを返すようにした。これらの処理は javascript によって行った。いずれかの項目に、半角一文字だけの文字列が入力されているとき、またはどの項目にも文字が入力されていないときに検索ボタンを押すとエラーを返すようにした。(資料 9 「kensaku.html」 参照)

(iv) 分類検索

蔵書検索機能の一つに、分類での検索がある。これを使えば、ある分類の本だけを全て表示したり、あるキーワードを含む文献のうちある分類のものを検索するために使うことができる。それにより、検索する人が必要でない分類の本が検索にかからないので、目的の文献が探しやすくなるのである。もともと、分類の項目はデータには存在しないので、データベースで先に定義しておき、編集機能を使って、後で一つ一つ追加する事になる。

しかし、分類での検索の部分で問題が起きたのである。分類のフィールドは数値のデータで作られているのだが、既定値を入力していなかったため、フィールドの中が全て NULL になっていた。ブラウザ上での表示に関しては、数値が 0 または NULL の時に未分類、としておけば問題ないのだが、検索にかけるときに、`bunrui = NULL` では検索で取得されなかった。この問題は `is null` 演算子を用いる事によって解決する事ができた。

なぜ、未分類のものを表示する必要があるのか。データベースが完成されていて、全ての文献に分類が割り当てられているならば、未分類のものはないので、この処理は必要ない。しかし、分類の項目はこれから一つずつ入れていかななくてはならないので、未分類のものだけを探して順番に入力していく事が考えられるからである。

これで、「未分類」で検索した時に、数値が 0 のものと NULL のものを両方表示する事に成功した。先ほど、分類のフィールドは数値になっているといったが、分類名は数値ではない。分類名を表示する方法は、プログラムの最初の方であらかじめ分類名の配列を作っておき、表示する際にその配列に数値を当てはめる事で、表示する分類名を決定してい

る。分類の項目はある程度表示するものが決まっているので、編集時にいちいち分類名を入れるよりは効率がよく、間違った分類名を入れてしまう心配も無い。

(v) データベースの修正のための編集と削除

これで、検索システムは完成したわけだが、データベースにはまだ不備が残っている。そのため、編集機能に加え、削除機能を追加した。まず、編集機能についてだが、最初に作った状態では、データの編集画面に行くためには、「検索結果」→「詳細画面」→「編集」の順で行く事になる。そうすると、データの編集を行った後、また編集結果に戻りたい場合は、再度検索する必要があるのである。これでは大変不便なので、詳細画面を出す時に新しいウィンドウで表示するようにした。こうすれば検索結果の画面を残したまま編集する事ができ、作業も効率よく進められるようになる。

次に、削除の実現方法について説明する。削除の方法は詳細画面で1データずつ行う事もできるが、チェックボックスを使い、検索結果に表示される1ページ単位での削除も可能にした。(資料10「削除の方法」参照)1ページに表示する件数は20, 50, 100件から選択できるので、一度の削除できる数もこの中のいずれかである。削除する際には、直前のページでチェックしたデータだけの一覧を表示して、削除の確認をする画面もある。これは、どのデータをチェックしたのかを分かりやすくするために実装した。

同様に、削除したデータを元に戻す機能も追加した。このプログラムの内容は、削除と書かれた表記と、削除するプログラムの部分を変えただけのものだ。なお、削除の方法としては、データ自体を消してしまうのではなく、削除されているかいないかを、true、falseで判断して表

示する項目を決定する方法である。これにより、データの削除と復帰が容易にできるようになった。さらに、この方法だと、復帰した時にきちんと元の位置にデータが戻る、すなわちデータの並び順が変わらないので、削除や復帰をしたことでこれ以上データがばらばらになることは無い。

しかし、削除の判定するフィールドは、最初は予定されていなかったもので、データベースで定義されていなかった。定義されていないので、プログラム中で判定ができず、一覧の表示さえ機能しなくなってしまう。そこで、データベースに追加の定義をする SQL ファイル「add.sql」を作成する事にした。まず「use book」でデータベースを選択し、次に「ALTER table book ADD del BOOLEAN」で削除の判定をするフィールドを追加した。既定値を削除していない状態、「false」にする必要があったので、「insert into book (del) values (false)」で全てのデータのフィールドに「false」を入力した。これで、一覧には削除していないデータが並び、削除一覧には削除されたデータが表示されるようになった。

(4) 言語の問題

このデータベースはチベット語文献のデータベースである。しかし、図書館からのデータにはチベット文字は含まれておらず、漢字表記（中国語）や英語表記、ローマ字表記（チベット語のローマ字転写）のデータだけである。しかし、現状どうなるものでもないので、このまま行く事になった。文字コードについては、特殊な字体のローマ字もあるので、どの言語でも表示できるよう unicode (utf-8) で作成する事にした。

チベット語文献データベースを作っている Mac には「utf-16」もある

のだが、普及率の高い Windows で利用できるように文字コードを「utf-8」に決めた。これにより、Windows 標準のブラウザである「Internet Explore」でもだいたい文字化けせずにこのデータベースを利用することができる。

3 まとめ

(1) データベース検索システムを作成してみる

チベット語文献だけのデータベースを作成し、チベットの専門家である先生達に見てもらい、話し合った結果、これまで見つからなかった多くの問題が浮上してきた。これまで述べてきた問題も、ほとんどがこのような専門家の方によって判明したものである。

チベット語文献のデータベースを作成するにあたって、技術的なものに関しては去年、授業でやってきたデータベースと PHP の応用のような位置付けだったので、スムーズに進めることができた。データベースの定義や MySQL の使い方はもちろんのこと、文字列での検索やデータの編集、削除も授業でやった事があった。XAMPP の使い方も授業でやっていたので、問題なく扱うことができた。授業でやってきた経験を今回の制作に生かす事ができたのがよかったと思う。

また、この制作を通してわかった事は、最初はデータベースを作ると聞いて、何かの資料を元に手作業でデータを入力していくものかと思っていたが、テキスト処理のプログラムを使い、自動的にデータベースを入力するためのファイルを生成していくというのがわかった。だからこそ、最初のファイルできちんとした基準を作り、データ全体を統一させる事が重要だと知る事ができた。

今回、図書館からもらったデータは、一つのデータが「”項目”, ”項

目”・・・」というように形式はきちんと統一されていたが、それぞれの項目の内容が統一されていなかった事が原因で、様々な問題が起こっていた。それにより、例えば、注記にタイトルの一部が移っていたり、「”」や「'」が不自然に残っているという現象が起こったのである。

(i) タイトルの問題（言語）

まず、タイトルにまつわる問題から挙げていきたいと思う。タイトルの言語に漢字のもの、英語のもの、ローマ字転写のチベット語のものなど、言語がバラバラになっているという点がある。ローマ字転写については、通常のキーボード入力で出せるローマ字のものと、文字に点や線のようなものが入っているローマ字のもの2種類が存在している。

さらに、つづり自体が間違っているものもあり、目的の文献を検索で簡単に出せない、または検索できない状態になっていた。つづり間違いは単純に誤字・脱字という事も含むので、検索で出せない以前の問題である。そして、この問題を解決するには、一つ一つ間違いが無いか地道に調べていかなければならない。

また、言語が違うが、チベット語でのタイトルが同じ本は、言語が違えば登録されている言語で検索しなければならないし、どのタイトルの本がどの言語で登録されているかは、実際にそのデータを見るまで分からない。しかもそれらは、色々な位置に点在しているので、一覧から探し出すのは困難である。

一覧の並び方が、アルファベット順など、特定の規則で並んでいれば、一覧から探す事も不可能ではないが、一覧に並んでいるデータの表示順はアルファベット順ではなかったので、もし、一覧から探すとなると全てのデータを順番に見ていかななくてはならない。これでは、せっかくの検索機能も何の役にも立たないのである。

そもそも一つのデータベースに複数の言語のデータが混在している事が問題であり、チベット語文献のデータベースであるのだから、漢字表記や英語表記では使えない。チベット文字、またはチベット語のローマ字転写での表示にするべきであり、ローマ字転写にするのであれば、転写方法を統一する必要がある。最初にテキストデータを作るときに、チベット語が分かる人物が加わっていれば、これらの問題は多少解決されていたかもしれない。

また、元々はチベット語の文献でも、他言語に訳されている本は、少なくともチベット語で書かれた文献ではない。これらの本をチベット語文献に含むのかどうか、含むとしたらそのままでは言語が違うので、どのようなデータにするのかを考えていかななくてはならない。

(ii) タイトルの問題 (表記)

他にもまだ、タイトルにまつわる問題は存在する。タイトルが途中で切れていて中途半端のものや、タイトルとしては相応しくないチベット語の単語がローマ字転写で書かれているものもある。タイトルが中途半端の文献の中には、注記に残りのタイトルが入ってしまっているものもある。このため、書名検索に掛からなかったり、注記が無駄に長くなったりしている事になる。しかしながら、研究所にある文献についてはタイトルが最後まで入っている。

また、ローマ字転写の仕方の問題に、文節の区切りがスペースのものとハイフン (-) のものがあるという問題がある。この場合、複数の単語をスペースで区切って検索した場合、文節の区切りがスペースのものは出るが、ハイフンのものは検索に掛からない。AND 検索は二語までしか対応していないし、複数の語で一つの言葉という場合もあるだろう。よって、この表記方法は統一しなければならないのである。

また、全集で一巻ごとにタイトルが少し違うものがある。そのデータはタイトルが完全に一致していないので、全て別データとして取られ、巻数分データが並んでしまっているのである。全集は巻数が多いものがあり、検索の一覧に表示されるデータの数が多くなってしまう。ものによっては、一部だけ別データになっているものもある。本来ならばもっと一覧になるデータの数は少なくなるはずだ。このままでは、ある全集の特定の巻を探し出すのに、共通した単語で検索すれば、大量の検索結果から探すことになり、また、所蔵データを見るためにも詳細データを一つずつ確認していく必要があり、それらのデータもほとんど同じ内容なので別々に分かれている必要性もない。このままでは不便で、まとめであった方が見やすいので、これも修正する必要がある。

タイトルテーブルと所蔵テーブルに分けたのにもかかわらず、タイトルテーブルに同じタイトル、または同じようなタイトルの文献が連続して並んでしまっている。どうやら、データがあちこちに散乱していたため、データの統一性が無かったのが原因のようだ。言語が統一されていないので、違う言語で同じタイトルのものは表記が全然違うので、別物扱いされているのである。

違う言語で同じタイトルのものの表記が違うというのは、例えば、りんごだと日本語では「りんご」だが、英語では「apple」である。このように、意味はまったく同じものでも、言語が違くと表記される文字は全然違う。本来のチベット語でのタイトルは全く同じであっても、文字が違うという事は検索しても一致しないので、同じ本を同じ方法で検索する事ができないのである。カードにはチベット語のタイトルが出ている。

(iii) その他の問題

チベット語文献には、どの目録にも出ていないものがあることを三宅先生が指摘された。どの目録にも出ていないということは検索で探す事は現状不可能という事になる。また、普段入れない部屋が配架場所になっている文献もあり、これは逆に検索で出す事ができても、利用しづらい。

他に、著者名が表記されていない、分類検索は今回作成したので分類が設定されていない、などの問題がある。分類はこれから入れていけばいいとして、元々、図書館の蔵書検索でのデータの詳細では、著者が記されていたので、今回それがないのは問題である。検索のシステムが出来ているのに対し、データベースが完成されていないので、データベース検索システムとしては成り立たないのである。

(2) 今後の課題

今後、解決すべき課題としては、この問題だらけのデータベースを作り直す必要があるだろう。一覧の表示順はエクセルでタブ区切りのファイルでソートするか、データベースを選択する SQL 文で出来るが、言語が統一されていない現状では、タイトル名によるソートができない。データに統一性を持たせないことには、基準にする項目が無く、ソートしようがないのである。ソートされていないという事は、全データ一覧から目的のデータを探そうとすると、どこにあるのかが全く分からないので、地道に最初から順番に見ていかななくてはならない。

文献のタイトルの言語の統一、ローマ字転写の統一、スペース部分の表記の統一はもう一度、データを一通り確認して修正していかななくてはならない。ローマ字はどちらの字体にするのか、スペース表記の仕方も

どちらにするのか、言語は何に統一するのか、そもそもチベット語の表記にしないのか、これらを決めないことには、データの修正作業に移る事もできない。

タイトルのつづりの間違いやタイトルの一部が注記に来てしまっているものも、一つ一つ確認して修正していかなければならない。そうしなければ、たとえ、言語や表記の仕方を統一したとしても、タイトル自体が間違っているのであれば、検索にかかりようがない。

著者の項目は全く入力されていないので、実際に文献を見たり、もともと大学の図書館にある蔵書検索システムで確認しなければならない。しかし、著者が不明の文献もあるので、これを完全にすることはできない。分類の項目は今回実装したものなので、これから順に入力していく必要がある。

上記のように、この問題の多いデータベースを修正するには、それぞれの文献についてのデータ作りからやり直し、間違った表記や、不統一な言語や表記方法を改善していかなければならない。タイトルの言語については、チベット語の元のタイトルを入れた項目を新たに増やす方法が考えられる。こうすれば、全ての文献にチベット語のタイトルがつくので、どの文献でもチベット語で検索できるようにすることができる。

しかし、チベット語の文献についてや、チベット語自体がわからないと、何が間違っているのか、どうすれば修正されるのかがわからない。これらの作業が困難で、図書館ではできないというのなら、チベット語の専門の先生達がやらなくてはならないのかもしれない。

(3) 終わりに

今回、大谷大学所蔵チベット語文献データベース検索システムを制作した事により、図書館にあるデータの問題が見つかった事で、今後どうすればよいかという事がある程度判明した。今回の制作をやらないでいたならば、データの問題はしばらく放置されていたかもしれないので、このデータベース制作は、意味のあるものだと思う。

データに問題があるため、検索システムはこのままでは利用する事ができないが、きちんとしたチベット語文献データベースを完成させるための進展になったのではないだろうか。今後、今回判明した問題を解決し、大谷大学のチベット語文献データベースが完成されることを期待したい。