

物産誌・地形図・データベースを用いた実習型授業の実践

—ジョイント・リサーチ「文化系データベースの構築と利用」の報告—

松森 智彦・福田 智子・矢野 環

1. はじめに

ジョイント・リサーチとは文化情報学部の3年次履修の必修科目である。これは、グループワークを中心とし、問題発見・解決能力を養い、発表・討論・レポート作成を実習して卒業研究の準備を行う科目である¹。2015年度に、新しく「文化系データベースの構築と利用—テキスト・画像・地理コンテンツ活用の実践—」というジョイント・リサーチが開講された。筆者らはその担当教員である。

この授業では人文学に軸足を置きながら、データベース、GIS、定量分析などを用いた資料研究を実践する。到達目標としては次の三点を掲げている。(1)活字化された歴史資料に親しみ、文化的・人文学的な考察力を身に付ける。(2)文化系データベースの作成、操作の基礎を学び、その能力を身に付ける。(3)研究対象の定量的分析、地理的分析を実践し、その能力を身に付ける。

コンピュータや地図を使った実習という事もあり、授業実施の準備に際しては、幾つかの技術的課題があった。その解決には、筆者らの議論と試行、また協力者からの助力が必要であった。本稿ではこの授業について、これら技術的課題の解決、授業構想、準備、気が付いた事など、授業実施に関わる記録として、報告を行う。

2. 授業の目的

諸論あると思うが、人文の文とは、文化や文学の文ではない。『日本大百科全書』の[人文地理学]の項には「人文は「にんもん」とも読まれ、古くから使われた天文(てんもん)、地文、水文と並ぶ人に関する学問の意味である。」とある²。古くは、人文学とは、人に関する学を指していた³。人に関する学問の多くは、過去に人間により作成された資料に基づいて行われてきた。そのため、人文学とは端的には「資料の学」と考えて良い。この授業では「資料の学」としての人文学のための研究技法、特にデータベースやGIS、定量分析など用いた文化と情報学による研究技法を学ぶ。なお、本授業の名称には「文化系」とある。初案は「人文系」であったのだが、先述の通り今日「人文」の語は多義的であり、分野名にもなっていて混乱を招く。一般に文化とは人間の持ち物であるため、文化でも良いであろうという結論となり、また本学部の学部名称との関連もあって、現行の名称となった。

ほか、本授業では「リアルな体験の重視」を授業構想の一つのテーマとした。分野を問わず、研究のリアリティのためには「本物に触れた体験」

¹ http://www.cis.doshisha.ac.jp/curriculum/pn_01.html より要約。

² 『日本国語大辞典』の[人文]の項には、「人に関する事柄。人間の社会。人事。また、人倫の秩序。じんもん。」とある。

³ 今日では人文学また人文科学の定義は多くあるが、学問の発展に伴い包含する分科は変化するので、これは自然なことである。しかし本授業では、人文学を「人に関する学」ひいては「資料の学」として扱った。この定義では大規模な社会調査や計画的な実験に基づく、仮説検証が可能な社会科学は、対象に含めない。

際には文字認識率を上げるために、字種を数字に限定し、整理作業を行った。字種の制限により、400dpiの画像でも十分な認識結果が得られ、担当教員のみでデータの作成を行うことが出来た。

この村落ごとに並び替えられた、整理済みの産物データを分割地域ごとに印刷し、学生に配付した。学生はこの表をもとに産物データの一つ一つを入力する。産物の種類は175品目である。デジタル化済みのデータがあるのだから、それを再度手入力するのは、一見無駄な作業のように思える。しかし、この再入力の作業を通し、学生は村落の名前、産物の名前、種類、前後の村落との産物構成の比較、極端に量の多い産物、少ない産物、単位の種類など、対象資料を研究する上で、最低限必要な知識を身に付けることができる。また、自分の手を動かし、苦勞して入力作業を行うことにより、入力経験とともに入力対象への愛着が湧くことが期待できる。研究対象への愛着は、学問を行う上での原動力として欠かせないものである。

小山ら1982には、産物名ごとに地図と生産量、概要をまとめた食品目録が所載されている。175品目について解説があり、一ページに対して一品目が割り当てられている。本授業では、この食品目録についても教材利用を行った。ページ枚数が多く、印刷配付することが難しかったため、目録ページのスキャン画像(PDF)をWebサーバーにアップロードし、授業中に学生が自由に閲覧できるように工夫した⁵(図4)。また『斐太後風土記』刊本については「近代デジタルライブラリー」より閲覧が可能であるが⁶、大正期発行の刊本についても教室に持ち込み、学生が自由に手に取って

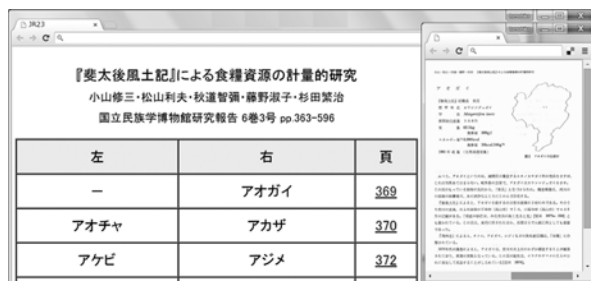


図4. 食品目録のPDF閲覧

読めるように配慮した。

4. システムの構成

本授業では、データベースの構築及び操作の実習を行う。その実習を実現するために、授業用のシステムの準備が必要である。ここでは、その準備の過程およびシステムの構成について報告する。なお、本システムのデータベースサーバーの利用において、同志社大学文化情報学部の波多野賢治氏、また深川大路氏に御助力頂いた。またネットワーク構成においては、同学情報企画課情報ネットワーク係に御協力頂いた。記して感謝する。

授業用システムの開発での要件は以下の4点である。(1)データベースは文化情報学部のサーバーに設置する。(2)学生は学部のコンピュータからサーバーに直接アクセスし、データを取り出す。(3)学生自身が入力データの登録を行う。(4)学生には挿入・更新・削除の命令は実行させない。これらの要件を満たすために、図5に示すようなシステムを設計・実装した。

要件(1)についてであるが、本授業はデータベースの構築と操作の実習を行う。授業で扱う村落・産物データの量は、表計算ソフトウェアで十分に扱いきれる量ではあるが、それでは授業の目的は達成できない。目的の達成のために、データベースソフトウェアを操作する必要がある。データベースのソフトウェアは多く存在し、無償のも

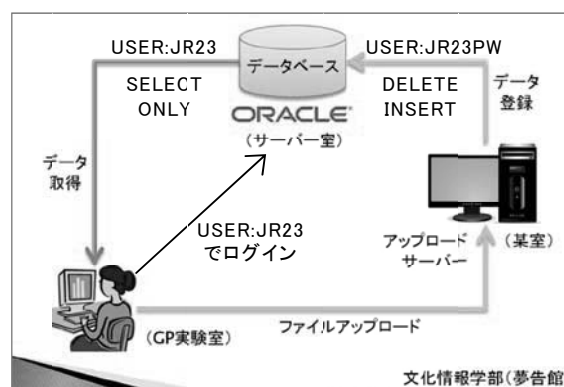


図5. 授業用システムの構成

⁵ Webサーバーは学部内に設置しており、学外からアクセスすることはできない。後述のアップロードサーバーと共用である。

⁶ <http://kindai.ndl.go.jp/info:ndljp/pid/952768>(上巻)
<http://kindai.ndl.go.jp/info:ndljp/pid/952769>(下巻)

の、コンパクトなものも多くある。しかし、本授業では文化情報学部設置されている商用のデータベースサーバーを利用することにした。商用のデータベースソフトウェアの操作経験を得ること、そしてサーバー・クライアント型システムの体験と理解を得ることが目的である。

要件(2)については、データ取り出し専用のデータベースユーザーを JR23 という名前で作成した。要件(4)により、学生からの挿入・更新・削除命令の実行を禁止しているため、このユーザーには CREATE SESSION 権限のみを持たせ、SELECT の他を実行不可とした。

要件(3)と要件(4)は一見矛盾しているが、学生が直接、挿入・更新・削除等のコマンドを実行できないよう制限すれば良い。学生には入力データを CSV ファイルで Web サーバーにアップロードさせて、Web サーバーが登録処理を行うようにシステムを構成した。この間接的な登録処理には、短い PHP のプログラムを用いて、規定の書式の CSV ファイルで、特定ユーザーの産物テーブルを上書きするように処理を組んだ。その際ユーザー ID とパスワードによって認証を行い、特定ユーザーの全産物データを削除、全挿入するトランザクション処理を行う。登録処理の際、データベースへの接続には、JR23PW というパワーユーザーを用い、そのユーザー領域に授業用システムのテーブルを保存した(図6)。

要件(2)を満たすために、JR23 ユーザーからの JR23PW の産物テーブルへのアクセスを許可した。そして、授業では JR23 ユーザーを用いたデータベースアクセスを行った。JR23 ユーザー内には、JR23PW 内のテーブル群と同名のシノニムを配置し、そのシノニムを通じて、JR23 ユーザーより JR23PW のテーブルを検索するように

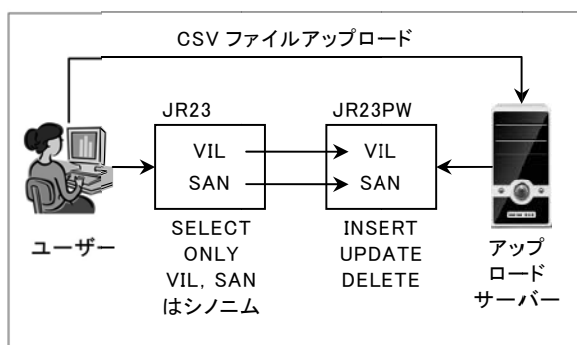


図6. データベースの構成

設定した。結果的に、学生からは JR23 ユーザー領域のテーブルを操作しているように体感されるが、SQL 文で指定している JR23 ユーザーのテーブル名は、実はシノニムであり、実際には JR23PW のテーブルを検索している。これは、耐攻撃のための簡単な仕組みであるが、このデータ登録のための Web サーバーは、授業時間の他はシャットダウンしてあり、攻撃を受けるリスクを小さくしている。

5. データの取り出し

授業ではデータの取り出しに SQL*Plus を用いた。これはデータベースソフトウェアに同梱してある、古くから使われているシンプルなアプリケーションである。データベースの業界でも、システムの開発や保守・メンテナンスの際に頻用される。キャラクタベースのユーザーインターフェースであり、一見扱い辛そうに見えるが、操作は単純で結果が明快である。授業では SQL*Plus のみを用い、他のクライアントアプリケーションは使わなかった。Excel や R 等にデータを移す際は、SQL*Plus の画面出力をコピーした、クリップボード経由でのデータ移送を行った。そのために、コマンドラインでの出力行数の設定、また SQL*Plus の列ヘッダ出力の削除など、諸設定を行っている⁷。

授業では、SQL 文のデータ操作言語 (DML) の中でも、データの取り出しを行う SELECT 文のみを学習する。授業の第1～4回までを、表データの入力及び SQL 文の学習に割り当てた。SQL 文は、基礎的な文法については極めて単純である。春学期では、SELECT 文の基本構文と WHERE 句、IN 句、ORDER BY 句、GROUP BY 句、COUNT・SUM・AVG・MEDIAN などの集計関数、そして副問い合わせまでを学習した(図7)。リレーショナル・データベースの重要項目であるテーブル結合は、ややイメージし難い部分がある

⁷ コマンドプロンプトの左上端アイコンをクリック→編集→選択より画面出力をコピーすることができる。出力行数の設定は、左上端アイコンより、プロパティを指定。SQL*Plusの列ヘッダは set pages 0の命令により非表示とすることができる。SQL*Plusより Excelへデータを貼り付ける際は、区切り文字にスペースを指定すれば良い。Rへの貼り付けは、SQL*Plusから、そのまま read.table("clipboard")より読み込むことができる。

ため、データベースの操作に慣れた秋学期での学習とした。

8. 集計関数: AVG, MEDIAN

▶ AVGは平均、MEDIANは中央値を取得。

```
select food, avg(value), median(value)
from san where id = 1
group by food;
```



FOOD	AVG(VALUE)	MEDIAN(VALUE)
アズキ	520	400
アユ	155	100
アウ	1332	1400
イノシシ	1.5	1
イワナ	442.857143	500
コサギ	2.33333333	3

図7. SQL文実行の例(第3回講義スライドより)

6. 旧版五万分一地形図の利用

本授業では、地理コンテンツの活用も授業の柱の一つとしている。地理コンテンツとは地理情報を含む資料を指しているが、本授業では特に「旧版五万分一地形図」を教材として利用した(図8)。これは、国土院の全身の一つである陸地測量部が作成した全国網羅の基本図であり、明治23年から大正5年にかけて全国の整備が行われている。飛騨地方全域の謄本は入手済みで、ロールスキャナで読み込んだデジタルデータも手元にあったのだが、本授業ではあえて紙地図での教材利用を行った。昨今デジタル地図の利用が普通であり、GPSと組み合わせて極めて利便性の良いもので

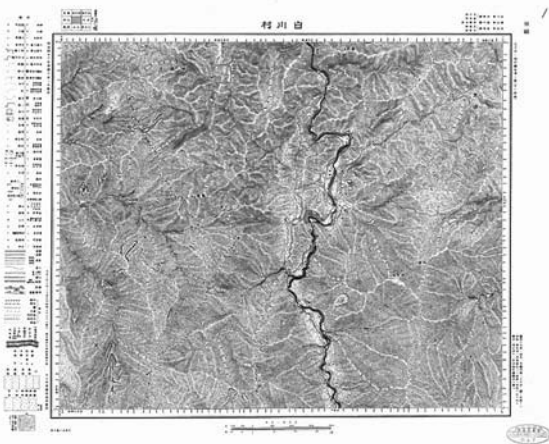


図8. 旧版五万分一地形図(白川村)

はあるのだが、地理学の基礎を学ぶには古い紙地図を自分の手で扱い、定規で測ったり、線を引くのが学習の方法として最良と思われる。謄本は枉版であるため、大型コピー機で謄本をA2用紙に等倍複写し利用した。授業の際は、学生の担当地域ごとに、対象村落をすべて含む紙地図の複写のセットを用意し、提供した。書き込みを許可し、地図に親しんでもらうことを再優先とした。

先述したが『斐太後風土記』には既にデータベースが存在し、村落の位置情報も網羅している。しかし、授業ではそれらの位置情報は用いず、旧版地形図より村落名を探し出し、定規で測って座標計算を行い、Excelに入力するという、位置入力の作業を実習した。手間のかかる作業のように思えるが、地形図と村落との関係を地図上で読み取り、理解するために、どうしてもこの作業が必要である。

旧版地形図には四隅に日本測地系での緯度経度が記載してあるので、これを元に図郭幅(度)を算出する。また定規を用いて図郭幅(cm)を測り、度の図郭幅とcmの図郭幅の比を取り、1cm当たりの度換算値を算出する。村落位置計測の作業は以下の手続きで行った。(1)村落を地図上で探し出す。(2)村落の位置を定規を用いて、図郭線からの距離(cm)で測る。(3)1cm当たりの度換算値を用いて、図郭線からの距離(cm)を度に換算する。(4)図郭線の座標値と(3)を合算し村落の座標を算出する。

計測した村落の座標は、日本測地系による緯度経度であるため、今日のGISソフトウェアで利用するためには、世界測地系に変換しておくのが良い。変換にはWeb版のTKY2JGDの一括変換機能を利用した⁸。変換後の座標を元にKMLファイルを作成し⁹、地理院地図¹⁰での閲覧を行った。変換後の緯度経度は、現行の地図とズレている事があったため、地理院地図の地点編集機能を用い、旧版地形図の地形、地物を見ながら村落位置の訂正を行った。これらの作業には授業の第5～8回を割り当てた。

⁸ <http://vldb.gsi.go.jp/sokuchi/surveycalc/ky2jgd/>

⁹ KMLファイルの作成は以下の手続きで行った。(1)ExcelファイルでKMLの雛形を作る。(2)座標を流し込む。オートフィル機能を用い、座標点をタグで挟む。(3)Excelからテキストをコピーし、テキストエディタに貼り付けて保存する。文字コードはUTF-8に指定する。

¹⁰ <http://maps.gsi.go.jp/>

なお、上記手続きにより村落の位置入力を行っていくと、最終的に出力された村落位置が、一様に300mほど西にズレる現象が見られた。授業中に、複数の学生においてこの現象が発現したため、調べてみると、図郭そのものが300mほどズレていることが分かった(図9)。この件について、国土院地理空間情報部に問い合わせたところ、大正7年9月以前発行の地図は経度10秒4のズレがあるため、これを加算して下さい、との回答を得た。日本測地系の測量原点には、最初の明治25(1892)年のものと、後の大正7(1918)年のものがあるとの事である。大正期の測量により、最初の明治期の測量原点に経度10秒405のズレがあることが分かったため、大正7年9月以前発行の図は、経度10秒4のズレを加算して計算する必要があった¹¹。本授業で使用した地図は明治期測量のものが多く、村落位置のズレが発現していた。なお、経度幅の距離は緯度により異なるが、飛騨では1秒およそ25mであるため、10秒4で260mほどのズレが出ていた。この計算手続きの誤りは、本授業において、日本の近代測量史の学びを深める良い機会・体験となった。

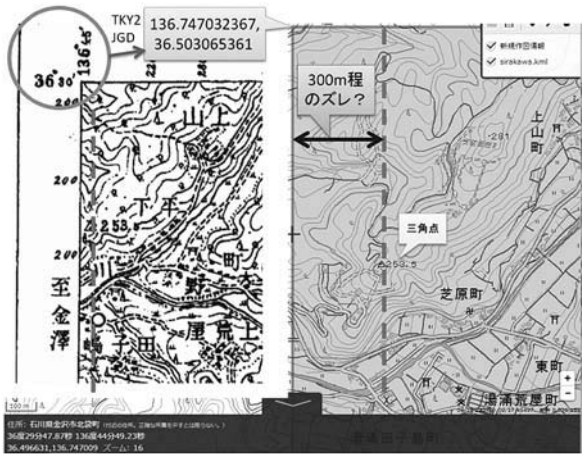


図9. 図郭のずれ(左:地形図、右:地理院地図)

¹¹ 明治25年設置の測量原点は東京天文台の子午環中心とされ、東経139°44'30".0970であった。大正7年には139°44'40".5020と改められ、10".405のズレが生まれている。

7. 定量分析

本授業では第10~12回を定量分析の学習に割り当てた。春学期では尺度水準の低い質的データ分析を2回行い、量的データの分析は1回のみとし、特に基本集計と単変量の分析に留めた。春学期ではデータ操作の基礎を学ぶために、単純で分かりやすい内容を実習し、多変量解析など知識や複雑な計算が必要な分析方法については、秋学期の項目とした。第10回の授業では、まずデータベースから取り出したリスト形式のデータより、Excelのピボットテーブルを用いて二値データを作成し、産物を記載頻度で並び替える作業を行った。自分が扱っている研究対象が、定量分析のためのデータとなる過程を体験し、情報を多く持つ属性とそうでない属性を認識、理解するためである。また、第11回では、二値データよりJaccard距離を算出し、距離行列を作成した。この距離行列をもとに階層的クラスター分析を行い、データに基づく分類を実習した(図10)。第12回では単変量の量的分析を行った。要約統計量、代表値の算出、一次元の散布図、ヒストグラム、箱ヒゲ図の5項目について、実習を行った。これらは基礎的で平易な内容であるが、取り扱うデータの理解のために、必ず調べる項目である。

春学期に実習した、SQL文によるデータベースからのデータ取り出しと、上記の基礎的な定量分析の方法を組み合わせれば、学生が各自でデータセットの作成と定量分析を行うことができ、自律的・自発的な研究活動を促すことができる。春学期ではそのための準備として、これらの基礎的なSQL文と定量分析を重点的に学習している。

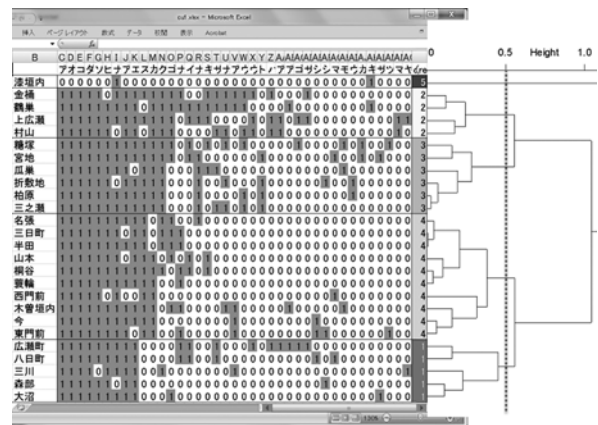


図10. 二値データによるクラスター分析

8. レポートの作成と口頭発表

本授業では第9回にレポートの作成方法について解説し、レポート提出をもって中間評価とした。文章の良し悪しについて、文の品質と文章の構成の二点から評価し、単文・短文での執筆を推奨した。事実の列挙に力点を置き、データに基づいたレポートの記述方法について解説した。

また第13回にはスライドの作成方法について解説し、14回には口頭発表の留意点について説明を行った。分かりやすいスライドの作成、階層的なスライド構成の設計、読み上げ原稿の作成と発表時間の調整、リハーサル的重要性、質疑応答への貢献など、発表に関わる基礎的な留意点について解説を行った。第15回は各地域ごとの発表会とし、プレゼンテーションと質疑応答を行って期末評価とした。

9. おわりに

本稿では、筆者らの担当したジョイント・リサーチについて、授業実践の構想、準備、技術的解決、気付きなど報告を行った。今後の大学教育では、コンピュータを利用した実習は増えて行くと考えられるため、本稿のように授業実施に当たっての準備、課題、解決などを報告し共有することは、益のある事だと思う。

春学期では、物産誌、地理学、情報学について一度に学ぶ事となり、学際研究の面白さと大変さを教員、学生ともに大いに味わうことが出来た。これは文化情報学部の特徴であると思うので、十分に苦労し、吸収し、学習に繋げていきたい。

参考文献

- 朝尾 紀幸 2010「観測機器が伝える歴史《5》—リーフラー時計—」『水路』第152号, pp.27-29
- 蘆田 伊人編 1915『大日本地誌大系 第七冊 斐太後風土記 上』大日本地誌大系刊行會
- 小山 修三・松山 利夫・秋道 智彌・藤野 淑子・杉田 繁治 1982「『斐太後風土記』による食糧資源の計量的研究」『国立民族学博物館研究報告』第6巻 第3号 pp.363-596
- 箱岩 英一 2004「日本の経度は金星日面経過観測から」『測量』2004. 4, pp.28-29
- 『日本国語大辞典』「人文地理学」の項
<http://japanknowledge.com/lib/display/?lid=1001000123067>
- 『日本大百科全書』「人文地理学」の項
<http://japanknowledge.com/lib/display/?lid=2002023c186fJa08b7re>