

シラバスデータベースシステムの実用化

— シラバスと履修科目表を中心とするデータベースの構築方法に関する考察 —

井田 正明[†] 芳鐘 冬樹[†] 野澤 孝之[†] 宮崎 和光[†] 喜多 一[†]
 大学評価・学位授与機構[†] 京都大学 学術情報メディアセンター[†]

1. はじめに

近年における情報ネットワークの整備とともに高等教育においても情報技術の活用が進展しつつある。これまでに我々は高等教育における教育課程の分析のため、電子化されたシラバスや履修科目表など教育課程に関する情報の収集を進めてきた。また収集情報の構造を解析することにより教育課程に関する情報の一般的なデータ構造を検討し、データベースをリレーショナルデータベース上で試作してきた。また、データベースの利用による教育課程の分析に関しても検討を行ってきた（以上、参考文献[1]-[3]）。本稿ではシラバスと履修科目表を中心とするデータベースの構築に向けて、データベース化の際の問題点について報告する。

2. 教育課程に関する情報の階層表現

教育課程の編成方針やシラバスの内容の概略についての記述はあるが（学校教育法、各種答申など）、その詳細に関しては各大学ごとに設計されている。本稿ではこれらをふまえ以下の考察を行う。教育課程に関する情報の階層構造をつぎのように表現する。

機関(大学) - 学部 - 学科 - 年度 - 教育課程 (プログラム) - 授業科目

これまでにこの情報構造の一般的な表現を検討してきた[1]。ここでは再検討を行った教育課程全体のデータ構造の概要を Table 1 に示す。これは XML Schema により表現され、データの入力の際の検証や階層構造を利用したデータベースの利用の際に用いられる。

Table 1 XML Schema on curriculum

institution	大学 (機関)
institutionID	大学コード
└faculty	学部 (研究科)
└facultyID	学部コード
└department	学科
└departmentID	学科コード
└year	年度
└program	(教育プログラム)
└programID	プログラムコード
└programName	プログラム
└purpose	目的
└characteristics	特徴
└termSystem	学期制 (2 学期制など)
└creditsForCompletion	修了要件単位数

└cRemarks	修了要件に関する備考 (授業科目リスト)
└courses	(授業科目)
└course	授業科目コード
└courseID	科目名
└title	英文科目名
└eTitle	達成される目標
└courseObjectives	授業概要
└abstract	キーワード
└keywords	使用言語
└language	開講学期
└term	曜日
└day	時限
└time	教室
└room	授業形式 (講義など)
└classType	(教員リスト)
└lecturers	(教員)
└lecturer	教員コード
└lecturerID	教員名
└name	教官連絡先電話
└tel	教員電子メール
└e-mail	教員 WebSite
└webSite	教員研究室
└office	オフィスアワー
└officeHour	(授業計画)
└plan	(各回授業計画)
└session	各回予習
└preparation	各回授業内容
└topics	各回宿題
└assignment	予備知識
└prerequisiteCompetences	事前に履修しておく科目
└prerequisiteCourses	成績評価方法
└evaluation	取得単位数
└credit	教科書
└textbooks	参考書
└references	備考
└remarks	

データベースの元となる資料は、Web、CD などより収集したシラバスや履修科目表などである（過去数年間で収集）。Table 1 のデータ構造をリレーショナルデータベースに格納するため、項目を分割して複数のテーブルを設計している：program テーブル（教育課程に関する項目）、linkage テーブル（教育課程と授業科目の関係）、course テーブル（各授業科目）。教育課程と授業科目は一般に多対多の関係にあるため、linkage テーブルによりそれらに関係付けている（授業科目と担当教員にも同様の関係あり。検討中）。これら以外に program や faculty に関するコード表のテーブルを設けている。また、実態として同じ授業科目が読み替えて別の名称を持っている場合などはその対応付けが必要。全学共通科目に関しては、大学

の直下の階層に対応付けるなどの対応が必要。

3. 既存の電子資料からのデータ移行

現時点でのデータベース整備状況として、入力済み件数は、教育プログラム数：約 100、授業科目数：約 5000、linkage テーブルのデータ数：約 5000 である。以下では上述の 3 つのテーブルについてそれらのデータベース化の際の問題点を述べる。

(1) program テーブル：

- ・ institutionID, facultyID, departmentID はコード表と照らし合わせながら人手で付与（JSPS のコードを援用）。institutionID と facultyID は一意に決まるので、自動変換スクリプトの作成が可能。departmentID は、適切なコードがない。
- ・ creditsForCompletion (cRemarks) は、教育プログラムごとにさまざまな条件が付加されており、一般的に表現することは困難。文書データとして対応すべき。

(2) linkage テーブル

- ・ linkage の作成には、当該教育プログラムの全科目が 1 ページにまとまっていないケースが多く、linkage の作成に多くの時間を要する（学部・学科・コースのページをすべて見ないといけないものや、基礎/専攻などの種別で履修科目表が複数ページに分割されている大学もある）。

(3) course テーブル

- ・ 収集したシラバスのデータから情報を抽出し course テーブルに格納する際には、（半）自動抽出システムを用いている [2]。このシステムは主にデータが [項目名]+[項目値] の繰り返しからなるテーブル構造であることを想定しているが、その利用において以下のような事項が問題点となる。
- ・ 項目名が省かれているデータ：科目名 (title, eTitle) などそのようなケースが多い（科目名については、項目名が見つからないときは、予め定義しておいた項目値のパターン（「～演習」など）にマッチングする回数が多い出現位置（行）を探すことで対処しているが、出現位置が一定でない場合（ヘッダの情報量が統一されていないなど）は、出現位置の特定に失敗し、誤って抽出してしまう。

- ・ 項目名+項目値の 1 対 1 のセットになっていないテーブル構造：「項目名 1, 項目名 2, 項目値 1, 項目値 2」のように、対応する項目名と項目値が連続していない場合に対応する必要がある。

- ・ 繰り返しありの（項目値が複数ありうる）項目：同じ学科・教育プログラムでも、デリミタが統一されていないケースがあり、分割してデータベースに格納することが難しい (lecturer 関係, prerequisiteCourses, textbooks, references, remarks, session 関係, keywords がこれに該当する)。現在は、できる範囲で、手作業で後処理を

加えている。

- ・ 判別が難しい項目：abstract, prerequisiteCompetences, courseObjectives, evaluation, remarks, session 関係などは、もとのデータの項目名の表現が多様であり、また他の項目と分けずに記述されている場合もあり、正確な抽出が難しい。

- ・ 元データにない項目：language など。

- ・ ID の付与：ID 類も基本的には元データにないためこちらで生成する必要がある。科目コードがないデータが多いので、courseID の採番ではファイル名の使用に統一している。prerequisiteCourseID は、後処理のスクリプトで、prerequisiteCourses と title のマッチングに基づいて自動的に付与している。前述のとおり prerequisiteCourses に複数の科目名が分割されずに格納されている問題があったり、表記のゆれの問題があったりするため、現在の付与はかなり不完全である。ID の付与は、現在、エディタや Excel 上での置換とスクリプトの使用を合わせて行っているため作業が煩雑である。

- ・ データベースの定義に合わせて変換する必要がある項目：credit は、後処理で float 型になるよう規格化する必要がある。全角数字、語尾（～単位）、イレギュラーな表現（「2 または 1 単位」など）に対応する必要がある。現在は手作業で変換している。

(4) 全般について

各々の処理はそれほど大きい作業量ではないが、ひとつひとつ状況に応じて判断しながら処理しており、また複数のツールを切り替えながら使用しているため、自動化が難しく作業が煩雑になっている。また、一連の処理（自動抽出した course データの修正、program データの入力、linkage データの作成）を、教育プログラムの数だけ繰り返さなければいけないため、大学数や学部数が増えると、かなりの作業量になる。可能な部分については自動抽出システムへの統合による合理化が望ましい。

参考文献

- [1] M. Ida, T. Nozawa, F. Yoshikane, K. Miyazaki, H. Kita: Syllabus database and web service on higher education, ICACT2005 (2005)
- [2] 渡辺将尚, 絹川博之, 井田正明, 芳鐘冬樹, 野澤孝之, 喜多一: Web 上のシラバス情報の収集と XML 変換, FIT2004 (2004)
- [3] 野澤孝之, 井田正明, 芳鐘冬樹, 宮崎和光, 喜多一: シラバスの文書クラスタリングに基づカリキュラム分析システムの構築, 情報処理学会論文誌, Vol. 46, No. 1 (2005)