

データ従事者の育成を目的としたデータ・リテラシー教育の試み

中鉢直宏^{†1}

概要：現在、ビックデータやビジネス・インテリジェンスなどの取り組みにより積極的なデータ利用が行われる様になってきた。データ従事者の日々の業務で入力されるデータの質が重要になってきている。そこで、本論文では、大学の一般情報教育において、データベースを学習するにあたって、情報システムにおけるデータベースの役割を理解するとともに、データ従事者として、彼らが入力するデータへの質について理解を促す学習方法を試みた。

Practice of data literacy education for development of data workers

NAOHIRO CHUBACHI^{†1}

1. はじめに

現在、ビックデータやビジネス・インテリジェンス (BI) などのデータを活用した予測や意思決定などが注目されている。例えば、教育の分野で、大学をはじめとする教育機関では IR (Institutional Research) と呼ばれる調査活動が注目され、教育機関で BI が実施されている。IR を促進するためにデータ従事者の育成が必要とされている[1]。

BI で分析されるデータは、多くは業務で使用する目的のものである。それらは業務システムに蓄積されているデータや Excel のデータである。しかし、それらのデータを組み合わせ分析しようとする、そのデータ自体分析に向かない場合がある。その前にデータの加工や名寄せという作業が発生する。実際分析にするまでの労力が問題となる。そのため、BI をするためには、極力加工や名寄せが発生しないようにデータの質の向上を目指し、データ作成、データ管理をする必要がある。

データの質が低い例として、例えば、大学などのシラバスなどのデータにおいて、分析に使用したいデータが、直接出版社へ提出してできた出版形式のものだったり、Web 化はされているが一括でデータを取得することは想定されていないシステムであったり、これらのデータを分析用にデータを加工するためにはかなりの労力が必要になる。

一般的に、業務データは帳票や個々の案件に対応するためにデータが使用されているので、これらのデジタル・データを利用している人であるデータ従事者は、目的外利用に使用できる組織の資産として考えることができていない。その問題を解決するために、本研究ではデータを業務とし

て扱う役割を担う人であるデータ従事者を BI などのデータが新しい価値を生む時代に合った形の人材育成方法を提案し、実際の大学の講義のデータベース教育の一環として試みた。

2. データ従事者の問題点

BI などの分析に使用されるデータは、業務システムやファイルサーバに蓄積される。それらは、自動的に蓄積されているデータもあれば、従業員が入力するデータもある。そして、業務に必要なデータは日々変化する。それらのニーズにこたえるために、システムの改修やマニュアルの改訂が行われる。しかし、正式に対応が行われなくて、現場最適で、新たなデータ入力のルールが決定される場合がある。これらのデータは日々の業務を遂行することが目的なので現場で問題が起きなければ、そのままシステムが使用される場合がある。その場合、データ入力の際にほかのデータとの整合性を意識した形で十分考慮してデータに関する作業が行われる必要がある。例えば、新しい商品に、勝手に新しい ID 付与する際に、対比表や付与するためのルールを明確化にして残しておく必要がある。しかしそのようなことを意識してデータ作業を行っているかという疑問である。

例えば、Excel などでレポートを作成した時点で、加工して、加工に使用したデータや元データが残されていないなどの問題がおこる。特に本来データを扱うはずの Excel を帳票やレポートを作成するツールとして使用し、データを二次利用できない形で残される場合がある。これは、ネ申す Excel 問題として、一般的に多く存在する問として指摘

^{†1} 帝京大学 高等教育開発センター
Center for Teaching and Learning, Teikyo University.

されている。そこでは、「高校」の「教科情報」や大学の一般情報教育では、たとえ Excel を扱う際に、データ・リテラシー教育の大切さ指摘されている。また「古くなったデータは不要なので削除するという行動様式も見られた。これはデータ文化にも関連するが「アーカイブズ」の文化が欠けているのが原因と思われる。」[2]

その一方で、データ・サイエンティスト育成などにみられる、高度なデータ分析などの人材育成が求められている[3]とあり、データに関する教育は必要不可欠と考える。

3. 本研究の目的

データ従事者は実際業務でデータを扱う人である。図 1 のように対象者は、直接データベースを設計したり、システムを構築したりする人数より多い。本研究は、データ従事者への教育が重要だと考え、データベースのためのデータ作成を通して、一般的なデータ・リテラシー教育を目指した。特に、一般情報教育におけるデータ従事者としてのアプローチという視点を重視した。データベース教育の試みは、sAccess などを利用した教育手法が提案されて、効果を上げている[4]。本研究では、データベースへ入力するデータの質に対して意識させるデータ・リテラシー教育を目的としている。データを入力時の配慮やそのデータがほかのデータと連携する可能性に考慮し、配慮できる能力を育成するアプローチを開発したいと考えている。

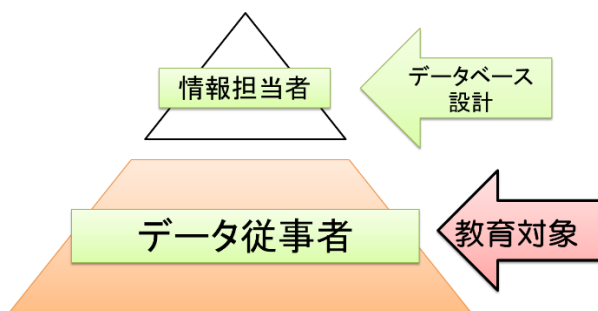


図 1. データ・リテラシー教育の対象者

4. 授業方法

4.1 授業設計の方針

データベース教育では予めデータを用意する方法が多く用いられる。与えられたデータを正規化、リレーショナル・データベースに読み込ませるような教育に重点が行われている。しかし、本講義の目的は、自分たちが作ったデータが再利用されることの意識を持ってもらうことである。

そこで、本授業では、データベース教育において、「料理データベース」を作成する目的で、受講生がデータ作成から学習する試み[5]を参考にした。この試みでは、料理データ

を各自持ち寄り、グループごとまとめることにより、データ表記の気づきが学習者にみられたという知見が報告されている。このことは、本研究で目指すデータ従事者に必要な気づきであり、この気づきを計画的に誘発したいと考えた。料理データベースを持ち寄るというアプローチは、学習者に関心を高める反面、データベース用に料理メニューのデータに条件を設ける工夫が必要となる。

そこで、本授業では、各自のデータを持ち寄ってデータ表記の揺れに気づかせること、データの質について理解すること目的とした。演習は、業務を想定してデータベース作成のシナリオにした。演習の目標は、データを使ったレポートを作成することとした。次のことを受講生には体験してもらうことにした。データ入力は全員が同じ業務としてデータ入力を体験する。他の受講生が入力したデータを一つにまとめることによって複数人数で行うデータ入力作業を体験する。与えられたデータ位階の想定外のデータ作業を体験させる。データ部分とレポート部分を分けることによる利点を考えさせる。入力されたデータに対し、ほかのデータと結合することを体験させる。これらの体験を通してデータ従事者としてどのようなことを配慮する点を学習してもらう。

4.2 授業概要

本研究の試みは、帝京大学の一般教養の講義である「情報処理Ⅲ」で実施された。本授業の目的は Access を使ったデータベース習得として開講されている。授業は 2 クラスで運用されている。授業の大まかな流れを下記に示す。授業は、本稿は、授業実施途中なので前半部分のみ示す

1. 店員として伝票（図 1）を配布し記入させる。
2. 伝票をシャッフルし、再配布する。
3. 店長として本部に紙のレポート（手書き）を作成させる。
4. Excel を使って伝票をまとめさせる。
5. 追加の伝票をわたす。
6. 紙のレポートと Excel のデータにそのデータを追加させる。
7. うまく追加できたか、感想を聞く
8. 値段用のシート（図 3）を配布して、Excel データに反映させる
9. データとレポートを分ける必要性を考えさせる
10. 非正規型のデータを作成させる
11. 全受講生のデータをまとめる。
12. 10 のデータをデータベースに読み込ませる
13. データベースでソートや検索を行う
14. 値段テーブルを配布し、結合させた、
15. うまく結合できたか確認する。
16. 必要があれば名寄せ作業を行う
17. 様々なクエリを試す。

今後は、様々なデータベースの操作、フォームの作成、レポート作成を実施予定。

担当者	<input type="text"/>	注文日	<input type="text"/>
定食屋「帝京」メニュー		伝票番号	<input type="text"/>
ラーメン		焼肉定食	
味噌ラーメン		焼き魚定食	
塩ラーメン		フライ定食	
とんかつラーメン		とんかつ定食	
かつ丼		納豆定食	
天丼		麻婆豆腐定食	
うな丼		生姜焼き定食	

図 2. 伝票シート

ラーメン	¥600	焼肉定食	¥800
味噌ラーメン	¥700	焼き魚定食	¥650
塩ラーメン	¥700	フライ定食	¥950
とんかつラーメン	¥800	とんかつ定食	¥1,000
かつ丼	¥750	納豆定食	¥500
天丼	¥900	麻婆豆腐定食	¥550
うな丼	¥1,200	生姜焼き定食	¥850
玉子丼	¥600	天ぷら定食	¥1,200
親子丼	¥700	ひじき	¥100

図 3. 値段シート

4.3 授業手順

まず、最初に各自 10 枚ずつ伝票シートを配布した。これは、食堂やレストランでみられる伝票を意識したものである。ほとんどの学生が一度は目にしたことがあるであろうことから採用した。まず、店員が 1 名から 4 名の注文を 10 件受けたと想定して、自由に伝票に記入させた。いったん伝票を回収し、伝票を混ぜ再配布する。そして、まず A4 用紙を 1 枚配布し、報告書（レポート）としてまとめてもらった。このことにより、業務の目的はレポート作成ということを意識してもらった。次にデータを Excel にまとめるように指示した。Excel では、必ずすべてのデータを Excel に反映させることを指示した。まとめる形式はここでも一切指定していない。次に処理する伝票が後から見つかったというシナリオとして一枚伝票を配布し、手書きのレポートと Excel に反映させるように指示した。これは、レポートという形はデータ操作ができないことや自分のまとめた Excel の形式がデータ追加をできる形で作成されていたかを考えさせる目的に行った。次に値段シートを配布し、Excel データに追加することをほかのデータと連携することを体験させ、このことにより、自分の作ったデータがほかのデータと結びつくことを理解してもらうことを目指した。次に Excel で、データシートとレポートを分けるように指示した。また、非正規形のデータを説明し、データを

まとめるように指示した。提出された非正規形のデータは、すべてまとめて再配布した。次にデータベースに読み込ませ、検索、ソートを実施させた。次に別のテーブルを配布し、データを結合させた。結合できない箇所の発見を指示し名寄せ作業を体験させた。本格的なデータベースの実習としてクエリなどを学習させた。

5. 授業の実施結果

まず、伝票を配って記入させると、名前の記入法方法、伝票番号の記入漏れ、現実ではありえない注文数の記入など様々な不備を用意する伝票がみられた。ここではあえて訂正の指示はしなかった。各自が伝票記入方式がことなることは、各店舗の方針が異なるなど、現実でも起こり得ることなので、あえて訂正せずそのままにした。実際他人の記入した伝票が配布されると、自分や他人の伝票の違いが起こることに気づいた生徒もいた。次に、紙のレポートを作成するにあたって、クロス表で集計したり、伝票を日付順にまとめたりと、伝票データをそのまま転記する例はあまり見られなかった。Excel にデータをまとめさせると、縦持ちの非正規型でまとめた例は 1 名しかおらず、多くはクロス集計やグラフなどを使ってまとめることが多かった。ただ、Excel 操作が得意な生徒は、グラフでまとめたり、合計を各所に入れたりするなど、元データが取り出しにくい形でまとめる例がみられた。そのため、あとから伝票を追加する演習に関しては、Excel でまとめる度合いが高いほど、データ追加が難しいことを体験させることができた。また、値段シートの追加は、vlookup などの Excel の関数を知らない限り、効率的にデータを追加することは難しい。また、データ部分が切り出されていないと vlookup すら厳しいことに気づいてもらった。ただ、特別に指示せずに、データ追加が容易な非正規形のデータを作成できる受講生は 1 名のみであった。伝票をまとめる制限として、列（フィールド）の項目を講師側から指定し、どのようにまとめるかを演習してもらった。今回は、伝票番号、注文日、記入者、メニュー名、個数という列のみでまとめることを指示した。この列でまとめると、非正規型のデータを作成することになる。ただ、その指示だけではデータ作成が困難な受講生には、1 枚の伝票をどのようにデータ入力する方法を例として提示した。ほとんどの受講生が非正規型のデータに作成することができた。講師は、これらの全学生の非正規型の伝票データを一枚のシートにまとめた。次に Excel のフィルター機能を使った並び替えと検索をさせた。フィルター機能を使うと、データが統一されて入力されていない部分が受講生は見つけることができる。この演習により、従業員名の表記の揺れ、メニューデータの表記の揺れ、伝票番号などの空白データの取り扱いについて考えさせることができた。Access にてテーブル作成の演

習後、まとめた Excel をインポートさせた。次にクエリにて検索・ソートの演習後、値段シートのデジタルデータを受講生に配布後、クエリのテーブルの結合機能を説明した。ほとんどの学生がクエリの操作に戸惑うところは見られなかった。しかし、結合結果の表示についての説明は必要だった。データ結合のうまくいかなかった現象を明示するため、学生が入力した伝票データを全レコード表示させる条件で値段シートのデータを結合させた。結合クエリの結果例を図4に示す。値段シートと異なる部分には値段が反映されていないことが明示できた。このことにより、入力されてデータがほかのデータと結びつくことを体験させるとともに、実際に入力されたデータの正確さがクエリの結果がうまく反映されるかどうかを学習させることができた。間違っって入力された名寄せという作業としてデータの訂正作業を体験させた。

データを結合させた演習後のアンケートでは、「再びデータを打ち間違えることのないようにする。確認が取れるのであればあいまいな名称はそれであっているか確認する。」「独自の名称にせず、誰でも分かるようにする点と正式名称で入力しなければならない」などの回答があり、多くはデータ入力への配慮の必要性について内容が多くみられた。

担当者	メニュー	値段
人見 勇太	納豆定食	¥500
桑原	とんかつ定食	¥1,000
桑原	生姜焼き定食	¥950
クロダ	みそ(ラ)	
クロダ	塩(ラ)	
クロダ	とん(ラ)	
クロダ	とん(定)	
宮尾	うな井	¥1,200
宮尾	焼肉定食	¥800
木村	カツ丼	¥750
木村	フライ定食	¥950
木村	とんかつ定食	¥1,000
安藤	生姜焼き定食	¥850
安藤	ラーメン	¥600
桑原	ラーメン	¥600
安藤	味噌ラーメン	¥700
クロダ	天丼	¥900
クロダ	納豆(定)	
クロダ	麻婆豆(定)	
宮尾	塩ラーメン	¥700
木村	ラーメン	¥600
木村	フライ定食	¥950

図4. 結合できなかったクエリ結果

6. 本授業の課題

一連の演習を通して、データベースでデータを読み込ませるまで、想定したより時間がかかった。特に紙のレポートを作成させる演習に関しては、思ったより多くの時間を要してしまった。また、多くデータを入力させるので、Excelなどの能力やキーボード入力の差が、演習の進度の差を生んでしまった。大規模の演習をこの形式で実施する際は、演習の進度の差の解消が必要になる。

7. まとめと今後の発展

受講生がデータ従事者としてどのようなことを配慮すべきかを演習で体験させることができた。演習を通してデータ入力に関する配慮の重要性を考えさせることができた。

今後は、クエリ操作により、元データがデータベースとして存在することによって、どのような形でも加工しデータ出力ができることを、理解させる演習を実施する予定になっている。このことにより、データベースとレポートを分けてデータを扱うことを学習する予定である。次に、データベース用にデータフォームを設計させる。このことにより、業務システムについて学習させる。このことにより、データ従事者として、入力と出力をデータベース中心としてシステムへの理解を促したいと考えている。

今後は、よりデータの質の向上につながるデータ従事者を育成することを演習など実施したいと考えている。

参考文献

- 1) 中鉢直宏, 教務データに基づく教学 IR の新展開: IR 支援システムによるマクロ型 IR からマイクロ型 IR の実践を目指して, 大学マネジメント, 国立大学マネジメント研究会, vol.11, 8-14 (2016)
- 2) 奥村 晴彦, 「ネ申Excel」問題, 情報教育シンポジウム2013論文集, pp93-98, 2013(2)
- 3) 桑田 修平, 中川 慶一郎 ビッグデータ時代のビジネス・インテリジェンス: 次世代ビジネス・インテリジェンス. 情報処理学会デジタルプラクティス / デジタルプラクティス編集委員会編, pp 38-43 (2013)
- 4) 兼宗 進, 白井 詩沙香, 竹中 一平, 長瀧 寛之, 島袋 舞子, 田邊 則彦, データベースを利用した高校におけるプログラミング授業の実践と評価, 研究報告コンピュータと教育 (CE), 2016-CE-134(22), 1-9 (2016)
- 5) 野部 緑, 料理メニューの作成から学ぶデータベース, 研究報告コンピュータと教育 (CE), 2016-CE-134(17), 1-4 (2016-02-27)