

日経 NEEDS 財務データにもとづく データベースサーバの構築

地 道 正 行

I はじめに

一般の科学の分野にとどまらず，社会科学系の学問分野においても，「安価」で「効率的」かつ「短時間」に「大量」のデータを入手する方法を検討することは，現在でも重要なテーマであろう．インターネット等のネットワーク環境が発達した現在において，金融関連の情報提供として米国における“Yahoo! Finance”¹⁾ や日本における「Yahoo! ファイナンス」²⁾ によるサービスが行われており，株価等のデータの入手は容易となりつつあるけれども，企業の財務情報のサービスとしては，過去から現時点までの時間的推移と各種の産業の横断的なもののバルクを対象に抽出できるフリーの環境は整備が始まった段階にあるといえよう³⁾．

本稿で扱うテーマは，比較的低いコストで，ある程度大きな規模の財務データを効率的に短時間で抽出できる学内向けのシステムを構築することである．実際には「日経 NEEDS 財務データ（一般事業会社）」の MT⁴⁾ 版をもとにデータベースを構築することによって，企業の財務データを効率的に取得

1) <http://finance.yahoo.com/>

2) <http://finance.yahoo.co.jp/>

3) 2008年に本格的にサービスが開始された金融庁による EDINET (<http://info.edinet-fsa.go.jp/>) 上の XBRL を利用した財務データの提供等があり，今後サービスの充実が期待される．

4) Magnetic Tape の略．

するためのシステム構築について述べる⁵⁾。

本稿は、以下のような構成によって成り立っている。まず、Ⅱ節ではサーバ環境の構築に関して述べる。ここでは、日経 NEEDS 財務データの MT 版にもとづくデータベースサーバの構築を扱い、特に Mac OS X Server, Linux, Windows といった各種のオペレーティング・システム (Operating System: OS) におけるデータベース管理システム MySQL⁶⁾ を用いたデータベースの構築に関して詳しく述べる。次に、Ⅲ節において Web サーバ Apache⁷⁾ と MySQL を PHP⁸⁾ を用いて連携させることによって Web ブラウザを使ってデータ抽出を可能とするシステムの構築について述べるとともに、Ⅳ節において、このデータ抽出サービスを受けるためのクライアントの設定に関して述べる。特に、データ解析環境 R⁹⁾ からデータベースに ODBC¹⁰⁾ を使ってアクセスするための環境設定に関して述べる。さらに、Ⅳ節において、データ抽出システムの実際の利用法に関して、具体例を与えながら述べることとする。その際、MT 版のデータベースから Web ブラウザを利用して抽出する例とともに、ODBC を利用することによってデータベースから R ヘダイレクトにデータを抽出する例も与え、データ解析ソフトとデータベースとの間をシームレスに利用する方法について言及する。最後に、Ⅵ節においてまとめと今後の展望について述べる。なお、付録において、LAMP¹¹⁾ 環境の構築法、ODBC 環境の設定法、日経 NEEDS 財務データ MT 版の解説とそれにもとづ

5) 本稿を執筆するにあたり豊原 (2004, 2009) が非常に参考になった。

6) <http://www.mysql.com/>

7) <http://www.apache.org/>

8) 正式名称は PHP: Hypertext Preprocessor であり、これを再帰的に略したものである。オープンソースの汎用スクリプト言語であり、HTML (Hyper Text Markup Language) に埋め込むことが可能である。なお、PHP という略語そのものは Personal Home Page Tools に由来する。詳細は、PHP マニュアル (<http://www.php.net/manual/ja/index.php>) を参照されたい。

9) The R Project (www.r-project.org) によって開発・配布が行われているオープンソースのデータ解析環境。

10) Open DataBase Connectivity の略。一般のアプリケーションからデータベースにネットワークを経由してアクセスするための標準的なインターフェース規格。

11) Linux, Apache, MySQL, PHP の略。

いて構築されたデータベースの構造や構築に利用される各種のスキプトのソースコードのリスト等が与えられる。

II サーバ構築

この節では、日経 NEEDS 財務データ MT 版にもとづくデータベースサーバの構築に関して述べる。今回構築した実際のデータベースは、連結本決算と単独本決算に関する両方が存在するが、ここでは連結本決算にもとづくもののみについて述べる。プラットフォームとしては、Mac OS X Server (10.5)、Linux (Ubuntu¹²⁾ 8.04 LST)、Windows Vista 上に LAMP 環境が設定済みであることを前提としていることに注意しよう。なお、LAMP 環境の構築に関しては付録 A を、日経 NEEDS 財務データ MT 版については付録 D を参照されたい。また、構築されたデータベースの構造については付録 G を参照されたい。

まず、データベースの設計に関して以下のことに注意する。

連結本決算のデータベースの設計項目

(設計項目 1) 連結本決算データ (日経によるコード: HIFSCA052) に対するデータベース (hifsc052) の構造として、レコード種別 FA01, ..., FG01 に対応するテーブル fa01, ..., fg01 を持つものを作成する。

(設計項目 2) 社名から日経会社コードなどを検索可能にするために社名リストのテーブル firmlist を作成する。

これらの点に留意して、プラットフォームに共通したデータベース構築の手順を以下に与える。特に、(設計項目 2) の必要性として、データベース hifsc052 を実際に利用する際、日経会社コードでデータを抽出することが考えられるけれども、社名等のキーワードから日経会社コードを検索する機能がなければ、そのことを実現するために膨大な社名リストの一覧を手元に持つ必要があり、利便性が損なわれるという理由による。なお、各手順は、

12) Ubuntu コミュニティ (<http://www.ubuntu.com/>) によって開発されている Linux ベースのオペレーティングシステム。

Unix のコマンドやシェルスクリプトを実行することによって実現され、そのためにはターミナル (terminal), コンソール (console) といったシェル (shell) を実行するためのアプリケーション上で作業が行われる。また, sed, Perl 等のインタプリタ¹³⁾, フィルタプログラム nkf¹⁴⁾ 等が適宜利用される¹⁵⁾。なお, 各手順で利用されるスクリプトファイルは以下のようなものである。

ファイル名	説明
script.sh	データファイルの前処理を行うシェルスクリプトからなるファイル
separate.pl	ファイルにおける適切な箇所にフィールドを区切るセパレータを挿入するための Perl スクリプトファイル
loadtoMySQL.sql	データベースとテーブルを作成し, script.sh によって作成されたテーブル単位のデータファイルを MySQL にロードするための SQL スクリプトファイル
loadfirmlisttoMySQL.sql	社名リストファイル firmlist.txt を MySQL へロードするための SQL スクリプトファイル

なお, これらのファイルのソースコードについては, 付録Eを参照されたい。
(手順1) ファイルの前処理

シェルスクリプト・ファイル script.sh と圧縮された連結本決算のデータファイル

```
HIFSCA052.1970.gz
HIFSCA052.1980.gz
HIFSCA052.1990.gz
HIFSCA052.2000.gz
```

を適当なディレクトリに保存後, ターミナルを起動し, そのディレクトリに移動した後, 以下のように script.sh を実行する。(なお, \$ はシェ

13) sed, Perl 等に関する詳細に関しては, ドウラティ&ロビンス (1997), ウォールら (1997) 等を参照されたい。

14) Network Kanji Filter の略。文字コードの変換等を行うプログラム。

15) Mac OS X, Linux には標準的にこれらの環境は用意されているので問題ないが, Windows にはデフォルトでこれらの環境は用意されていないので, Cygwin (<http://www.cygwin.com/>) 等の環境をあらかじめインストールしておく必要がある。

ルプロンプトである.)

```
$ /bin/sh script.sh
```

このスクリプトは、以下の操作を行う。

- (a) unzip コマンドでデータファイルを展開した後、リダイレクション (>, >>) を使ってこれらを1つのファイル `renketsu.all` にまとめる。
- (b) `grep`, `sed`, `Perl` 等を使って、このファイルから `FA01, ..., FG01` という文字列を含む行を個別に抜き出す。
- (c) `Perl` スクリプトファイル `separate.pl` を読み込んで、各ファイルのフィールド間に `tab ('\\t')` をセパレータとして挿入したファイル `FAtab.data, ..., FGtab.data` を作成する。(図1も参照。)

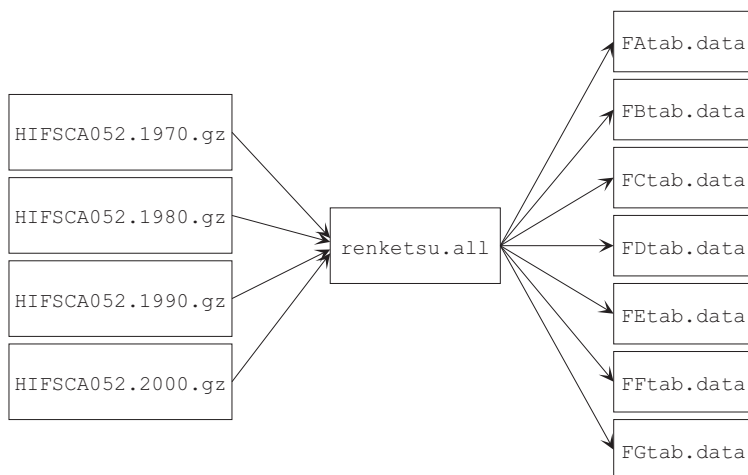


図1：データファイルの前処理

なお、Mac OS X Server と Ubuntu においてファイル `FAtab.data, ..., FGtab.data` を作成する最終段階で `nkf` を用いて文字コードを UTF-8¹⁶⁾

16) UCS Transformation Format 8 の略。(ここで、UCS は Universal Multiple-Octet Coded Character Set の略。) 文字コード系 (character code system) である Unicode の文字

に変換していることに注意しよう。(付録Eのシェルスクリプト・ファイル `script.sh` も参照.)

(手順2) ファイルの移動

(手順1)で作成したファイル `FAtab.data`, ..., `FGtab.data` を MySQL のルートディレクトリに移動 (またはコピー) する. なお, OS 毎の MySQL のルートディレクトリはデフォルトで以下のようなものである.

OS	MySQLのルートディレクトリの場所
Mac OS X Server	<code>/var/mysql</code>
Ubuntu	<code>/var/lib/mysql</code>
Windows	<code>C:\VertrigoServ\MySQL\data</code>

たとえば, Mac OS X Server の場合は以下のように実行する.

```
$ sudo mv F*tab.data /var/mysql
```

(手順3) データベースとテーブルの作成

ターミナルで `loadtoMySQL.sql` が存在するディレクトリに移動し, `mysql` コマンドを使って, MySQL モニタを `root` 権限で起動後, `SOURCE` コマンドを以下のように実行する.

```
mysql> SOURCE loadtoMySQL.sql
```

ここで, `mysql>` は MySQL のプロンプトである. `loadtoMySQL.sql` は以下の操作を行うスクリプトである.

- `CREATE` コマンドでデータベース `hifscsca052` とテーブル `fa01`, ..., `fg01` を作成する.
- `LOAD` コマンドを使って, (手順1) で作成したデータファイル `FAtab.data`, ..., `FGtab.data` をテーブル `fa01`, ..., `fg01` に読み

符号化方式 (Character Encoding Scheme: CES) の一つ. 文字コードに関する詳細については, たとえば, 深沢 (2003) を参照されたい.

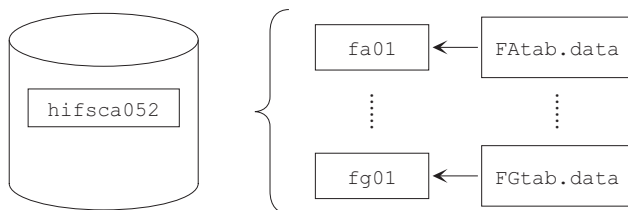


図 2：データファイルのロード

込む。(図 2 も参照。)

(手順 4) 社名リストのテーブルの作成

社名リストのテーブルを作成する際に、データベース hifsca052 に含まれるすべての企業の社名リストを含むテキストファイルを作成する必要があるが、その作業自体は Windows 上でを行い、作成されたファイルを個々のプラットフォームのデータベースに読み込む手順で行う。

以下の手順で社名リストのテーブル firmlist を作成する。

(L1) 会社情報のファイルへの出力

一旦作成したデータベース hifsca052 のテーブル FA01 から、会社の日経コード (a04), 株価コード (a05), 日経業種コード (a22), 社名に関する情報 (b071, ..., b080) を抽出しファイルへ書き出す。実際の作業は、MySQL モニタで以下のコマンドを実行する。(紙面の都合上、スクリプトの途中で改行していることに注意。)

```
mysql> USE hifsca052
mysql> SELECT a04,a05,a22,b071,b072,b073,b074,b075,b076,b077,
           b078,b079,b080 INTO OUTFILE 'temp.txt'
           FIELDS TERMINATED BY '\t' FROM FA01;
```

ここで、出力されたファイル temp.txt は、フォルダ C:\¥Vertrigo Serv\¥mysql¥data に保存されているので、次の手順を他の場所で行う際にはその場所に移動 (またはコピー) しておくこと。なお、MySQL の設定 (my.ini で設定) で文字コードが Shift_JIS¹⁷⁾ または

cp932¹⁸⁾になっていないと、文字コードに関するトラブルが起こる可能性があることに注意しよう。また、Rで処理する段階で、プライム（'）があると読み込む際に失敗するので、エディタまたはインタプリタで削除しておく必要がある¹⁹⁾。

(L2) Rによるファイル処理

(R1) Rを起動後、以下のスクリプトを実行しファイルをオブジェクト temp に読み込む

```
> temp<-read.table("temp.txt",
  colClasses=c("factor","factor","factor",rep("character",9)),
  sep="\t")
```

ここで、>はRのプロンプトである。

(R2) 社名の重複排除を行い、オブジェクト x に代入する。

```
> x<-temp[!duplicated(temp[,1]),]
```

(R3) 以下のRスクリプトを実行し、日経コード nikkeicode、株価コード stockcode、社名（ローマ字表記 shamei.en、日本語表記 shamei.jp、カタカナ表記 shamei.jpj）のオブジェクトを作成し、それらをデータフレームとして結合したものをオブジェクト x へ代入する。

17) 日本工業規格（Japanese Industrial Standards: JIS）で策定されている文字コードに関する規格である JIS X 0201 と JIS X 0208 をもとに作られた文字コード系。「JIS X 0208 漢字」の1バイト目をずらして（シフトさせて）配置したことからこのような名称となった。現在は JIS X 0208 の附属書1で規定されている。詳細については、たとえば、深沢（2003）を参照のこと。

18) Code Page 932 の略。Shift_JIS をもとに Microsoft 社によって Windows 用に拡張された文字コード系。なお、932は「日本」に割り当てられた番号である。

19) ちなみに、プライム記号は“K'S DEKNKI”等の社名で利用されている。


```
> nikkeicode<-as.factor(x[,1])
> stockcode<-as.factor(x[,2])
> nikkeigyousyucode<-as.factor(x[,3])
> shamei.en<-paste(x[,4],x[,5],x[,6],sep="")
> shamei.jp<-paste(x[,7],x[,8],x[,9],sep="")
> shamei.jp<-paste(x[,10],x[,11],x[,12],sep="")
> x<-data.frame(nikkeicode,stockcode,
  nikkeigyousyucode,shamei.en,shamei.jp,shamei.jp)
```

(R4) ファイルへの書き出し

以下のスクリプトを実行し、社名リストのテキストファイル `firmlist.txt` を出力する。

```
> write.table(x,"firmlist.txt",row.name=F,sep="\t")
```

ここで、ファイル `firmlist.txt` には1行目に列名が出力されているので、適当なエディタで削除しておくこと。また、ダブルクオテーション (") も削除しておくこと。

(L3) 社名リストファイルのデータベースへの読み込み

Windows の場合は、データベースの読み込みのためにファイル `firmlist.txt` を `C:\¥VertrigoServ¥MySQL¥data` へコピーし、MySQL モニタから `SOURCE` コマンドを用いてスクリプトファイル `loadfirmlisttoMySQL.sql` を実行する。

```
mysql> SOURCE loadfirmlisttoMySQL.sql
```

(図3も参照。)

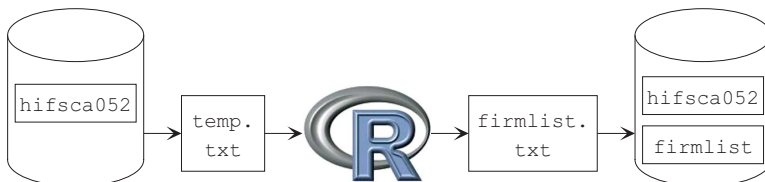


図3：社名リストファイルのロード

なお、Mac OS X Server と Ubuntu の場合は、社名リストファイル `firmlist.txt` の行末コードを `dos2unix` 等のコマンドを用いて Unix 用に変換し、さらに文字コードを `nkf` を用いて UTF-8 に変換したものを、MySQL のルートディレクトリに移動しておく必要があることに注意しよう。

(手順5) データ抽出専用ユーザ `needs` の登録

データベース `hifasca052` からデータを抽出することは、管理者ユーザ `root` でももちろん可能であるが、このユーザは抽出のみならず、すべてのデータベースの変更や削除まで可能であるため、セキュリティの観点から問題がある。このような理由から、多くのユーザがデータベースを利用することを考え、データ抽出を専用とするユーザを登録することが望ましい。データベース `hifasca052` のすべてのテーブルからデータ抽出を可能とする専用のユーザ `needs` を登録するための実際の設定は、MySQL モニタを起動し、`CREATE`、`GRANT` コマンドを用いて以下のように行う。

```
mysql> CREATE USER needs IDENTIFIED BY '*****';  
mysql> GRANT SELECT ON hifasca052.* TO 'needs'@'localhost';
```

ここで、`*****` は、ユーザ `needs` のパスワードを与える。

III データ抽出システムの構築

この節では、II 節で構築されたデータベースサーバから Web 経由でデータを抽出するためのシステムの構築について述べる。なお、このシステムは、以下の手順によってサービスを行う：

データ抽出システムのサービス手順

- (手順1) Web ページ上でデータベースからデータを抽出するための命令を入力し、データベースサーバに送信
- (手順2) 命令の処理
- (手順3) 抽出結果を受け取る

それを実現するための仕様が以下のようなものである。(図4も参照のこと.)

各仕様で利用される主要なファイルの説明を以下の表に与える。なお、これらのファイルのリストを付録Fに与えるのであわせて参照されたい。

データ抽出システムの仕様 (概略)

- (仕様1) HTML の FORM 機能を使って SQL 問合せをデータベースサーバへ送信 (POST)
- (仕様2) PHP スクリプトを使って、受け取った SQL 問合せをファイルへ出力し、データの抽出を実行する
- (仕様3) 抽出を行った結果の HTML ファイルをクライアントへ転送

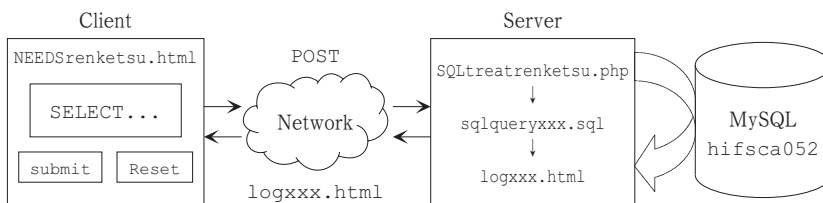


図4：データ抽出システムの概要

ファイル名	説明
NEEDSrenketsu.html	SQL 問合せのスク립トをサーバへ送信するための HTML ファイル。FORM の機能を使って PHP スクリプト SQL renketsu.php へ SQL 問合せを送信 (POST) する。
SQLtreatrenketsu.php	以下の役割を果たす PHP スクリプトファイル。 <ol style="list-style-type: none"> 1. NEEDSrenketsu.html によって送信された SQL 問合せの受け取り 2. SQL 問合せの内容をシェルのプロセス ID 付きで SQL スクリプトファイル sqlqueryxxx.sql へ出力 3. ファイル sqlqueryxxx.sql の内容を引数として mysql を実行することによって抽出結果を logxxx.html へ出力 ここで、xxx にはシェルのプロセス ID が与えられる。
sqlqueryxxx.sql	SQLtreatrenketsu.php によって出力される SQL スクリプトファイル。
logxxx.html	SQLtreatrenketsu.php によって出力される HTML ファイル。

ここで、抽出システムに関するこれらのファイル群は KGUSBADES²⁰⁾ というディレクトリ (またはフォルダ) に一括して納められており、そのディレクトリは Web システムのドキュメントルートの直下に設置している。なお、OS に応じてドキュメントルートはデフォルトで以下のようになっている：

OS	ドキュメントルート
Mac OS X Server	/Library/WebServer/Documents
Ubuntu	/var/www
Windows	c:\VertrigoServ¥www

また、一時的に書き出される sqlqueryxxx.sql や logxxx.html 等のファイルは OS を問わずシステムのトップディレクトリ (/または¥) における tmp ディレクトリに保存されるように設定されている²¹⁾。なお、抽出結果の

20) Kwansai Gakuin University, School of Business Administration, Data Extraction System (関西学院大学商学部データ抽出システム) の略。

21) Windows の場合にはこれらのファイル名には (諸般の事情から) プロセス ID を付加する設定にはしていない。

ファイル `logxxx.html` は、Mac OS X Server と Ubuntu では、最終的な出力の段階でフィルタプログラム `nkf` によって文字コードが `Shift_JIS` に変換されていることに注意しよう²²⁾。

IV クライアント設定

IV.1 Web による利用

Web 経由でデータベースを利用することに関しては、Internet Explorer (IE)、Firefox 等の Web ブラウザが利用可能であればよく、特別な設定を行う必要はない。

IV.2 ODBC による利用：R の場合

通常のデータベースの利用としては、一旦データをデータベースから抽出し、ファイルなどに保存した後に、MS²³⁾ Excel 等の一般のアプリケーションに読み込んで処理することが考えられるが、データベースから直接アプリケーションにデータを読み込んで処理することができれば、データの流れがシームレスとなり利便性が向上するであろう。オープン・データベース・コネクティビティ (Open Database Connectivity; ODBC) は、このことを実現するための仕組みであり、一般のアプリケーションからデータベースにネットワークを経由してアクセスするための標準的なインターフェース規格である。(図5も参照。)

ODBC を使ってデータベースに直接接続可能なアプリケーションの例として、ここでは R を取り上げ、ODBC を利用するための設定について述べる。なお、Windows における ODBC の設定に関する詳細については付録 B を参照されたい。また、R とデータベースとのインターフェースに関しては付録 C を参照されたい。

22) Windows の場合は LAMP 環境構築の際に文字コードを `cp932` で運用する設定にしておき、デフォルトのキャラクターセットを `Shift_JIS` に設定して `mysql.exe` を実行しているのがこの限りではない。

23) Microsoft の略。

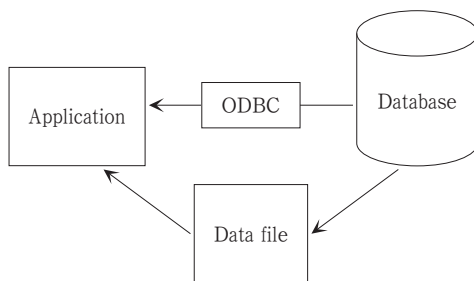


図 5 : ODBC の役割

Windows, Mac OS X, Linux 等のプラットフォームに共通する R に関する ODBC 環境の設定は以下のような手順で行う：

RODBC の設定

- (手順 1) パッケージ RODBC のダウンロード
- (手順 2) パッケージ RODBC のインストール

ここでは設定の具体例として、Windows 上の R にパッケージ RODBC をインストールする場合について以下に与える。なお、R に関するパッケージは CRAN²⁴⁾ サイトからダウンロードすればよい。

(RODBC-W 1) RODBC ドライバ (パッケージ) のダウンロード先の設定

R の [パッケージ] メニューから [CRAN サイトの設定……] を選択すると、[CRAN mirror] というダイアログが開くので、適当なサイト (ここでは筑波大学のサイト) を選択する。

24) Comprehensive R Archive Network の略。R のソースコードや Unix, Linux, Windows, Mac OS 用のバイナリファイル, ドキュメント, アドオンパッケージなどを提供するサイト。URL (Uniform Resource Locator) は、<http://cran.r-project.org/> である。

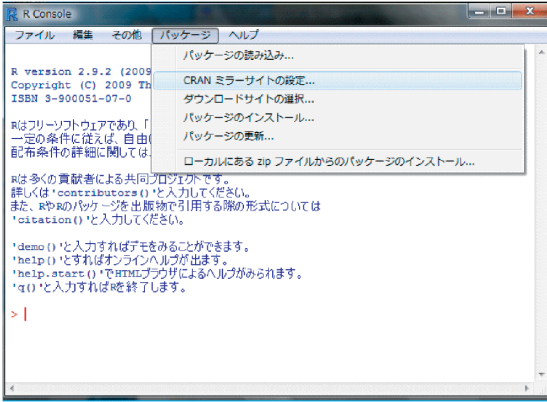


図 6 : CRAN ミラーサイトの設定



図 7 : CRAN ミラーサイトの一覧

(RODBC-W 2) RODBC ドライバ (パッケージ) のインストール

R の [パッケージ] メニューから [パッケージのインストール……] を選択すると, [Packages] というダイアログが開くので, [RODBC] を選択

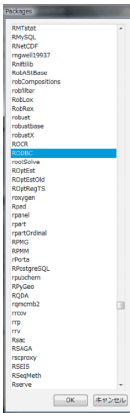


図 8 : インストール
パッケージ
RODBC の選択

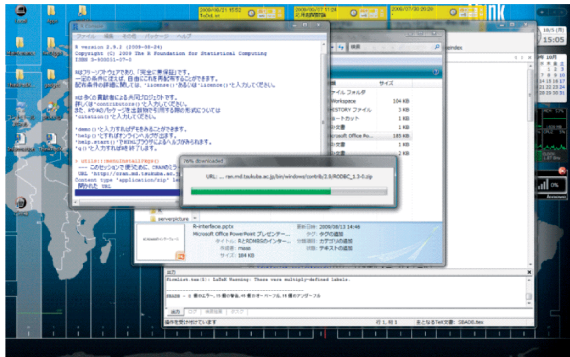


図 9 : パッケージ RODBC をインストールしている
画面

すると、パッケージが手順 (RODBC-W 1) で指定された場所からダウンロードされ、インストールされる。

V データベースの利用

この節では、データベースを Web 経由で利用する場合と、R から ODBC 経由で利用する場合について述べる。なお、両方ともデータの抽出には SQL 問合せ (SQL query) を記述することによって利用する方式をとることに注意しよう。

V.1 SQL 問合せの例

まず、SQL 問合せを行うために最も重要な SELECT コマンドの基本的な用法を与える。

SELECT コマンド

```
SELECT カラム名 1, カラム名 2, ... FROM テーブル名 WHERE 条件 ORDER BY カラム名;
```

ここで、SELECT の次に与える引数はデータベースに用意されているテーブルにおける抽出対象となる「カラム名」であり、複数のものを与えることが可能である。なお、すべてのカラムを抽出対象とする場合はアスタリスク (*) を指定する。次に、FROM 以降に抽出対象となる「テーブル名」を与える。オプションとして WHERE 以下には抽出する際の「条件」を与え、ORDER BY に続いて抽出の順序に利用する「カラム名」を与える²⁵⁾。

では、典型的な SQL 問合せの例を以下に与える。

(SQL-1) 社名 (英文) に TOYOTA という文字列を含む企業の日経コードや社名を抽出

25) 一般に、RDBMS 毎に SQL の実装は異なる可能性があることに注意しよう。なお、MySQL に実装されている SQL に関する詳細に関しては、たとえば、デュボワ (2003)、西沢 (2007) を参照されたい。


```
SELECT * FROM firmlist WHERE shamei_en LIKE '%TOYOTA%';
```

[解説] 企業の社名や日経コードが納められているテーブル `firmlist` におけるカラム `shamei_en` (英文表記の企業名) において `TOYOTA` という文字列を含むフィールドをすべてのカラム (*) に対して抽出する (`SELECT`) ことを意味する。

(SQL-2) トヨタ自動車の「決算年月日」, 「決算月数」, 「売上高・営業収益」を抽出 (決算年月日順)

```
SELECT a02,a07,b001 FROM fc01 WHERE a04='0001353' ORDER BY a02;
```

[解説] 損益計算書のテーブル `fc01` におけるカラム `a04` (日経会社コード) において `0001353` (トヨタ自動車の日経会社コード) という文字列を含むフィールドを `a02` (決算年月日), `a07` (決算月数), `b001` (売上高・営業収益) に対して, 決算年月日順 (`ORDER BY a02`) に抽出する (`SELECT`) ことを意味する。

(SQL-3) 2008年3月31日決算の東京証券取引所1部上場企業の売上高・営業収益と売上総利益の抽出 (日経会社コード順)

```
SELECT a04,b001,b006 FROM fc01  
WHERE a02='2008-03-31' AND a23='1' ORDER BY a04;
```

[解説] 損益計算書のテーブル `fc01` におけるカラム `a02` (決算年月日) において `2008-03-31` (2008年3月31日) であり, かつカラム `a23` (上場情報:東京) において `1` (1部上場) であるフィールドを `a04` (日経会社コード), `b001` (売上高・営業収益), `b006` (売上総利益) に対して, 日経会社コード順 (`ORDER BY a04`) に抽出する (`SELECT`) ことを意味する。

(SQL-4) 2008年3月31日決算の東京証券取引所1部上場企業の売上高・営業収益の抽出 (日経会社コード順, 企業名付き)

```
SELECT fc01.a04, firmlist.shamei_jp, fc01.b001 FROM fc01
      JOIN firmlist ON fc01.a04=firmlist.nikkeicode
      WHERE a02='2008-03-31' AND a23='1' ORDER BY a04;
```

[解説] 損益計算書のテーブル fc01 におけるカラム a02 (決算年月日) において 2008-03-31 (2008年3月31日) であり, かつカラム a23 (上場情報:東京) において 1 (1部上場) であるレコードに対して損益計算書のテーブルにおける日経会社コードのカラム (fc01.a04) と企業リストのテーブルにおける日経会社コードのカラム (firmlist.nikkeicode) とを結合 (JOIN) した結果から, fc01.a04 (損益計算書のテーブルにおける日経会社コードのカラム), firmlist.shamei_jp (企業リストのテーブルにおける日本語の企業名), fc01.b001 (損益計算書のテーブルにおける売上高・営業収益) に対して日経会社コード順 (ORDER BY a04) に抽出する (SELECT) ことを意味する。

(SQL-5) 東京証券取引所1部上場企業の売上高・営業収益と売上総利益のパネルデータ抽出 (企業名付き)

```
SELECT fc01.a04, firmlist.shamei_jp, fc01.a02, fc01.b001, fc01.b006 FROM fc01
      JOIN firmlist ON fc01.a04=firmlist.nikkeicode
      WHERE a23="1" ORDER BY a04, a02;
```

[解説] 損益計算書のテーブル fc01 におけるカラム a23 (上場情報:東京) において 1 (1部上場) であるレコードに対して損益計算書のテーブルにおける日経会社コードのカラム (fc01.a04) と企業リストのテーブルにおける日経会社コードのカラム (firmlist.nikkeicode) とを結合 (JOIN) した結果から, fc01.a04 (損益計算書のテーブルにおける日経会社コードのカラム), firmlist.shamei_jp (企業リストのテーブルにおける日本語の企業名), fc01.b001 (損益計算書のテーブルにおける売上高・営業収益) と fc01.b006 (損益計算書のテーブルにおける売上総利益) に対して日経会社コードと決算年月日順 (ORDER BY a04, a02) に抽出する (SELECT) ことを意味する。

V.2 Web による利用

サーバへのアクセスは、学内から IE などの Web ブラウザで以下の URL にアクセスすればよい：

`http://thorin.kwansei.ac.jp/KGUSBADES/` (Ubuntu)
`http://aule.kwansei.ac.jp/KGUSBADES/` (Mac OS X Server)

URL に指定されている KGUSBADES は Kwansei Gakuin University, School of Business Administration, Data Extraction System (関西学院大学商学部データ抽出システム) の略であり、単に SBADES と略する場合もあるので注意されたい²⁶⁾。図10は、IE を使って `thorin.kwansei.ac.jp` のデータ抽出システムのトップページにアクセスしたときのイメージである。

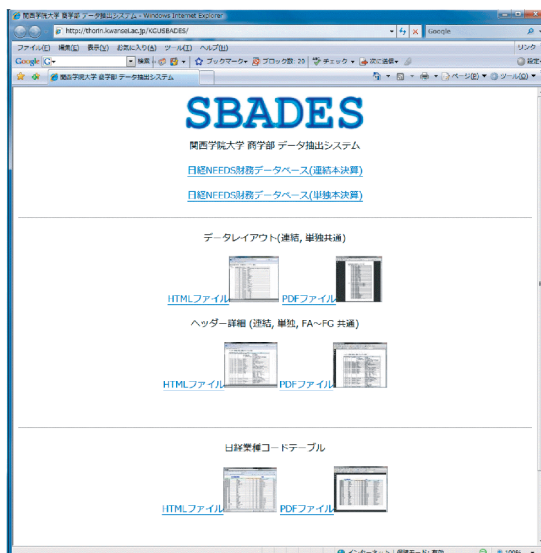


図10：IE7 で `http://thorin.kwansei.ac.jp/KGUSBADES/` にアクセスしたときのイメージ

26) 今回提供する日経 NEEDS 財務データの抽出サービスは、このシステムにおける最初のサービスとして位置づけることができる。

現時点でのデータベースに関するサービスとしては、トップページに用意された以下のリンクに関するものである：

- ・日経 NEEDS 財務データベース（連結本決算）
- ・日経 NEEDS 財務データベース（単独本決算）

また、データベースからデータを抽出する際に必要となる情報である。

- ・データレイアウト（連結，単独共通）
- ・ヘッダー詳細（連結，単独，FA～FG 共通）
- ・日経業種コードテーブル

の各種のテーブルのファイルが HTML, PDF 形式でそれぞれ用意されているので、適宜参照されたい。なお、付録Gにデータベースの構造と各種のテーブルと同等の表を与えているのであわせて参照されたい。

データ抽出システム KGUSBADES のトップページから、たとえば「日経 NEEDS 財務データベース（連結本決算）」というリンクをクリックすることによって、連結本決算のデータを抽出するためのページにアクセスすることができる。図11は、IE を使って thorin.kwansei.ac.jp の連結本決算

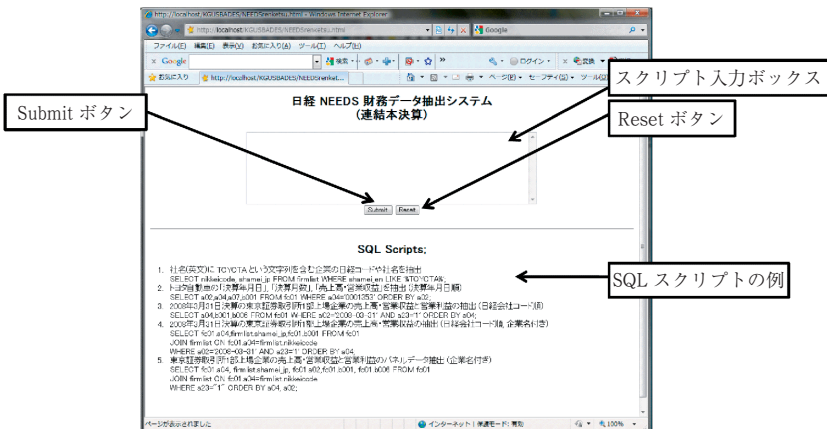


図11：IE7 で連結本決算に関するデータ抽出システムのページ <http://thorin.kwansei.ac.jp/KGUSBADES/NEEDSrenketsu.html> にアクセスしたときのイメージとその説明

に関するデータ抽出システムのページにアクセスしたときのイメージとその説明である。

ここで、実際にデータ抽出を行うには、**スクリプト入力ボックス**にSQL問合せ（SQL query）のスクリプトを入力し、**Submit** ボタンをクリックすることによって、サーバに命令が送信され、結果が画面に表示される。なお、**Reset** ボタンを押すことによって**スクリプト入力ボックス**に入力された文字列をすべて消去することができることに注意しよう。また、SQL問合せのスクリプトの入力例が用意されているのでコピー＆ペーストすることによって利用可能である。

ここでは、トヨタ自動車の日経コードと社名を抽出することを例として取り上げ、データ抽出システムのデモンストレーションを与える。まず、IE等のWebブラウザを使って `thorin.kwansei.ac.jp` の連結本決算に関するデータ抽出システムのページにアクセス後、SQLスクリプトの例1をスクリプト入力ボックスへコピー＆ペーストし、**Submit** ボタンをクリックする。（図12も参照せよ。）

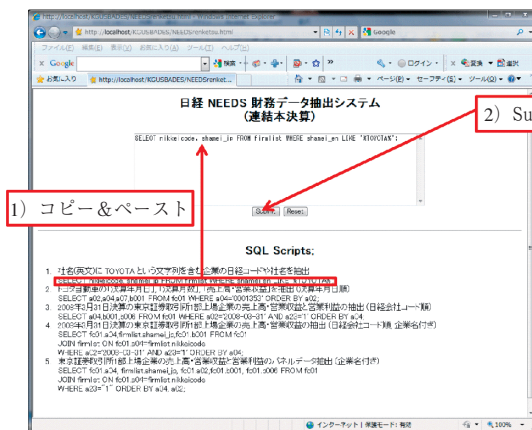


図12：データ抽出の例：トヨタ自動車に関する検索

トヨタ自動車の日経コード

日経コード	会社名
0001353	トヨタ自動車
1011994	富士通
3001016	豊田自衛機
3001956	トヨタ車体
3002585	トヨタソフトウェア
3100026	トヨタ自動車
3000322	トヨタ紡織
3011001	豊田自衛機
301614E	トヨタファイナンス

図13：データ抽出の結果：トヨタ自動車に関する検索

この操作によって、ブラウザ上に図13のような結果が表示され、結果としてトヨタ自動車の日経コードが0001353であることがわかる。

以上のような操作手順によってデータを抽出することが可能であり、これらのデータを手軽に再利用する方法としては、IE等のWebブラウザの[編集(E)]メニューから[すべて選択(A)]を指定し、MS Excelへコピー&ペーストすることであろう。また、HTML形式やテキスト形式のファイルとして保存することも可能であるので、適当なソフトウェアに読み込んで利用することも可能である。

V.3 Rによる利用

ここでは、RのRODBCパッケージを利用してデータベースhifaca052からデータを抽出する例を与える。なお、RODBCパッケージの詳細に関しては付録Cを参照されたい²⁷⁾。

データの抽出に関しては、以下のような入力を行ってパッケージの読み込

27) Rにおけるデータの一般的な取り扱いに関してはスペクター(2008)が詳しい。

みとコネクションの確立を行っておく必要があることに注意しよう。具体的には、R を起動し、[R Console] において library 関数を使って RODBC パッケージを読み込み、odbcConnect 関数を利用してデータベース hifaca052 とコネクションを確立する。

```
> library(RODBC)
> con <- odbcConnect('MyNEEDS')
```

ここで、> は R のプロンプトであり、MyNEEDS は ODBC ドライバのインストール時に登録した、[DSN] 名である。(詳細は、付録Bを参照されたい。) この入力によって、con オブジェクトに R とデータベース hifaca052 との間に確立したコネクションに関する情報が格納される。この後の操作は、オブジェクト con の情報をもとに R の関数 sqlQuery を利用し、SQL 問合せをデータベースに送信することによってデータ抽出が行われることに注意しよう。

V.3.1 トヨタ自動車の売上高・営業収益の抽出 (年月日順抽出)

トヨタ自動車の売上高・営業収益を収録されている年度すべてに関して抽出することを例として与える。以下のように入力すればよい。(なお、誌面の都合上、行の途中で折り返していることに注意されたい。)

```
> x<-sqlQuery(con,'SELECT a02,a07,b001 FROM fc01
                WHERE a04="0001353" ORDER BY a02')
```

ここでは、SQL 問合せをデータベースに送信し、結果をオブジェクト x に代入している。なお、x は、日時、決算期間、売上高・営業収益額を列に持つ以下のようなオブジェクトである²⁸⁾。

28) 正確にはデータフレーム (data.frame) オブジェクトである。

```
> x
      a02 a07      b001
1 1984-06-30 12 5908973
2 1985-06-30 12 6770250
3 1986-06-30 12 6646243
4 1987-06-30 12 6675410
5 1988-06-30 12 7215798
6 1989-06-30 12 8021042
7 1990-06-30 12 9192838
8 1991-06-30 12 9855132
9 1992-06-30 12 10163376
10 1993-06-30 12 10210749
11 1994-06-30 12 9362732
12 1995-03-31 9 8120976
13 1996-03-31 12 10718739
14 1997-03-31 12 12243834
15 1998-03-31 12 11678397
16 1999-03-31 12 12749008
17 2000-03-31 12 12879561
18 2001-03-31 12 13424423
19 2002-03-31 12 15106297
20 2003-03-31 12 16054290
21 2004-03-31 12 17294760
22 2005-03-31 12 18551526
23 2006-03-31 12 21036909
24 2007-03-31 12 23948091
25 2008-03-31 12 26289240
```

このように抽出されたデータは以下のように入力することによって時系列プロットを描くことができる。

```
> plot(as.Date(x[,1]),x[,3],type="l",ylab="売上高・営業収益")
> title("トヨタ自動車の売上高・営業収益")
```

なお、このような入力によって [R Graphics] ウィンドウに図14のようなプロットが描画される。

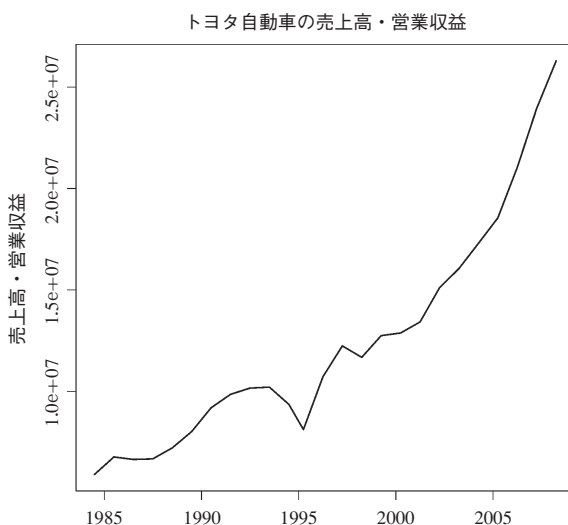


図14：トヨタ自動車の売上高・営業収益の時系列プロット：1995年3月31日の決算は6月30日決算から年度末に決算日の変更があったため、9ヶ月分のものとなっていることに注意されたい。

V.3.2 東証1部上場企業の売上高・営業収益と売上総利益の抽出

2008年3月31日決算の東京証券取引所1部上場企業の売上高・営業収益と売上総利益を日経会社コード順にデータベースから抽出することは以下のように入力することによって実行できる。

```
> y<-sqlQuery(con,'SELECT a04,b001,b004 FROM fc01
WHERE a02="2008-03-31" AND a23="1" ORDER BY a04')
```

ここで、 y は1202社の企業を行とし、日経会社コード、売上高・営業収益、売上総利益を列としてもつオブジェクトである。このように抽出されたデータについて、売上高・営業収益と売上総利益の散布図を描くための命令が以下のようなものである。

```
> plot(y[,c(2,3)],xlab="売上高・営業収益",ylab="売上総利益",pch=20)
> title(main="売上高・営業収益と売上総利益の散布図",
       sub="2008年3月31日決算の東証1部上場企業")
```

なお、このような入力によって [R Graphics] ウィンドウに図15のようなプロットが描画される。

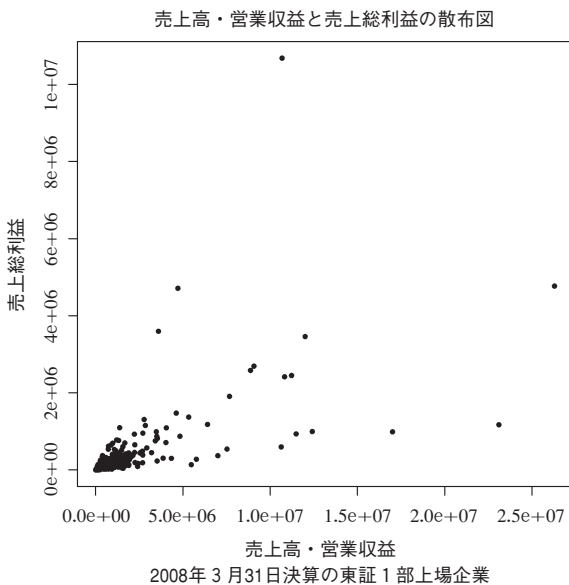


図15：2008年3月31日決算の東証1部上場企業の売上高・営業収益と売上総利益の散布図

このプロット結果から、売上高・営業収益と売上総利益の間には「線形性」(直線関係)が成り立っていないことがわかる。このことは、このデータセットに対して単回帰モデル(回帰直線)を当てはめることが成功しないことを示唆している。

そこで、実験的ではあるが、売上高・営業収益と売上総利益²⁹⁾のそれぞれを対数変換したものの散布図を描き、回帰直線を当てはめることを考える。

```

> plot(log(y[,2]),log(y[,3]+4000),
       xlab="売上高・営業収益",ylab="売上総利益",pch=20)
> title(main="売上高・営業収益と営業利益の散布図:対数スケール",
       sub="2008年3月31日決算の東証1部上場企業")
> abline(lm(log(y[,3]+4000)~log(y[,2]),y),lwd=2)

```

以上のような入力によって図16のようなプロットが得られる。

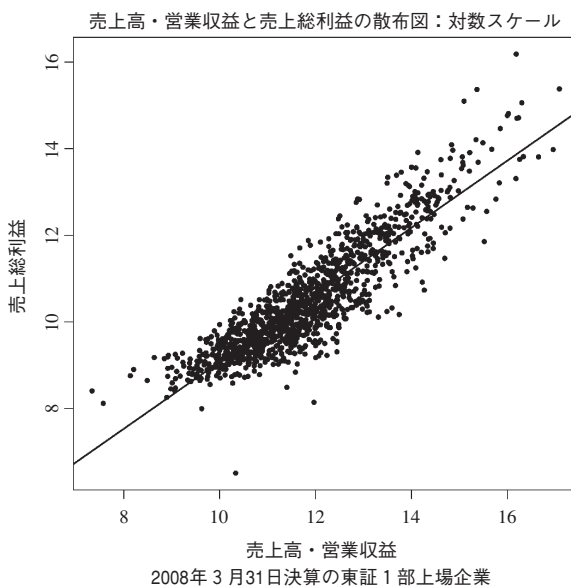


図16: 2008年3月31日決算の東証1部上場企業の売上高・営業収益と売上総利益の散布図:対数スケール

この結果から、いくつかの外れ値が存在することは否めないけれども、(少なくとも粗データに直接、回帰直線を当てはめるよりは)回帰直線の当てはまりが改善されていることが確認できることに注意しよう³⁰⁾。

29) 売上総利益に関しては負のデータが含まれるため、ここでは4000を加えて正值してから対数をとっていることに注意しよう。

30) ちなみに、粗データと対数変換後のデータのそれぞれに単回帰モデルを当てはめたときに、自由度調整済み決定係数は0.44から0.79に増加した。

VI おわりに

本稿は、日経 NEEDS 財務データ MT 版にもとづいたデータベースサーバの構築とその利用法に関して述べてきた。今後の課題としては、インデックスの作成等を検討しデータベースサーバのチューニングを行うことと、金融関連企業の財務データ等の抽出サービスの充実が考えられる。また、前節で扱った東証1部上場企業の売上高・営業収益と売上総利益のデータにもとづく散布図から類推されることであるけれども、財務データは、一般には、非正規・非線形であると考えられるため、2次元対数正規分布等の非正規モデルにもとづく非線形回帰モデルの構築とモデル選択、そしてそれらのモデルにもとづく実証分析等を検討することが課題として考えられる。

(筆者は関西学院大学商学部教授)

謝辞

今回のデータベース環境の構築に関して、幸運にも以下のような資金面での多くの援助を受けることができたことに心から感謝したい。

- ・ 関西学院大学からハードウェア購入のための資金面（研究科特別経費）での援助
- ・ 関西学院大学商学部から日経 NEEDS 財務データの購入のための資金面での援助
- ・ 関西学院大学産業研究所から日経 NEEDS 財務データの購入のための間接的な援助

また、データベース構築に関するハードウェア・ソフトウェアの両面に関して有益なコメントを頂いた関西学院大学経済学部豊原法彦教授に心から感謝したい。さらに、今回のプロジェクトメンバーである関西学院大学商学部の梶浦昭友教授・岡村秀夫教授には様々なコメント・ご助言をいただいたことに、そして様々な面で協力いただいた関西学院大学および商学部の構成員の方々に心から感謝の意を述べたい。以上のような援助や支援なしには今回の

データベースの構築は実現できなかったことをここで改めて述べて本稿の結びとさせて頂く。

参考文献

- [1] P. デュボワ著, 赤井誠他訳 (2003) 『MySQL クックブック VOLUME 1, 2』, オライリー・ジャパン.
- [2] D. ドウラティ, A. ロピンス共著, 福崎俊博訳 (1997) 『sed & awk プログラミング 改訂版』, オライリー・ジャパン.
- [3] 深沢千尋著 (2003) 『文字コード「超」研究』, ラトルズ.
- [4] 金宏和實著 (2007) 『データベースの知識と実務』, 翔泳社.
- [5] 増永良文著 (2003) 『リレーショナル・データベース入門 [新訂版]』, サイエンス社.
- [6] 日本経済新聞社デジタルメディア編 (2009) 『NEEDS 財務データ一般事業会社レイアウト Ver. 4.11』, 日経メディアマーケティング株式会社.
- [7] 西沢夢路著 (2007) 『基礎からのMySQL』, SoftBank Creative.
- [8] R Development Core Team (2009) *R Data Import/Export*, URL <http://www.r-project.org/>, R Foundaition for Statistical Computing, Vienna, Austria, ISBN 3-900051-10-0.
- [9] P. スベクター著, 石田基広, 石田和枝共訳 (2008) 『R データ自由自在』, シュプリンガー・ジャパン.
- [10] 豊原法彦著 (2004) 『地域総合データ読出しシステムの構築』, 経済学論究, 第58巻3号, pp 159-175.
- [11] 豊原法彦著 (2009) 『SQL を用いた日経総合経済ファイル読出しシステムの開発』, 経済学論究, 第63巻第3号掲載予定 (ドラフト版).
- [12] L. ウォール, R. L. シュワルツ, T. クリスチャンセン共著, 近藤嘉雪訳 (1997) 『プログラミング Perl 改訂版』, オライリー・ジャパン.

付録

A LAMP 環境の構築

現在, データベース環境の構築は, Linux 上で Web サーバ Apache とリレーショナル・データベース・マネージメント・システム (RDBMS) である MySQL を導入し, PHP でそれらを連携しながら運用することが標準になっている. これらを略して, LAMP³¹⁾ とよぶ. これらのソフトウェア間の関係については図17を参照されたい.

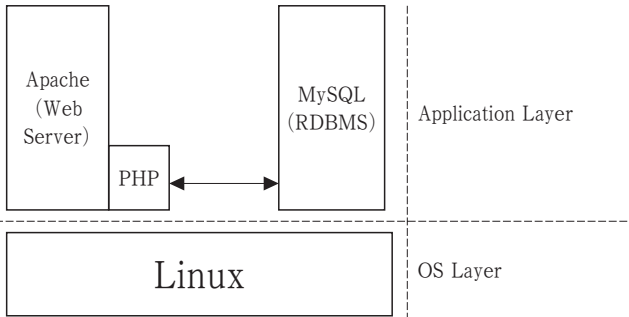


図17：LAMP 環境のイメージ図

ここでは LAMP 環境を，Mac OS X Server (10.5)，Linux (Ubuntu 8.04)，Windows (Windows Vista) といった OS 上に構築することを考える。

A.1 Mac OS X Server 上での LAMP 環境の構築

Mac OS X Server (10.5) に LAMP 環境を構築する際に新たにソフトウェアをインストールする必要はなく，以下の手順に沿って行えばよい。

(手順1) デスクトップ上の [サーバー管理] アイコンをクリックする。

(手順2) 起動させたいプロトコル (ここでは，MySQL, Web 等) にチェックマークを入れ，[保存] ボタンをおす。

(手順3) MySQL の設定に関しては，[サーバー管理] の MySQL を選択し，[MySQL の開始] ボタンをおす。

設定の際に，MySQL における root (管理者ユーザ) のパスワードを聞かれるので設定する必要があることに注意しよう。なお，インストールが正常に完了したかどうかはターミナルから MySQL モニタを起動して以下のように表示されればよい。

31) Windows や Mac OS では，正確には WAMP, MAMP となろうが，ここでは LAMP という用語を統一的に使う。

```
$ mysql -u root -p
Enter password:*****
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 1
Server version: 5.0.67-log Source distribution
Type 'help;' or '\h' for help. Type '\c' to clear the buffer.
mysql>
```

(手順4) Webサーバの設定に関しては、[サーバー管理]で項目[Web]を選択し、[Webの開始]ボタンをおす。なお、Webサービスが行われているかどうかは、適当なWebブラウザでサーバにアクセスし確認する。Web (HTTP) ドキュメントのルートは、/Library/WebServer/Documents であることに注意しよう。

(手順5) PHPの設定と動作確認

i) Webサーバ (Apache) の設定ファイル /etc/apache2/httpd.conf³²⁾において、デフォルトではPHP関連の設定が#によってコメントアウトされているので、適当なエディタ³³⁾を用いて削除し、設定を有効にする。(以下も参照せよ。).

```
#LoadModule php5_module libexec/apache2/libphp5.so
->(修正)
LoadModule php5_module libexec/apache2/libphp5.so
```

ii) 以下の内容のファイル testphp.php をドキュメントルート /Library/WebServer/Documents に作成する。

```
#ファイルtestphp.php
#=====
<?php phpinfo(); ?>
#=====
```

32) Apacheのサービスは、実質的には、Hyper Text Transfer Protocol (HTTP) のデーモン (daemon) プログラムによって行われているため、httpdという略語が利用され、このことからその設定 (configuration) ファイル名が httpd.conf となっている。

33) vi エディタを使って修正することをすすめる。なお、その際 sudo コマンドを併用しなければいけないことに注意しよう。


```
$ sudo su www-data
[sudo] password for masa: ***** # password を入力
$ # 変更できた。
```

ここで、\$ は、Ubuntu のコマンドプロンプトであり、masa はユーザ名、# 以下はコメントであることに注意しよう。

(手順2) PHP の動作確認

i) 以下の内容のファイル testphp.php をドキュメントルート /var/www に作成する。

```
#ファイルtestphp.php
#=====
<?php phpinfo() ; ?>
#=====
```

ii) Web ブラウザでドキュメントルートにある testphp.php を閲覧し、PHP の各種状態を表示した図19のようなページが表示されれば、正常にインストールされていることがわかる。

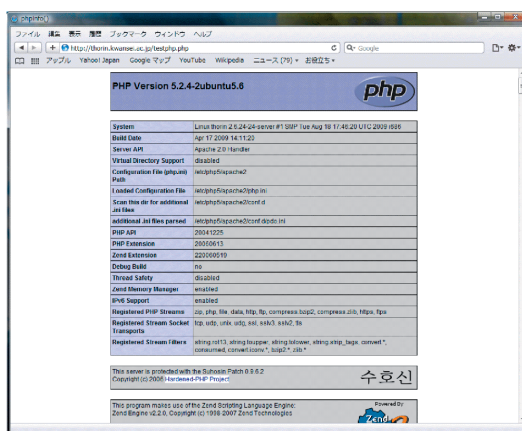


図19: Web ブラウザで Ubuntu のドキュメントルートにある testphp.php を開いたところ

(手順3) MySQL のインストール

ターミナルから、以下のスクリプトを実行する。

```
$ sudo apt-get install mysql-server
```

インストールの途中でパスワードの設定を促されるので指示に従って入力する必要がある。以上でインストールが完了する。インストールが正常に完了したかどうかはターミナルから MySQL モニタを起動して以下のように表示されればよい。

```
$ mysql -u root -p
Enter password: *****
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 35
Server version: 5.0.51a-3ubuntu5.4 (Ubuntu)
Type 'help;' or '\h' for help. Type '\c' to clear the buffer.
mysql>
```

A.3 Windows 上での LAMP 環境の構築

Windows として、ここでは Windows Vista を取り上げ、その上で LAMP 環境を構築する方法について述べる。Windows 上に LAMP 環境を構築するフリーソフトウェアとしては、いくつか存在する³⁴⁾けれども、ここでは VertrigoServ を取り上げる。これは、Apache, PHP, MySQL, SQLite, SQLiteManager, PhpMyAdmin, Zend Optimizer といった一連のソフトウェアを一括インストールすることによって Windows 上に LAMP 環境を構築するソフトウェアである。実際のインストールは以下の手順に沿って行う。

(手順1) Vertrigo Project のページ³⁵⁾から、セットアップ・ファイル³⁶⁾をダウンロードし、管理者権限で実行する。

34) たとえば XAMPP (<http://www.apachefriends.org/en/xampp-windows.html>) 等がある。

35) <http://vertrigo.sourceforge.net/>

36) ここでは、Vertorigo 221.exe を利用している。

- (手順2) インストーラーが起動し、インストール用の言語をきかれるので、[Japanese] を選択する。
- (手順3) ダイアログ [セットアップウィザード] が起動するので、[次へ] を選択する。
- (手順4) ライセンスの同意に関するダイアログが表示されるので [同意する] を選択する。
- (手順5) インストールするコンポーネントを選択するためのダイアログが表示されるので [次へ] (デフォルト) を選択する。
- (手順6) インストール先のフォルダをきかれるので、デフォルトではなく C:\¥VertrigoServ に変更する³⁷⁾。
- (手順7) [スタートメニューフォルダ] の選択をきかれるので、[インストール] (デフォルト) を選択する。
- (手順8) インストールが開始され、(問題がなければ) 完了する。
- (手順9) VertrigoServ のアイコンが [VertrigoServ] グループに登録されるので、ダブルクリックすることによって、VertrigoServ の起動を示すウィンドウが表示される。(図20参照。)

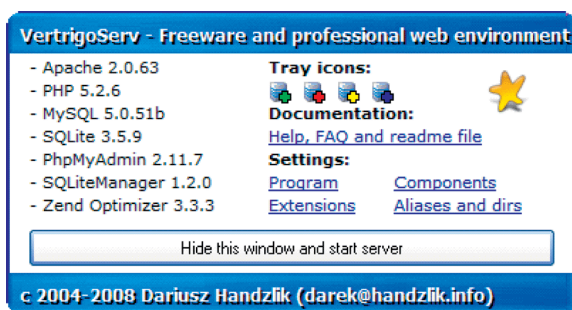


図20：VertrigoServ の起動画面

37) デフォルトのインストール先は、C:\¥Program Files¥VertrigoServ であるけれども、空白 (半角) を含むフォルダ名はトラブルの原因となるため、このような修正を行うことを推奨する。

ここで、[Hide this window and start server] ボタンをクリックすることによって、VertrigoServ の起動を示すアイコンがタスクトレイに表示される。なお、タスクトレイに常駐している VertrigoServ のアイコンにカーソルを合わせて、マウスの右ボタンをクリックすることによって以下のような VertrigoServ のメニューが表示される。

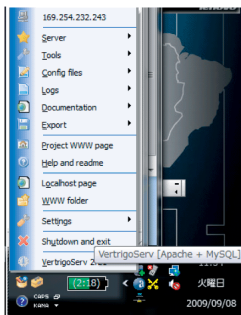


図21 : VertrigoServ のメニュー

このメニューから、MySQL モニタの起動や、各種のプロトコルの設定や再起動等の操作を行うことができることに注意しよう。なお、VertrigoServ

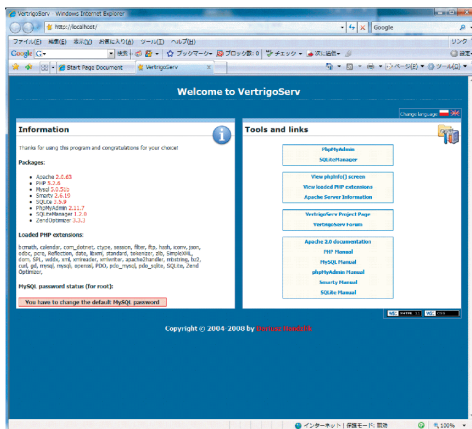


図22 : VertrigoServ がインストールされたサーバの初期トップページ

が正常にインストールされたかどうかを確認する方法は、インストールしたマシン上で Web ブラウザを起動し、`http://localhost/` にアクセス後、図22のように表示されればよい。

(手順10) 文字コードの変更

MySQL に利用される文字コードとして cp932 を利用する場合には、以下のような手順で設定を行う。

- i) VertrigoServ のメニューにおいて [Config Files] から [my.ini] を選択する。
- ii) エディタ（デフォルトは「メモ帳」）が起動し、MySQL の設定ファイル my.ini が編集可能となるので、以下の2行を追記する。

```
[mysqld]
basedir=C:\VertrigoServ\Mysql\
datadir=C:\VertrigoServ\Mysql\data\
port = 3306
key_buffer = 16M
max_allowed_packet = 1M
table_cache = 32
sort_buffer_size = 512K
net_buffer_length = 8K
read_buffer_size = 256K
read_rnd_buffer_size = 512K
myisam_sort_buffer_size = 8M
character-set-server = cp932 # 追記
skip-character-set-client-handshake # 追記
:
: (省略)
:
```

- iii) VertrigoServ のメニューにおいて [Server] から [restart] を選択し、VertrigoServ を再起動する。
- iv) VertrigoServ のメニューにおいて [Tools] から [MySQL Console (for root)] を選択することによって MySQL モニタを起動する。

```
Enter password: *****
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 1
Server version: 5.0.51b-community MySQL Community Edition (GPL)
Type 'help;' or '\h' for help. Type '\c' to clear the buffer.
mysql>
```

- v) MySQL モニタにおいて status コマンドを実行し, charsetset の箇所が cp932 になっていることを確認する.

```
mysql> status
-----
C:\VertrigoServ\mysql\bin\mysql.exe Ver 14.12 Distrib 5.0.51b, for Win32 (ia32)

Connection id:          1
Current database:
Current user:           root@localhost
SSL:                   Not in use
Using delimiter:       ;
Server version:        5.0.51b-community MySQL Community Edition (GPL)
Protocol version:      10
Connection:            localhost via TCP/IP
Server charsetset:     cp932
Db charsetset:         cp932
Client charsetset:     cp932
Conn. charsetset:      cp932
TCP port:              3306
Uptime:                7 hours 45 min 30 sec

Threads: 1  Questions: 7  Slow queries: 0  Opens: 12
Flush tables: 1  Open tables: 0  Queries per second avg: 0.000
-----
```

A.4 TIPS

ここでは, LAMP 運用上で留意すべき点を与える.

1. Windows 環境で VertrigoServ を利用している場合に, 文字コードを Shift_JIS で運用している場合にカタカナの「ソ」などの直後に円マーク (¥) が挿入されるなどの問題が起こる場合がある. これは, いわゆる「5C コード問題」とよばれるものであり³⁸⁾, 環境設定ファイルなど

では対処できないことがある。このような問題がある場合には、MySQL モニタの起動の際にオプション `--default-character-set` に `sjis` を指定することによって問題を回避することができる場合がある：

```
mysql -u root -p***** --default-character-set=sjis
```

2. Web ベースの抽出環境下で以下のようなエラーメッセージが出力され、抽出が失敗する場合がある：

```
Fatal error: Maximum execution time of 30 seconds exceeded in xxx
```

このときは、PHP の設定ファイル `php.ini` における `max_execution_time` を適切な値に設定する必要がある。（初期設定は 30sec である。）

B ODBC 環境の設定

オープン・データベース・コネクティビティ（Open Database Connectivity; ODBC）とは、一般のアプリケーションからデータベースにネットワークを経由してアクセスするための標準的なインターフェース規格である。開発当初は Windows であったが、現在では Linux や Mac OS X 等を含む Unix 系 OS にも実装され一般的に利用されている。ODBC はクライアント・サーバシステムであり、Windows と一般の Unix サーバ上の DBMS に相互接続を可能とする。Windows 上の ODBC 環境はデフォルトでインストールされているが³⁹⁾、Unix/Linux 上では ODBC ドライバ・マネージャが必要である⁴⁰⁾。

38) カタカナ「ソ」の Shift_JIS 文字コード（16進コード）は 835C である。一方、5C は Shift_JIS では円マーク¥に割り当てられているため、ソースコードや文字データの処理において Shift_JIS を想定していないプログラミング環境ではエスケープシーケンスと認識されることから問題が起こる場合がある。

39) 最新のバージョンは <http://www.microsoft.com/data/odbc/> から入手可能である。

40) たとえば、unixODBC (<http://www.unixODBC.org>) や iODBC (<http://www.iODBC.org>) がある。

なお、FreeODBC プロジェクト⁴¹⁾ から ODBC に関する情報が提供されている。

一般に ODBC の環境設定は以下の手順で行われることに注意しよう。

ODBC 環境設定の共通手順

- (手順1) ODBC ドライバのインストール
- (手順2) DSN (Data Source Name) の設定

ここでは、Windows (Vista) に MySQL がインストールされている場合に ODBC 環境を設定する方法について述べる。

(手順1) Windows 用 ODBC ドライバのダウンロード&インストール

MySQL のサイト⁴²⁾ からドライバソフトウェア⁴³⁾ のファイルをダウンロード後、ダブルクリックしてインストールする。なお、インストールは通常のソフトウェアと同様である。

(手順2) ODBC データソースアドミニストレータの起動

[コントロールパネル] から [ODBC データソース] アイコンをダブルクリックすると、[ODBC データソースアドミニストレータ] ダイアログが開くので、[ドライバ] タブで [MySQL ODBC3.51 Driver] が表示されていることを確認する。

(手順3) ドライバの選択

[システム DSN] タブで [追加] ボタンを押すと、[データソースの新規追加] ダイアログが開くので、[MySQL ODBC3.51 Driver] を選択し、[完了] ボタンを押す。

(手順4) DSN 設定

[システム DSN] の項目に対し以下のように入力する。

41) <http://www.jepstone.net/FreeODBC/>

42) <http://dev.mysql.com/downloads/connector/odbc/>

43) 2010年1月現在で最新バージョンは、mysql-connector-odbc-5.1.6-win32.msi である。


```
Data Source Name: MyNEEDS
Description: NEEDS Database by MySQL
Server: localhost
User: needs
Password: *****
Database: hifsc052
```

(手順5) 設定の終了

[ODBC データソースアドミニストレータ] ダイアログを [OK] ボタンを押して閉じる。

C RにおけるRDBMSインターフェースパッケージ

RにはRDBMSと通信するためのいくつかのパッケージが用意されており、それぞれ異なった水準でデータベースとの間でデータの入出力をサポートする。これらのうちRODBCパッケージはODBCインターフェースをサポートしているデータベースに対するインターフェースを提供する。(図23参照。)

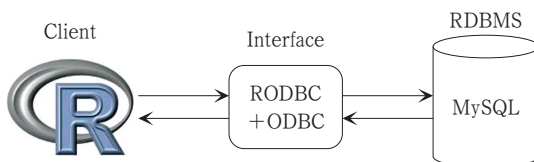


図23：RODBCの役割のイメージ図

RODBCはUnix, Linux, Windows上で利用でき、ほとんどすべてのデータベースシステムがODBCに対するサポートを行っている。このことはRODBCを利用することによって、共通のRのコードで異なったデータベースシステムにアクセスすることが可能となることを示している。なお、RODBCの動作は、Windows上では、Microsoft SQL Server, Access, MySQL, PostgreSQL等であり、Linux上では、MySQL, Oracle, PostgreSQL, SQLite等で確認されている。

データベースとの接続は、関数 `odbcConnect` または `odbcDriverConnect` を利用することによって開始される。なお、MS AccessやMS

Excel のファイルとの接続には `odbcConnectAccess`, `odbcConnectExcel` という関数が用意されている。また、ODBC を使った接続において、`odbcGetInfo` によってクライアントとサーバーの詳細を得ることができる。一方、関数 `odbcClose` によってデータベースとの接続は終了する⁴⁴⁾。接続中のデータベースのテーブルの詳細は関数 `sqlTables` を利用することによって確認できる。関数 `sqlSave` は R のデータフレームをデータベースにおけるテーブルにコピーし、`sqlFetch` はデータベースにおけるテーブルを R のデータフレームにコピーする。SQL 問合せを `sqlQuery` でデータベースに送ることができ、結果は R のデータフレームとして返される。また、`sqlCopy` によって SQL 問合せをデータベースに送り、結果をデータベースのテーブルとしてセーブすることができる。なお、RODBC の詳細は R Development Core Team (2009) を参照されたい。

D 日経 NEEDS 財務データ MT 版に関する解説

D.1 提供データ

本稿で利用したデータセットは、日経メディアマーケティング株式会社から提供されている「日経 NEEDS 財務データ MT 版」である。ここで、MT は Magnetic Tape (磁気テープ) の略であり、当初は磁気テープ媒体を使って配布されていたためにこのような名称になったものと思われる⁴⁵⁾。

なお、収録されているデータは「財務確報」に関する「決算期データ」である。具体的には、以下のようなものである。

収録会社: 全国上場会社 (ヘラクレス, マザーズ含む, 外国部除く), ジャスダック上場会社, 非上場有価証券報告書提出会社

更新タイミング: 各項目を更新タイミングにより基本項目 (貸借対照表・損益計算書・キャッシュフロー計算書など), 付属項目 (各種明細表など) に分類されている。なお、それぞれの更新タイミングは下記の通りである:

44) R を終了させることによっても接続は終了する。

45) 現在は DVD によって配布されていることに注意しよう。

表 1：更新タイミング

項目	上場	店頭	有報（単独）	有報（連結）
基本項目	決算期 3 ヶ月後	決算期 3 ヶ月後	決算期 4 ヶ月後	決算期 4 ヶ月後
付属項目	決算期 4 ヶ月後	決算期 4 ヶ月後	決算期 4 ヶ月後	決算期 4 ヶ月後

提供ファイル：「決算期データ」

表 2：提供ファイル

収録会社	決算種別	ファイル名	更新頻度	収録開始期間
上場・店頭・有報会社	連結本決算	HIFSCA052	月次（月初第一営業日）	1984年4月期～
上場・店頭・有報会社	単独本決算	HIFSUA052	月次（月初第一営業日）	1964年4月期～

レコード長：3040バイト

D.2 データファイル

今回提供されたデータファイルは表 3 ようなものであり、gzip 形式で圧縮されている：

表 3：日経 NEEDS データ MT 版ファイル一覧

ファイル名	決算種別	説明
HIFSCA052.1970.gz	連結本決算	1970年代の企業に関するデータファイル（圧縮済）
HIFSCA052.1980.gz	連結本決算	1980年代の企業に関するデータファイル（圧縮済）
HIFSCA052.1990.gz	連結本決算	1990年代の企業に関するデータファイル（圧縮済）
HIFSCA052.2000.gz	連結本決算	2000年代の企業に関するデータファイル（圧縮済）
HIFSUA052.1960.gz	単独本決算	1960年代の企業に関するデータファイル（圧縮済）
HIFSUA052.1970.gz	単独本決算	1970年代の企業に関するデータファイル（圧縮済）
HIFSUA052.1980.gz	単独本決算	1980年代の企業に関するデータファイル（圧縮済）
HIFSUA052.1990.gz	単独本決算	1990年代の企業に関するデータファイル（圧縮済）
HIFSUA052.2000.gz	単独本決算	2000年代の企業に関するデータファイル（圧縮済）

このファイルを展開すると、以下のような形式のテキストファイルが得られる。なお、ここでは2000年代の企業に関するデータファイル HIFSCA052.2000.gz を展開後に得られるファイル HIFSCA052.2000_525.code 中で、2007年度決算の「極洋」のデータ 1 行分を与えている。

なお、-99999999999999 は「欠損値」を表すことに注意しよう。これらのデータの配列等の詳細は、日経 NEEDS 財務データ MT 版に付属のマニュアル [6] を参照されたい。

E データベース構築に関するスクリプトファイル

E.1 OS 共通スクリプトファイル

ファイル loadtoMySQL.sql

```
DROP DATABASE IF EXISTS hifscsca052;
CREATE DATABASE hifscsca052;
USE hifscsca052
DROP TABLE IF EXISTS fa01;
DROP TABLE IF EXISTS fb01;
DROP TABLE IF EXISTS fc01;
DROP TABLE IF EXISTS fd01;
DROP TABLE IF EXISTS fe01;
DROP TABLE IF EXISTS ff01;
DROP TABLE IF EXISTS fg01;
CREATE TABLE fa01 (
a01 VARCHAR(4),
a02 DATE,
a03 VARCHAR(1),
(中略)
a32 VARCHAR(4) ,
b001 VARCHAR(14),
b002 VARCHAR(14),
b003 VARCHAR(14),
(中略)
b210 VARCHAR(14)
);
CREATE TABLE fb01 LIKE fa01;
CREATE TABLE fc01 LIKE fa01;
CREATE TABLE fd01 LIKE fa01;
CREATE TABLE fe01 LIKE fa01;
CREATE TABLE ff01 LIKE fa01;
CREATE TABLE fg01 LIKE fa01;
LOAD DATA INFILE './Fatab.data' INTO TABLE fa01 FIELDS TERMINATED BY '\t';
LOAD DATA INFILE './FBtab.data' INTO TABLE fb01 FIELDS TERMINATED BY '\t';
LOAD DATA INFILE './Fctab.data' INTO TABLE fc01 FIELDS TERMINATED BY '\t';
LOAD DATA INFILE './FDtab.data' INTO TABLE fd01 FIELDS TERMINATED BY '\t';
LOAD DATA INFILE './FEtab.data' INTO TABLE fe01 FIELDS TERMINATED BY '\t';
LOAD DATA INFILE './FFtab.data' INTO TABLE ff01 FIELDS TERMINATED BY '\t';
LOAD DATA INFILE './FGtab.data' INTO TABLE fg01 FIELDS TERMINATED BY '\t';
```

ファイル loadfirmlisttoMySQL.sql

```
USE hifscsca052
DROP TABLE IF EXISTS firmlist;
CREATE TABLE firmlist (
nikkeicode VARCHAR(7),
stockcode VARCHAR(4),
nikkeigyousyucode VARCHAR(6),
shamei_en VARCHAR(42),
shamei_jp VARCHAR(42),
shamei_jpk VARCHAR(42)
);
LOAD DATA INFILE './firmlist.txt' INTO TABLE firmlist FIELDS TERMINATED BY '\t';
```

E.2 Mac OS X Server, Ubuntu 用スクリプトファイル

ファイル script.sh

```

echo Start unzip!
#
gunzip -c HIFSCA052.1970.gz > renketsuall.txt
gunzip -c HIFSCA052.1980.gz >> renketsuall.txt
gunzip -c HIFSCA052.1990.gz >> renketsuall.txt
gunzip -c HIFSCA052.2000.gz >> renketsuall.txt
#
echo Complete unzip!
#
grep FA01 renketsuall.txt | sed '/^$/d' > FA.data
grep FB01 renketsuall.txt | sed '/^$/d' > FB.data
grep FC01 renketsuall.txt | sed '/^$/d' > FC.data
grep FD01 renketsuall.txt | sed '/^$/d' > FD.data
grep FE01 renketsuall.txt | sed '/^$/d' > FE.data
grep FF01 renketsuall.txt | sed '/^$/d' > FF.data
grep FG01 renketsuall.txt | sed '/^$/d' > FG.data
#
echo Complete Separate!
#
perl -l separate.pl FA.data; sed '/^$/d' temp | nkf --utf8 > FAtab.data; rm temp;
perl -l separate.pl FB.data; sed '/^$/d' temp | nkf --utf8 > FBtab.data; rm temp;
perl -l separate.pl FC.data; sed '/^$/d' temp | nkf --utf8 > FCtab.data; rm temp;
perl -l separate.pl FD.data; sed '/^$/d' temp | nkf --utf8 > FDtab.data; rm temp;
perl -l separate.pl FE.data; sed '/^$/d' temp | nkf --utf8 > FEtab.data; rm temp;
perl -l separate.pl FF.data; sed '/^$/d' temp | nkf --utf8 > FFtab.data; rm temp;
perl -l separate.pl FG.data; sed '/^$/d' temp | nkf --utf8 > FGtab.data; rm temp;
#
echo Complete Tables!
#
rm renketsuall.txt
rm FA.data FB.data FC.data FD.data FE.data FF.data FG.data
#
echo Finish!

```

ファイル separate.pl

```

# code ファイルを読み込んで、項目ごとの列に分割する。
# セパレータはタブ (\t) とする。
#####
open(IN, "@ARGV");
open(OUT, ">temp");
#####
do {
    read(IN, $xx, 3042);
    $a[1]=substr($xx,0,4);
    $a[2]=substr($xx,4,8);
    $a[3]=substr($xx,12,1);
    (中略)
    $a[32]=substr($xx,96,4);
    $b[1]=substr($xx,100,14);
    $b[2]=substr($xx,114,14);
    $b[3]=substr($xx,128,14);
    (中略)
    $b[210]=substr($xx,3026,14);
    print OUT
    "$a[1]\t$a[2]\t$a[3]\t$a[4]\t$a[5]\t (中略) \t$a[32]\t
    $b[1]\t$b[2]\t$b[3]\t$b[4]\t$b[5]\t (中略) $b[209]\t$b[210]\n";
} while (eof(IN)!=1);
close (IN);

```

E.3 Windows 用スクリプトファイル

ファイル script.sh

```

echo Start unzip!
gunzip -c ../../HIFSCA052.1970.gz > renketsuall.txt
gunzip -c ../../HIFSCA052.1980.gz >> renketsuall.txt
gunzip -c ../../HIFSCA052.1990.gz >> renketsuall.txt
gunzip -c ../../HIFSCA052.2000.gz >> renketsuall.txt
#
echo Complete unzip!
#
grep FA01 renketsuall.txt | sed '/^$/d' > FA.data
grep FB01 renketsuall.txt | sed '/^$/d' > FB.data
grep FC01 renketsuall.txt | sed '/^$/d' > FC.data
grep FD01 renketsuall.txt | sed '/^$/d' > FD.data
grep FE01 renketsuall.txt | sed '/^$/d' > FE.data
grep FF01 renketsuall.txt | sed '/^$/d' > FF.data
grep FG01 renketsuall.txt | sed '/^$/d' > FG.data
#
echo Complete Separate!
#
perl -l separate.pl FA.data; sed '/^$/d' temp > FAtab.data; rm temp;
perl -l separate.pl FB.data; sed '/^$/d' temp > FBtab.data; rm temp;
perl -l separate.pl FC.data; sed '/^$/d' temp > FCtab.data; rm temp;
perl -l separate.pl FD.data; sed '/^$/d' temp > FDtab.data; rm temp;
perl -l separate.pl FE.data; sed '/^$/d' temp > FEtab.data; rm temp;
perl -l separate.pl FF.data; sed '/^$/d' temp > FFtab.data; rm temp;
perl -l separate.pl FG.data; sed '/^$/d' temp > FGtab.data; rm temp;
#
echo Complete Tables!
#
rm renketsuall.txt
rm FA.data FB.data FC.data FD.data FE.data FF.data FG.data
#
echo Finish!

```

ファイル separate.pl

```

# code ファイルを読み込んで、項目ごとの列に分割する。
# セパレータはタブ (\t) とする。
#####
open (IN, "@ARGV");
open (OUT, ">temp");
#####
do {
read (IN, $xx, 3041);
$a[1]=substr($xx,0,4);
$a[2]=substr($xx,4,8);
$a[3]=substr($xx,12,1);
(中略)
$a[32]=substr($xx,96,4);
$b[1]=substr($xx,100,14);
$b[2]=substr($xx,114,14);
$b[3]=substr($xx,128,14);
(中略)
$b[210]=substr($xx,3026,14);
print OUT
"$a[1]\t$a[2]\t$a[3]\t$a[4]\t$a[5]\t (中略) \t$a[32]\t
$b[1]\t$b[2]\t$b[3]\t$b[4]\t$b[5]\t (中略) $b[209]\t$b[210]\n";
} while (eof(IN)!=1);
close (IN);

```

F データ抽出システムに関するファイル

F.1 OS 共通 HTML ファイル

ファイル NEEDSrenketsu.html

```
<HTML>
<BODY>
<H2 align="center">
日経 NEEDS 財務データ抽出システム<br>
(連結決算)</H2>
<FORM ACTION="SQLtreatrenketsu.php" METHOD="post">
  <div align="center">
    <p>
      <TEXTAREA NAME="SQLcode" ROWS="10" COLS="80"></TEXTAREA>
      <BR>
      <INPUT TYPE="submit" VALUE="Submit">
      <INPUT TYPE="reset" VALUE="Reset">
    </p>
  </div>
</FORM>
<hr>
<h2 align="center">SQL Scripts; </h2>
<ol>
  <li>社名 (英文) に TOYOTA という文字列を含む企業の日経コードや社名を抽出<br>
    SELECT nikkeicode, shamei_jp FROM firmlist WHERE shamei_en LIKE '%TOYOTA%'; <br>
  </li>
  <li>トヨタ自動車の「決算年月日」,「決算月数」,「売上高・営業収益」を抽出 (決算年月日順)<br>
    SELECT a02,a04,a07,b001 FROM fc01 WHERE a04='0001353' ORDER BY a02;</li>
  <li>2008年3月31日決算の東京証券取引所 1 部上場企業の売上高・営業収益と営業利益の抽出 (日経会社コード順)<br>
    SELECT a04,b001,b006 FROM fc01 WHERE a02='2008-03-31' AND a23='1' ORDER BY a04;</li>
  <li>2008年3月31日決算の東京証券取引所 1 部上場企業の売上高・営業収益の抽出 (日経会社コード順, 企業名付き)<br>
    SELECT fc01.a04,firmlist.shamei_jp,fc01.b001
    FROM fc01 <br>
    JOIN firmlist ON fc01.a04=firmlist.nikkeicode <br>
    WHERE a02='2008-03-31' AND a23='1' ORDER BY a04;</li>
  <li>東京証券取引所 1 部上場企業の売上高・営業収益と営業利益のパネルデータ抽出 (企業名付き)<br>
    SELECT fc01.a04, firmlist.shamei_jp, fc01.a02,fc01.b001, fc01.b006 FROM fc01 <br>
    JOIN firmlist ON fc01.a04=firmlist.nikkeicode <br>
    WHERE a23='&quot;l&quot;' ORDER BY a04, a02;<br>
    <br>
  </li>
</ol>
<p><br>
</p>
</BODY>
</HTML>
```


F.2 Mac OS X Server, Ubuntu 用スクリプトファイル

ファイル SQLtreatrenketsu.php

```
<?php
$pid = exec('echo $$');
$sqlquery = "/tmp/sqlquery".$pid.".sql";
$loghtml = "/tmp/log".$pid.".html";
?>
<?php
$fp = fopen($sqlquery, 'w');
fwrite($fp, stripslashes($_POST['SQLcode']));
fclose($fp);
?>
<?
system("cat $sqlquery | /usr/bin/mysql -u needs -p***** -H hifsc052 | /usr/local/bin/nkf -s > $loghtml");
?>
<?php
include($loghtml);
?>
```

F.3 Windows 用スクリプトファイル

ファイル SQLtreatrenketsu.php

```
<?php
$sqlquery = "/tmp/sqlquery.sql";
?>
<?php
$fp = fopen($sqlquery, 'w');
fwrite($fp, stripslashes($_POST['SQLcode']));
fclose($fp);
?>
<?
system(
"C:\VertrigoServ\Mysql\bin\mysql.exe -u root -p***** --default-character-set=sjis -H hifsc052
< /tmp/sqlquery.sql > /tmp/log.html
")
?>
<?php
include('/tmp/log.html');
? >
```

G データベースの構造

構築された日経 NEEDS 財務データベースは以下のようなものである：

- ・2008年版日経 NEEDS 財務データ MT 版にもとづく連結本決算のデータベース（コード：hifscs052）
- ・2008年版日経 NEEDS 財務データ MT 版にもとづく単独本決算のデータベース（コード：hifsua052）

それぞれのデータベースには以下のようなテーブルが用意されている：

表5：テーブル一覧

テーブル名	内容
fa01	属性
fb01	貸借対照表
fc01	損益計算書
fd01	キャッシュフロー計算書
fe01	その他
ff01	その他2
fg01	株主資本等変動計算書
firmlist	収録企業リスト

テーブル $fx01$ ($x=a, b, \dots, g$) のカラム（列）は以下のような共通の構造を持っていることに注意しよう：

テーブル = ヘッダー部 + データ部

$fx01$ = a01～a32 + b001～b210

ここで、ヘッダー部は、テーブル $fx01$ ($x=a, b, \dots, g$) に関して共通であり、カラムの詳細は表7で与えられる。また、各テーブル $fx01$ ($x=a, b, \dots, g$) に関するデータ部の詳細は表8～14を参照されたい⁴⁶⁾。さ

46) 紙面の都合上、これらの表の項目の一部を省略していることに注意されたい。なお、すべての項目を含むリストのファイル（PDF, HTML 形式）は、商学部のデータ抽出システム KGUSBADES のトップページからダウンロード可能であることに注意しよう。また、すべての項目とその詳細は、日経 NEEDS 財務データ MT 版に付属のマニュアル

らに、テーブル `firmlist` は、表 6 で与えられるようなカラムを持つ収録企業のリストである。最後に、表 15 に日経業種分類表を与えている。

表 6：テーブル `firmlist` のカラム

カラム名	内容	データ型
<code>nikkeicode</code>	日経コード	VARCHAR(7)
<code>stockcode</code>	株式コード	VARCHAR(4)
<code>nikkeigyousyucode</code>	日経業種コード	VARCHAR(6)
<code>shamei_en</code>	社名 (英文表記)	VARCHAR(42)
<code>shamei_jp</code>	社名 (和文表記)	VARCHAR(42)
<code>shamei_jpk</code>	社名 (カタカナ表記)	VARCHAR(42)

表 7：ヘッダー詳細 (単独, 連結, `fa01`~`fg01` 共通)

項番	位置	桁数	カラム名	データ型	項目名	内容
1	1	4	<code>a01</code>	VARCHAR(4)	レコード種別 #: E:財務速報, F:確報 *: 一般事業会社:0, 銀行:1, 証券:2, 保険:3, 信用金庫:4	#A*1:属性 #B*1:貸借対照表 #C*1:損益計算書 #D*1:キャッシュフロー表 #E*1:注記・明細 1 (有価証券注 記, 退職給付会計等) #F*1:注記・明細 2 (リース会計, デリバティブ注記等) #G*1:株主資本等変動計算書
2	5	8	<code>a02</code>	DATE	決算年月日	YYYYMMDD
3	13	1	<code>a03</code>	VARCHAR(1)	識別フラグ	新規・修正: 1, 削除: 9
4	14	7	<code>a04</code>	VARCHAR(7)	日経会社コード	日経が定める会社コード
5	21	4	<code>a05</code>	VARCHAR(4)	株式コード	証券コード協議会が定める 4 桁 の会社コード
6	25	5	<code>a06</code>	VARCHAR(5)	予備	新株式コードのための予備 (スベ ース)
7	30	2	<code>a07</code>	VARCHAR(2)	決算月数	決算月数
8	32	1	<code>a08</code>	VARCHAR(1)	連結・単独フラグ	単独: 1, 連結: 2
9	33	1	<code>a09</code>	VARCHAR(1)	本・中間決算フラグ	本決算: 1 中間決算: 2 9ヶ月累計四半期: 3 第1四半期: 4 第2四半期: 5 第3四半期: 6 第4四半期: 7

ル [6] を参照されたい。

項番	位置	桁数	カラム名	データ型	項目名	内容
10	34	1	a10	VARCHAR(1)	決算期フラグ	新規収録分:1, 次期における修正分:0
11	35	8	a11	VARCHAR(8)	発表決算期	このデータが発表された決算期 YYYYMM00
12	43	17	a12	VARCHAR(17)	予備	スペース
13	60	1	a13	VARCHAR(1)	連結基準フラグ	日本基準:1, SEC基準:2(単独時は0)
14	61	1	a14	VARCHAR(1)	上場フラグ	上場中:1, 未上場・上場廃止:0
15	62	2	a15	VARCHAR(2)	上場場部	東証1部:11 東証マザーズ:13 東証2部:12 ヘラクレス 大証1部:21 スタンダード:91 大証2部:22 グロース :94 名証1部:31 名証2部:32 京都:41 広島:51 福岡:61 新潟:71 札幌:81 未上場:99 ※上場廃止会社は、廃止時の場部
16	64	1	a16	VARCHAR(1)	ジャスダックフラグ	ジャスダック上場:1, ジャスダック未上場:0
17	65	2	a17	VARCHAR(2)	ジャスダック市場	ジャスダック上場:11 ジャスダック未上場:99 ※上場廃止会社は、廃止時のジャスダック市場
18	67	1	a18	VARCHAR(1)	有報フラグ	有報提出会社:1, 非有報提出会社:0
19	68	1	a19	VARCHAR(1)	未上場収録フラグ	未上場会社として収録を継続している:1 収録廃止会社, 上場会社:0
20	69	5	a20	VARCHAR(5)	予備	スペース
21	74	8	a21	DATE	データ作成日	YYYYMMDD
22	82	6	a22	VARCHAR(6)	日経業種コード	ABBCC A…製造業:1, 非製造業:2B…日経業種中分類コード C…日経業種小分類コード
23	88	1	a23	VARCHAR(1)	上場情報:東京	1部上場:1, 2部上場:2, マザーズ:3, 未上場:0
24	89	1	a24	VARCHAR(1)	〃 :大阪	1部上場:1, 2部上場:2, 未上場:0
25	90	1	a25	VARCHAR(1)	〃 :名古屋	1部上場:1, 2部上場:2, 未上場:0
26	91	1	a26	VARCHAR(1)	〃 :京都	1部上場:1, 未上場:0
27	92	1	a27	VARCHAR(1)	〃 :広島	1部上場:1, 未上場:0
28	93	1	a28	VARCHAR(1)	〃 :福岡	1部上場:1, 未上場:0
29	94	1	a29	VARCHAR(1)	〃 :新潟	1部上場:1, 未上場:0
30	95	1	a30	VARCHAR(1)	〃 :札幌	1部上場:1, 未上場:0
31	96	1	a31	VARCHAR(1)	〃 :ヘラクレス	スタンダード:1, グロース:4
32	97	4	a32	VARCHAR(4)	予備	スペース

表 8：属性 (fa01)：データ部 (単独, 連結共通)

項番	桁数	カラム名	データ型	項目名
1	14	b001	VARCHAR(14)	レコードFA01/EA01収録フラグ
2	14	b002	VARCHAR(14)	レコードFB01/EB01収録フラグ
3	14	b003	VARCHAR(14)	レコードFC01/EC01収録フラグ
4	14	b004	VARCHAR(14)	レコードFD01/ED01収録フラグ
5	14	b005	VARCHAR(14)	レコードFE01/EE01収録フラグ
6	14	b006	VARCHAR(14)	レコードFF01/EF01収録フラグ
7	14	b007	VARCHAR(14)	レコードFG01/EG01収録フラグ
8	14	b008	VARCHAR(14)	予備
9	14	b009	VARCHAR(14)	予備
10	14	b010	VARCHAR(14)	予備
(中略):	:	:	:	:
210	14	b210	VARCHAR(14)	予備

表 9：貸借対照表 (fb01)：データ部 (単独, 連結共通)

項番	桁数	カラム名	データ型	項目名
1	14	b001	VARCHAR(14)	流動資産合計
2	14	b002	VARCHAR(14)	当座資産合計
3	14	b003	VARCHAR(14)	現金・預金
4	14	b004	VARCHAR(14)	受取手形・売掛金
5	14	b005	VARCHAR(14)	(うち受取手形)
6	14	b006	VARCHAR(14)	(うち売掛金)
7	14	b007	VARCHAR(14)	非連結子会社関連会社 受取手形・売掛金
8	14	b008	VARCHAR(14)	有価証券
9	14	b009	VARCHAR(14)	棚卸資産合計
10	14	b010	VARCHAR(14)	商品・製品
(中略):	:	:	:	:
210	14	b210	VARCHAR(14)	予備

表10：損益計算書 (fc01)：データ部 (単独, 連結共通)

項番	桁数	カラム名	データ型	項目名
1	14	b001	VARCHAR(14)	売上高・営業収益
2	14	b002	VARCHAR(14)	売上原価・営業原価
3	14	b003	VARCHAR(14)	割賦販売未実現利益・返品調整引当金差額
4	14	b004	VARCHAR(14)	売上総利益
5	14	b005	VARCHAR(14)	販売費および一般管理費
6	14	b006	VARCHAR(14)	営業利益
7	14	b007	VARCHAR(14)	営業外収益合計
8	14	b008	VARCHAR(14)	受取利息・割引料・有価証券利息
9	14	b009	VARCHAR(14)	受取配当金
10	14	b010	VARCHAR(14)	資産処分益・評価益
(中略):	:	:	:	:
210	14	b210	VARCHAR(14)	予備

表11：キャッシュフロー計算書 (fd01)：データ部（単独，連結共通）

項番	桁数	カラム名	データ型	項目名
1	14	b001	VARCHAR(14)	税金等調整前当期純利益
2	14	b002	VARCHAR(14)	減価償却費
3	14	b003	VARCHAR(14)	固定資産売却損益 (▲売却益)
4	14	b004	VARCHAR(14)	固定資産除却損益 (▲除却益)
5	14	b005	VARCHAR(14)	有価証券および投資有価証券売却益 (▲)
6	14	b006	VARCHAR(14)	繰延税金
7	14	b007	VARCHAR(14)	社債発行差金償却額
8	14	b008	VARCHAR(14)	連結調整勘定償却額
9	14	b009	VARCHAR(14)	貸倒引当金の増加額 (▲減少額)
10	14	b010	VARCHAR(14)	退職給付引当金の増加額・年金費用
(中略)：	：	：	：	：
210	14	b210	VARCHAR(14)	予備

表12：その他 (fe01)：データ部（単独，連結共通）

項番	桁数	カラム名	データ型	項目名
1	14	b001	VARCHAR(14)	受取手形割引高
2	14	b002	VARCHAR(14)	受取手形裏書譲渡高
3	14	b003	VARCHAR(14)	貸倒引当金 (棚外注記分)
4	14	b004	VARCHAR(14)	(棚卸資産のうち販売用不動産)
5	14	b005	VARCHAR(14)	(仕掛品のうち未成工事支出金)
6	14	b006	VARCHAR(14)	コマーシャル・ペーパー当期発行額
7	14	b007	VARCHAR(14)	コマーシャル・ペーパー当期償還額
8	14	b008	VARCHAR(14)	税効果会計の対象となった繰越欠損金
9	14	b009	VARCHAR(14)	有形固定資産減価償却累計額
10	14	b010	VARCHAR(14)	減価償却範囲額 (有形無形その他の合計)
(中略)：	：	：	：	：
210	14	b210	VARCHAR(14)	予備

表13：その他2 (ff01)：データ部（単独，連結共通）

項番	桁数	カラム名	データ型	項目名
1	14	b001	VARCHAR(14)	製品売上高
2	14	b002	VARCHAR(14)	商品売上高
3	14	b003	VARCHAR(14)	その他営業収益
4	14	b004	VARCHAR(14)	(▲)売上値引・戻り高
5	14	b005	VARCHAR(14)	売上高・営業収益明細合計
6	14	b006	VARCHAR(14)	期首製品・商品棚卸高
7	14	b007	VARCHAR(14)	当期製品製造原価
8	14	b008	VARCHAR(14)	当期商品仕入高
9	14	b009	VARCHAR(14)	小計
10	14	b010	VARCHAR(14)	期末製品・商品棚卸高
(中略)：	：	：	：	：
210	14	b210	VARCHAR(14)	予備

表14：株主資本等変動計算書（fg01）：データ部（単独，連結共通）

項番	桁数	カラム名	データ型	項目名
1	14	b001	VARCHAR(14)	資本金：前期末残高
2	14	b002	VARCHAR(14)	資本金：新株の発行
3	14	b003	VARCHAR(14)	予備
4	14	b004	VARCHAR(14)	予備
5	14	b005	VARCHAR(14)	予備
6	14	b006	VARCHAR(14)	予備
7	14	b007	VARCHAR(14)	予備
8	14	b008	VARCHAR(14)	資本金：資本金から準備金または剰余金への振替
9	14	b009	VARCHAR(14)	資本金：準備金から資本金または剰余金への振替
10	14	b010	VARCHAR(14)	資本金：剰余金から資本金または準備金への振替
(中略)：	：	：	：	：
210	14	b210	VARCHAR(14)	予備

表15：日経 NEEDS 業種分類表

大分類	中分類	小分類	大分類	中分類	小分類	中+小分類
非製造業	水産	水産	2	35	341	35341
非製造業	鉱業	石炭鉱業	2	37	361	37361
非製造業	鉱業	その他鉱業	2	37	362	37362
非製造業	建設	大手建設	2	41	401	41401
非製造業	建設	中堅建設	2	41	402	41402
非製造業	建設	土木・道路・浚渫	2	41	403	41403
非製造業	建設	電設工事	2	41	404	41404
非製造業	建設	住宅	2	41	405	41405
非製造業	建設	その他建設	2	41	406	41406
製造業	食品	飼料	1	01	001	01001
製造業	食品	砂糖	1	01	002	01002
製造業	食品	製粉	1	01	003	01003
製造業	食品	食油	1	01	004	01004
製造業	食品	酒類	1	01	005	01005
製造業	食品	製菓・パン	1	01	006	01006
製造業	食品	ハム	1	01	007	01007
製造業	食品	調味料	1	01	008	01008
製造業	食品	乳製品	1	01	009	01009
製造業	食品	その他食品	1	01	010	01010
製造業	繊維	化合繊	1	03	021	03021
製造業	繊維	綿紡績	1	03	022	03022
製造業	繊維	絹紡績	1	03	023	03023
製造業	繊維	毛紡績	1	03	024	03024
製造業	繊維	繊維二次加工	1	03	025	03025
製造業	繊維	その他繊維	1	03	026	03026
製造業	パルプ・紙	大手製紙	1	05	041	05041
製造業	パルプ・紙	その他パルプ・紙	1	05	042	05042
製造業	化学	大手化学	1	07	061	07061
製造業	化学	肥料	1	07	062	07062

大分類	中分類	小分類	大分類	中分類	小分類	中+小分類
製造業	化学	塩素・ソーダ	1	07	063	07063
製造業	化学	石油化学	1	07	064	07064
製造業	化学	合成樹脂	1	07	065	07065
製造業	化学	酸素	1	07	066	07066
製造業	化学	油脂・洗剤	1	07	067	07067
製造業	化学	化粧品・歯磨	1	07	068	07068
製造業	化学	塗料・インキ	1	07	069	07069
製造業	化学	農業・殺虫剤	1	07	070	07070
製造業	化学	その他化学	1	07	071	07071
製造業	医薬品	大手医薬品	1	09	081	09081
製造業	医薬品	医家向医薬品	1	09	082	09082
製造業	医薬品	大衆向医薬品	1	09	083	09083
製造業	石油	石油精製及び販売	1	11	101	11101
製造業	石油	石炭石油製品	1	11	102	11102
製造業	ゴム	タイヤ	1	13	121	13121
製造業	ゴム	その他ゴム製品	1	13	122	13122
製造業	窯業	ガラス	1	15	141	15141
製造業	窯業	セメント一次	1	15	142	15142
製造業	窯業	セメント二次	1	15	143	15143
製造業	窯業	陶器	1	15	144	15144
製造業	窯業	耐火煉瓦	1	15	145	15145
製造業	窯業	カーボン・その他	1	15	146	15146
製造業	鉄鋼	鉄鋼一貫	1	17	161	17161
製造業	鉄鋼	平電炉・単圧	1	17	162	17162
製造業	鉄鋼	特殊鋼	1	17	163	17163
製造業	鉄鋼	合金鉄	1	17	164	17164
製造業	鉄鋼	鋳鍛鋼	1	17	165	17165
製造業	鉄鋼	ステンレス	1	17	166	17166
製造業	鉄鋼	その他鉄鋼	1	17	167	17167
製造業	非鉄金属製品	大手精練	1	19	181	19181
製造業	非鉄金属製品	その他精練	1	19	182	19182
製造業	非鉄金属製品	アルミ加工(含ダイカスト)	1	19	183	19183
製造業	非鉄金属製品	電線・ケーブル	1	19	184	19184
製造業	非鉄金属製品	鉄骨・鉄塔・橋梁	1	19	185	19185
製造業	非鉄金属製品	その他金属製品	1	19	186	19186
製造業	機械	工作機械	1	21	201	21201
製造業	機械	プレス機械	1	21	202	21202
製造業	機械	織維機械	1	21	203	21203
製造業	機械	運搬機・建設機械・内燃機	1	21	204	21204
製造業	機械	農業機械	1	21	205	21205
製造業	機械	化工機械	1	21	206	21206
製造業	機械	ミシン・編機	1	21	207	21207
製造業	機械	軸受	1	21	208	21208
製造業	機械	事務機	1	21	209	21209
製造業	機械	その他機械	1	21	210	21210
製造業	電気機器	総合電機	1	23	221	23221
製造業	電気機器	重電	1	23	222	23222
製造業	電気機器	家庭電器(含音響機器)	1	23	223	23223

大分類	中分類	小分類	大分類	中分類	小分類	中+小分類
製造業	電気機器	通信機（含通信機部品）	1	23	224	23224
製造業	電気機器	電子部品	1	23	225	23225
製造業	電気機器	制御機器	1	23	226	23226
製造業	電気機器	電池	1	23	227	23227
製造業	電気機器	自動車関連	1	23	228	23228
製造業	電気機器	その他電気機器	1	23	229	23229
製造業	造船	造船	1	25	241	25241
製造業	自動車	自動車	1	27	261	27261
製造業	自動車	自動車部品	1	27	262	27262
製造業	自動車	車体・その他	1	27	263	27263
製造業	輸送用機器	車両	1	29	281	29281
製造業	輸送用機器	自転車	1	29	282	29282
製造業	輸送用機器	その他輸送用機器	1	29	283	29283
製造業	精密機器	時計	1	31	301	31301
製造業	精密機器	カメラ	1	31	302	31302
製造業	精密機器	計器・その他	1	31	303	31303
製造業	その他製造	印刷	1	33	321	33321
製造業	その他製造	楽器	1	33	322	33322
製造業	その他製造	建材	1	33	323	33323
製造業	その他製造	事務用品	1	33	324	33324
製造業	その他製造	その他製造業	1	33	325	33325
非製造業	商社	総合商社	2	43	421	43421
非製造業	商社	自動車販売	2	43	422	43422
非製造業	商社	食品商社	2	43	423	43423
非製造業	商社	繊維商社	2	43	424	43424
非製造業	商社	機械金属商社	2	43	425	43425
非製造業	商社	化学商社	2	43	426	43426
非製造業	商社	建材商社	2	43	427	43427
非製造業	商社	電機関連商社	2	43	428	43428
非製造業	商社	その他商社	2	43	429	43429
非製造業	小売業	百貨店	2	45	441	45441
非製造業	小売業	スーパー	2	45	442	45442
非製造業	小売業	月販店	2	45	443	45443
非製造業	小売業	その他小売業	2	45	444	45444
非製造業	銀行	長期信用銀行	2	47	461	47461
非製造業	銀行	都市銀行	2	47	462	47462
非製造業	銀行	地方銀行	2	47	463	47463
非製造業	銀行	信託銀行	2	47	464	47464
非製造業	銀行	相互銀行	2	47	465	47465
非製造業	銀行	証券金融	2	47	466	47466
非製造業	証券	証券	2	49	481	49481
非製造業	保険	保険	2	51	501	51501
非製造業	その他金融	その他金融業	2	52	511	52511
非製造業	不動産	賃貸	2	53	521	53521
非製造業	不動産	分譲	2	53	522	53522
非製造業	鉄道・バス	大手私鉄	2	55	541	55541
非製造業	鉄道・バス	中小私鉄	2	55	542	55542
非製造業	鉄道・バス	バス・その他	2	55	543	55543

大分類	中分類	小分類	大分類	中分類	小分類	中+小分類
非製造業	陸運	陸運	2	57	561	57561
非製造業	海運	大手海運	2	59	581	59581
非製造業	海運	内航	2	59	582	59582
非製造業	海運	外航・その他	2	59	583	59583
非製造業	空運	空運	2	61	601	61601
非製造業	倉庫	倉庫	2	63	621	63621
非製造業	倉庫	運輸関連	2	63	622	63622
非製造業	通信	通信	2	65	641	65641
非製造業	電力	電力	2	67	661	67661
非製造業	ガス	ガス	2	69	681	69681
非製造業	サービス	映画	2	71	701	71701
非製造業	サービス	娯楽施設	2	71	702	71702
非製造業	サービス	ホテル	2	71	703	71703
非製造業	サービス	その他サービス業	2	71	704	71704