関西学院大学経済データ抽出システムの構築

<table>
<thead>
<tr>
<th>著者</th>
<th>豊原 法彦</th>
</tr>
</thead>
<tbody>
<tr>
<td>雑誌名</td>
<td>経済学論究</td>
</tr>
<tr>
<td>巻</td>
<td>□</td>
</tr>
<tr>
<td>号</td>
<td>□</td>
</tr>
<tr>
<td>ページ</td>
<td>□</td>
</tr>
<tr>
<td>発行年</td>
<td>□</td>
</tr>
<tr>
<td>年月</td>
<td>□</td>
</tr>
</tbody>
</table>
KG-EDENS;
関西学院大学 経済データ抽出システムの開発

豊原 法彦

Our purpose of this paper is to introduce the data extraction system of the Japanese macro economic data (kg-edens). We designed this system for novice and veteran users, so we can use this system through homepage style interface without a code table and traditional command line. Furthermore, by using a free language under GPL, a central part has a wide portability from unix-base workstations to personal computers that run gawk.

JEL: C82; C87
Key Words: Internet-based methods

1. はじめに
2. kg-edens の構築
3. データの抽出方法
4. すすむ

1. はじめに
経済学部の教育・研究では、まず現実を数値的に把握しここから抽象化することにより、理論モデルを構築した後でさらにその妥当性について実証的なデータを用いて検討するという手法の重要性は、情報教育の進展とあいまって、一層高まってている。日経データ社より発売されている経済マクロデータからユーザが指定する項

1) Kwansei Gakuin Economic Data Extraction for Network System
2) 本システムは、その開発段階において旧 情報処理研究センター（現 情報メディア教育センター）、産業研究所の技術的・資金的サポートを受けている。
目、期間のデータを抽出するためのシステムは、関学では汎用ホスト上で以前より開発され、すべての処理をサーバ上で行うというトータルシステムと抽出のみサーバに依存しこれ以降を端末で行うという2通りの方法が存在していたが、端末台数不足、厳密な端末管理に加えて、前者の場合には操作の特殊性、後者の場合には抽出データの転送手続きの繁雑さ、さらには端末サイドのアプリケーションソフトの能力などの問題が指摘された。

いわゆる1980年代後半からのダウンサイジングという潮流に従って、学内にもワークステーションが多数導入され、その中で統計関係のサーバも含まれており、さらに全学的にネットワークが整備されたことから、日経NEEDSのマクロデータのネットワークに対応した抽出システムを開発することとなった4)5)。その際に考慮した点は以下の3つである。1) ユーザが利用する際に、サーバであるワークステーションとクライアントであるパソコンの間をできる限りシームレスにする。2) 開発費を低廉に抑え、ライセンス料など管理が繁雑にならないようにした上で、メンテナンスなどを容易にすること。3) ライセンスなどの問題が許されればパソコンなどで運用できるようにすること。

データベースなどの特定のアプリケーションに依存した環境では、操作性やアクセススピードなどのメリットを認めつつも、そのソフトの制限に拘束される可能性があることやそのためにサーバ、クライアントとも追加的な投資が必要であることを考慮して、本システムは繰り返しデータ抽出を行うベテラン層から、初歩的な実証分析を始めたノービス層まで簡単なオペレーションで利用できることを目指に開発された。まず、ベテランユーザにはコマンドラインからの利用を想定し、繰り返し抽出を容易にするためにリダイレクトによる入力も考慮している。また、ワークステーションでもパソコンでも最少の変更で利用できるようにするために、いずれ

3) 日立製作所製M680H
4) 同様のコンセプトに基づくもので、日経のマクロデータに関するものとしては豊原（1995）、日本開発銀行のデータに関するものとしては豊原（1999）がある。
5) 同様のシステムは神戸大学経済・経営研究所の安田、阿部によるRIEB(1999)などがある。
のプラットフォームでも入手可能なフリーな言語である日本語版の gawk\(^6\)を本体部分で採用した。

以上のことから、本稿では、われわれが開発した日本経済のマクロデータ抽出システムを紹介することを目的とする。まず第２章では kg-edens の構造について、第３章ではデータ抽出の方法について述べた後で、第４章では今後の課題も含めたまとめを行う。

2. kg-edens の構造

経済のマクロデータを抽出するにはそのデータを特定するコード、期種（暦年，年度、半期、四半期、月次）、抽出開始期、末期、出力先を設定する必要がある。そのためのインターフェイスは gui\(^7\)のようにノービスユーザに便利なものであればそれだけ一層システムへの負荷が大きく、さらにベテランユーザに対して不要な操作を強いることになり、逆にコマンドラインのみのインターフェイスであれば初心者への障壁を高めることになる。

そこで、われわれは次のようにシステムを２つに分割するという手法を採用した。具体的には図１にある本システムの構造の C とマークされた部分がもっぱら抽出を行うパートであり、ベテランユーザはワークステーションにログインしてコマンドラインからもっぱら gawk のスクリプトであるこの部分のみを起動することによってデータ抽出を行う。その際には MT コードの検索といったサービスを受けることはできない。これはこのレベルのユーザであれば同じようなデータを期間や期種を変更して繰り返して抽出したり、以前の資料などを参照できると考えているためである。この場合には極めて軽快に動作することから、数百メガバイト程度の空き容量があるパソコンであれ、システムそのものを移行することも可能である\(^8\)。さらに、unix や ms-dos などのコマンドラインでリダイレクト機能を活用すれば、あ

---

6) awk 言語については Alfred V.Aho 他 (1989)を参照のこと。
7) graphical user interface の略
8) 作動させるプラットフォームによってはディレクトリ構造が異なる場合があるので、その際には若干のスクリプトの変更が生じる場合もある。
らかじめ作成しておいたファイルに上述の各種パラメータを入力しておくことによっ
tて、毎回パラメータを入力することなく、類似した抽出操作を行うことができる。

ホームページを利用するノービスユーザの場合には、MT コード表が手元がない
状況を想定しており、8 つまでの系列を同時に抽出できる。具体的には図1 の A,B
において、期種、抽出開始期と末期を選んだ後で、MT コードを検索するモードに
入り、キーワードを入力するとそれにマッチする項目がリストアップされ、該当す
るものがあればその前にチェックボックスに印をつける。この処理は、unix の grep
コマンドを使って各データのヘッダ部分を検索しているので、正規表現にも対応し
ているが、出力制御をしていないために入力する単語によっては画面に納まりきら
ない場合も考えられる。

図1 kg-edens の処理フロー

```plaintext
ユーザ

ホームページ(html)^A
http://stathost.kwansei.ac.jp/edens
項目のMTコードを検索
（シェル、grepコマンドで記述）

ホームページ(html)^B
年度、項目、出力先を指定
（cgi、perlで記述）

edens本体(gawkで記述)^C

抽出結果の出力（テンポラリーファイル）

端末
exceなどアプリケーションへの読込
```

このように、構造としては項目コードなどを設定する部分（図1 の A,B）と実際
にデータ抽出の処理を行う部分（図1 の C）、さらには出力結果を転送する部分（図
1 の D）からなる。ユーザによって抽出されたファイルはテンポラリーファイルと

— 152 —
してサーバ上に作成される。これを画面に表示するか、端末に転送するかはユーザの選択による。

また、データ部分は717個のファイルからなり、MTコードごとに100の位ごとにまとめて、マイナス数値および欠損値の表現をパソコンなどで利用しやすい形式に変更した上でそれぞれにヘッダをつけたものを1レコードしたものである。これは完全なシーケンシャルファイルであればアクセスに時間がかかり、また1コードごとにファイルを作製したのではファイル数が多くなりすぎるすることを考慮したためである。また、MTコード検索の際に用いられる、約3メガバイトのヘッダーリストファイルも別途用意されている。いずれも、日経データ社より購入した8mmテープのファイル（固定長形式で約92メガバイト）から必要なフィールドを抜き出して作成したものである。これらの作成にはperlを用いた。これは固定長ファイルの処理が容易であるためである。

3. データの抽出方法

まずノービスユーザの場合について述べる。

最初にAにおいてhttp://stathost.kwansei.ac.jp/edensにあるホームページにアクセスし（図2）、そこでBにおいて抽出するデータの期種（年次、半年、四半期、月次）、抽出期間を設定する（図3）。ここでは、年次の場合のみ暦年と年度の選択が必要となる。また元のデータより長い期種を設定することは可能であるが、その逆の処理は行っていない。従って、元のデータが月次のときにはユーザはすべての期種を設定できるが、元データがたとえば暦年のときにはそれ以外の設定は行えず、その旨のエラーが表示されその処理はスキップされる。次に各データのMTコードを検索し、項目名のMTコードを指定して抽出を実行する。

なお、上記の処理の結果は、データ部分とヘッダ部分に別れて出力される。このデータ部分は、たとえば先の場合では図5のような形式になっているために、excelなどの表計算ソフトで容易に読込むことができる。

次に、コマンドラインで利用した場合について述べる。メディア教育センターに
図 2 起動直後の kg-edens

図 3 抽出の条件設定画面
図4 コード検索の画面

図5 kg-edens による出力形式

```
"period","001000016","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","","",...
図6 コマンドラインによる実行

出力するファイル名：不要な場合は空行入力　： testout
期種を入力（1：年次、2：半期、4：四半期、12：月次）　： 4
抽出開始期を入力（70年第1四半期なら7001）　： 7501
最終期を入力　： 9501
対象のMTコードを入力（0で終了）　： 100016
対象のMTコードを入力（0で終了）　： 0

について述べると、図6のEではこの処理の結果を出力するファイル(testout)を設定している。ここでの名前に関するルールはunixに従うのでロングネームも可能であるが、端末に転送することを考えると、2バイト文字（日本語など）ではなく英数8文字以下にしておくことが望ましい。ここで空行入力(何も入力せず
にエンターキーを押すこと)すると、抽出結果は画面に表示される。ターミナルソフトによっては画面の履歴を取得できるのでそれをカットアンドペーストし、若干の修正を行えばそのままパソコンのアプリケーションに取り込むこともできる。

次に期種を指定する。これは年次、半年、四半期、月次の中から選ぶものであるが、年次を選んだときには暦年（1月から12月）か年度（4月から3月）かのいずれかを選択できる。

開始期、最終期については、現時点では西暦の下2桁と2桁の期内番号を設定する。具体的には年次データのときには01のみ、半期では01と02、四半期では01～04、月次では01～12といった数値が入る。年次のときも01が必要である。開始期と最終期が入れ代わっているときには何も出力されないので注意を要する。

最後にMTコードを入力する。このモードではあらかじめ抽出したい項目のコードは判明していることを前提にしているので、それをひたすら入力することになる。そして最後に0を入力すると、抽出を開始する。なおこのモードのときには抽出個数の制限はない。

最後に、コマンドラインでリダイレクトを利用する方法を述べる。先と同様に、
gawk のスクリプトを edens06.awk とし、読込させるべきファイルを testin，出力ファイルを testout とするとき，あらかじめ図 7 のような testin を作っておくと，

```
gawk -f /home/toyohara/data/edens06.awk /home/toyohara/dum<testin
```

と入力しても同様の結果が得られる。さらにこの手続きそのものをシェルスクリプトとして登録し実行属性を与えておけば，繰り返しが一層容易になる。

図 7 コマンドラインからの読込み用ファイル

<table>
<thead>
<tr>
<th>testout</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
</tr>
<tr>
<td>7001</td>
</tr>
<tr>
<td>9504</td>
</tr>
<tr>
<td>100016</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>

なお，コマンドラインから MT コードなどを検索するときには，grep コマンドをもちいて直接ヘッダーファイル (gncent.hed) を調べればよい。具体的に GNP という単語を調べる場合について述べる。

```
grep GNP /stat/edens/data/gncent.hed
```

これは，grep というコマンドで gncent.hed というファイルの中にある GNP という語を含む行を表示せよというものである。なお，/stat/edens/data/は実際にこのヘッダーファイルが置かれているディレクトリを指名している。従って，ユーザがパソコンにシステムを移植している場合には，この部分の記述は変更されねばならない。また，多数の表示が想定されるときには，さらに別のキーワードも追加できる。たとえば 90 年価格を示す 90Y を追加するには

```
9) 上記のシェルスクリプトを含んだファイルを goedens とするとき，以下のようにコマンドラインで入力する。
   chmod +x goedens

--- 157 ---
grep GNP /stat/edens/data/gncent.hed | grep 90Y

とすればよい。なお, unix では大文字と小文字は厳密に区別されるので, その点の注意は必要である。また, ヘッダファイルには日本語のフィールドもあるので日本語による検索も可能であるが, 端末からの日本語入力にはかなり煩瑣な作業が生じるのでここでは述べない。

以上のことから, コマンドラインで実行するためのパラメータ設定 (MT コード検索・設定, 期種, 抽出開始期, 期末の設定など) をノービスモードでは gui によって行っていることが明らかとなりろう。

4. むすび

われわれのシステムを用いることによって, 日経データ社のマクロ経済データからユーザの指定するデータを抽出できることになる。本システムの特徴としては, gawk という unix でも ms-dos でもフリーで利用可能な言語を利用していることに加えて, ときに, ベテランユーザではリダイレクトを用いることによって効率的に作業が行えるといった点があげられる。実際に抽出されたデータは csv10) と言われるテキスト形式であることから, パソコンの表計算アプリケーション, 計量分析に用いられる tsp, rats などにも簡単に取り込むことができる。

今後の課題としては, 教育・研究活動をサポートする点からも収録データベースを増やすことに加えて, 抽出そのものを容易にするための手法, つまりデータベースソフトを導入し, ODBC の活用によって excel などマイクロソフト社製アプリケーションソフトとの親和性も改善する必要がある。

（参考文献）
豊原法彦, 「ネットワークに対応した立命館大学経営データ教育システム: ws-RUBIES の開発」, 『立命館経営学』第 34 巻第 4 号 67～88 ページ, 1995 年 11 月 10 日発行。
豊原法彦, 「DERFI-KG; Data Extraction for Ritsumeikan Financial Institute-

10) comma separated variable の略
KalGin data の構築」. 『立命館経営学』第 37 巻第 5 号 125～154 ページ, 1999 年 1 月 10 日発行.
安田 豊, 阿部茂行. 『RIEB データベースの研究』。神戸大学経済経営研究所刊行、研究叢書 52, 1999.
Alfred V.Aho, Brian W.Kernighan, Peter J.Weinberger (足立高徳 訳). 『プログラマリング言語 AWK』。アジソンウェスレイ/トッパン, 1989 年 11 月
Larry Wall and Randal L.Schwartz (近藤 嘉雪訳). 『Perl プログラミング』。ソフトバンク, 1993 年 9 月発行