

# 統計パッケージ SHAZAM の WWW インタフェース

一橋大学\* 中野純司

University of British Columbia\*\* Kenneth J. White

(1996年6月 受付)

## 1. はじめに

インターネットは世界的規模の計算機ネットワークである。その利用者数は近年爆発的に増加しているが、それは WWW (World Wide Web) 技術の発達に負うところが大きい。

WWW の最大の特徴は、その親しみやすいユーザインタフェースである。WWW から情報を得るためには、Mosaic や Netscape などに代表されるブラウザと呼ばれるソフトウェアを用いる。その操作はほとんどがマウスやトラックボールなどのポインティングデバイスで画面上の場所を指示し、それをクリックするだけであり、キーボードからコマンドを打ち込むという“面倒な”操作はほとんど必要なく、これまでのソフトウェアに較べると格段に使いやすい。さらに、その結果は文字情報だけではなく、写真や動画、音声などいわゆるマルチメディアで提供されるので、直観的にわかりやすい。このように、容易で美しく多角的な GUI (Graphical User Interface) は、計算機の専門家でない多くの人々の強い支持を得たのである。

WWW が支持を得たもうひとつの理由として、情報提供の容易さがあげられる。基本的な情報提供手段であるホームページをつくるには、HTML (HyperText Markup Language) に従ってテキストファイルを作ればよい。これは中学生でも容易に習得できるくらいの簡単なものである。その作成をさらに容易にするためのツールも多く提供されている。作成したホームページを、インターネットに接続された HTTP (HyperText Transfer Protocol) サーバとなつていゝる計算機に置いておけば、そのホームページは世界のインターネットユーザに公開されるわけである。インターネットはこれまでの計算機ネットワークに較べると、計算機を接続するための制限がほとんどないと言えるし、費用も非常に安い。1 個人でも大企業や政府と対等の立場で、世界中に情報を発信することが可能である。

そのような魅力的な媒体であるので、WWW 上で蓄積されたホームページの情報量は、短期間の間に莫大なものとなった。質的にはいろいろなものが混在しているとはいえ、価値の高いものも非常に多く、それらをほとんど無料で見ることができるのである。また、HTTP サーバの機能として、CGI (Common Gateway Interface) が標準的に利用できる。これは双方向的なやりとりを可能にする仕組みであり、これを利用して WWW のデータベース的な利用法なども実現されている。それらにより WWW の価値がますます高いものとなっている。WWW の普及は計算機の歴史において、専門家のための数値計算を行う“計算機”から、一般の人のための日常生活用の“情報端末”への質的転換のはじまりとさえ言える。

このような WWW の機能をいろいろなシステムが利用しようとすることは当然である。統

\* 経済学部：〒186 東京都国立市中 2-1.

\*\* Department of Economics: Vancouver, CANADA V6T 1Z1.

計パッケージにおいてもその利用法はいくつか考えられる。例えば Thioulouse and Chevenet (1996) は主として WWW を用いた統計パッケージの GUI を考察している。本論文では WWW の機能を利用して分散処理を行うためのユーザインタフェースを、統計パッケージ SHAZAM 用に作成する。

SHAZAM は White (1997) らによって作成された、主として計量経済学を念頭においた統計パッケージである。計量経済学で用いられる最新の手法を、多くの計算機で同じように使用できることをひとつの目標として作成されている。比較的安価であり、サイトライセンスを取得すれば構成員にはそのコピーを自由に配布可能である。UNIX 版を購入すればソースファイルが見られることも、研究者には魅力的であろう。ただし、ユーザインタフェースはコマンド言語によるものだけであり GUI はない。グラフ作成機能も非常に限定的なものである。

これまでの統計パッケージにおいては、ネットワーク機能に対応することや分散処理はあまり考慮されていなかったと言ってよい。しかし、今や WWW の機能を用いれば、これらは容易に実現できる。われわれは SHAZAM 言語を拡張し、2つの新しいコマンドを付け加えるとともに、2つのファイル入力コマンドにおいて http サービスに対する URL (Universal Resource Locator) を扱えるような WWW ブラウザ上のユーザインタフェースを作成した。この機能を用いれば、SHAZAM プログラムを複数の計算機で同時に実行したり、異なる計算機上のファイルを共有したりすることが簡単なプログラムで可能になる。

## 2. 拡張されたコマンドとユーザインタフェース

われわれの作成した SHAZAM インタフェースは HTML で書かれたホームページであり、それをブラウザで見るとその中に SHAZAM プログラム用の簡易エディタが埋め込まれている。現在のところブラウザのセキュリティ保護制限のため、このエディタのファイル入出力には、ウィンドウシステムに備わっているカット&ペースト機能を利用しなければならない。

この中では普通の SHAZAM 言語に加えて、拡張された SHAZAM コマンドが使える。SHAZAM 言語では原則として大文字・小文字の区別がなく、一行がひとつの SHAZAM 命令を構成する。われわれは新しい SHAZAM コマンドとして \*run\*, \*all\* を付け加え、既存のコマンド file input, read の用法を拡張した。

\*run\*, \*all\* コマンドを用いることにより、実行するプログラムおよび計算機を指定できる。プログラムのうち、最初の \*run\* コマンドが現れるまでの各行は後の \*run\* コマンドで指定されるすべての URL に送られる。\*run\* コマンドでは実行を行いたい計算機上にある SHAZAM インタフェース CGI プログラムの URL を指定する。その直後から次に出現する \*run\* または \*all\* コマンドの直前までの行は指定された URL に送られる。そして \*all\* コマンドよりあとの各行は指定されたすべての URL に送られる。

なお、\*run\* コマンドがない (したがって \*all\* コマンドもない) 場合は、そのプログラムはホームページの置かれている WWW サーバ上の SHAZAM インタプリタで処理される。

例えば、次のようなプログラムが書ける。

```
* Set sample size to T observations
*run* http://shazam.econ.ubc.ca/~nakanoj/shazam.cgi
      gen1 T=400
*run* http://www.stat.hit-u.ac.jp/cgi-bin/shazam.cgi
      gen1 T=40
```

```
*all*
  sample 1 T
  genr x=nor(1)**2+nor(1)**2+nor(1)**2
  ?nl 1 / ncoef=1 logden
  eq (v-2)/2*(log(x))-x/2-v/2*log(2)-lgam(v/2)
  coef v 1
  end
  stop
```

普通の SHAZAM 言語では、\* で始まる行はコメント行であり、SHAZAM インタプリタには無視される。したがって、このプログラムは文法的に正しい SHAZAM プログラムである。

genl コマンドはスカラー変数の代入を示し、sample コマンドは変数の長さを指定する。genr コマンドは変数の代入を示し、ここでは変数  $x$  の  $T$  個の成分それぞれに対して標準正規乱数 3 個の 2 乗の和を代入している。nl コマンドは非線形モデルのパラメータ推定を行うもので、end コマンドまでの部分で発生させたデータに対して  $\chi^2$  分布の自由度の最尤推定を行っている。記号 ? は出力を抑制するためのものである。

このプログラムをホームページの簡易エディタ内に書き、“Run SHAZAM” ボタンをクリックすると、インタフェースプログラムは次のように動作する。まず、

```
* Set sample size to T observations

  genl T=400

  sample 1 T
  genr x=nor(1)**2+nor(1)**2+nor(1)**2
  ?nl 1 / ncoef=1 logden
  eq (v-2)/2*(log(x))-x/2-v/2*log(2)-lgam(v/2)
  coef v 1
  end
  stop
```

というプログラムが、ドメイン名 shazam.econ.ubc.ca の計算機に送られ、そこで実行される。同時にこのプログラムの

```
genl T=400
```

という行が

```
genl T=40
```

に置き換えられたものが、ドメイン名 www.stat.hit-u.ac.jp の計算機に送られ、そこで実行される。両方の計算機の計算結果は WWW ブラウザの出力ウィンドウに、計算が終了した順に、実行した計算機の名前などを最初に付け加えられたかたちで順次表示される。図 1 の左側のブラウザで、この結果が表示されている。

この 2 つの拡張コマンドは、この例のようなシミュレーションを行う場合に特に有効であろう。\*run\* 行に続く行ではそこで指定された計算機で実施されるシミュレーションで用いられるパラメータの値などを指定し、\*all\* 行以下の部分でシミュレーションのメインプログラム

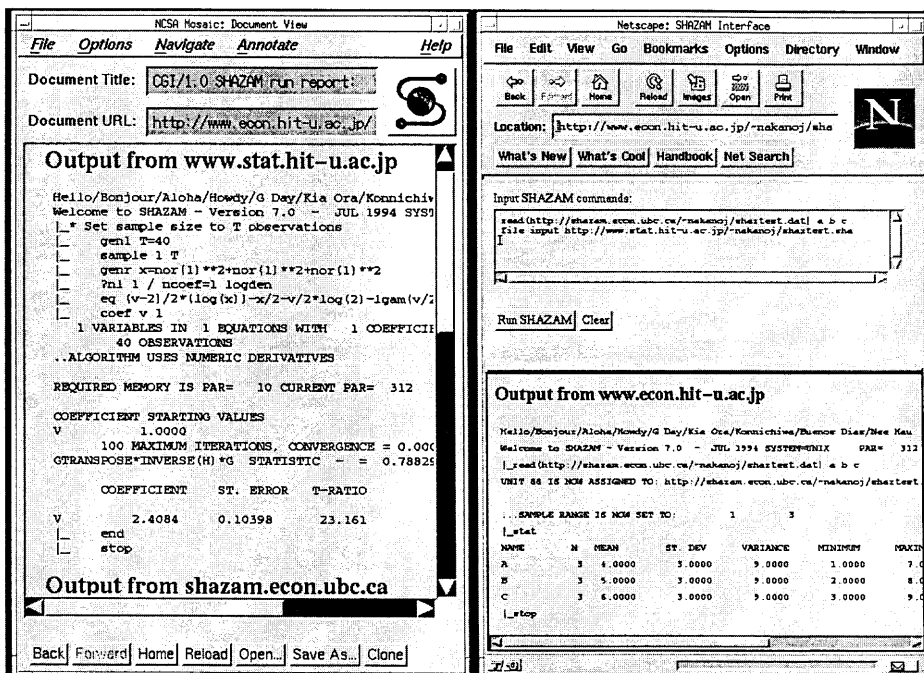


図1. WWW インタフェースの実行例。

のようなものを記述する。このようにすると、普通の SHAZAM プログラムにわずかな変更を加えるだけで、計算量の多いシミュレーションを複数の計算機で同時に分散処理することができる。

もうひとつの拡張として、プログラムファイル読み込みのための file input およびデータファイル読み込みのための read コマンドにおいて、ファイル名のかわりに http サービスに対する URL を指定することができるようにした。例えば、プログラム

```
read(http://shazam.econ.ubc.ca/~nakanoj/shaztest.dat) a b c
file input http://www.stat.hit-u.ac.jp/~nakanoj/shaztest.sha
```

の第1行ではドメイン名 shazam.econ.ubc.ca の計算機上にあるファイルからデータを変数 a, b, c に読み込み、第2行ではドメイン名 www.stat.hit-u.ac.jp の計算機上にあるプログラムファイルを読み込んで実行を行う。この機能は教育において便利である。プログラム例やデータ例を WWW サーバ上においておけば、世界中の SHAZAM プログラムからこの例のように簡単にそれらを参照することができる。

このプログラムの実行例は図1の右側のブラウザで示されている。このブラウザでは HTML タグとして、ひとつの画面を分割して表示するための FRAME タグが利用できる。ここではそれを用いてプログラムと出力を同時に表示している。FRAME タグをサポートしていないブラウザでこのホームページを見た場合には、最初はプログラムエディタの部分があれば、計算結果が得られれば図1の左側のブラウザのように、出力ウィンドウに切り替わる。

### 3. インタフェースの実現

前節で述べたような機能を実現するために HTTP サーバの CGI 機能を利用する。すなわち、SHAZAM インタフェースを含むホームページで HTML 言語の FORM タグを用いて、拡張 SHAZAM 言語をプログラムエディタに入力した後、“Run SHAZAM” ボタンをクリックすることにより、それらの情報をそのホームページのある計算機（ローカルコンピュータと呼ぶことにする。ブラウザを起動している計算機のことではないので注意してほしい）上の CGI スクリプトに送る。ここでの CGI スクリプトは shazam.cgi というプログラムであり、以下のように拡張 SHAZAM 言語を処理する。

1. まず file input 行をチェックする。その行が見つかり、ローカルファイル（ローカルコンピュータ上のファイルのこと）が指定されていれば、その行をローカルファイルの内容と置き換える。もし、ローカルファイルのかわりに http サービスに対する URL が指定されていれば、最初にそれを適当な名前のローカルファイルにコピーし、URL をそのファイル名で置き換える。そしてそれをローカルファイルとして処理する。この処理は繰り返し行われるので、file input で読み込まれるファイルの中に再び file input があってもよい。
2. 次に \*run\* 行をチェックする。それがあつた場合には、そこで指定されたドメイン名を持つ計算機（これをリモートコンピュータと呼ぶ）に送るためのプログラムを一時ファイルとして格納する。このとき \*all\* 行もチェックし、それぞれのリモートコンピュータのための完全なプログラムを作成する。これらの一時ファイルにはプログラムの他に \*run\* 行で指定されたドメイン名と CGI スクリプト名が保持される。それらは、sendshazam.tcl プログラムによってそれぞれのリモートコンピュータ上の shazam.cgi スクリプトに送られる。
3. もし \*run\* 行がなければ、shazam.cgi スクリプトは全行を調べて read について http サービスに対する URL があるかどうか検査し、それがあれば 1. と同様にローカルファイルにコピーするなどの処理を行う。
4. この段階で普通の SHAZAM システムが処理できるプログラムが得られているのでそれを SHAZAM インタプリタに渡す。その計算結果は、read や file input における一時ファイル名をもとの URL に変更し、先頭にリモートコンピュータの名称を示す部分を付け加えたあと HTML 形式で返される。ローカルコンピュータの shazam.cgi スクリプトはローカルおよびリモートコンピュータからの結果を集め、それを完全な HTML 言語の形式に整え、ブラウザの出力ウィンドウに返す。

われわれはこのインタフェースに関するすべてのプログラムを TclX 言語（バージョン 7.4a-p2）で作成した。TclX (extended Tcl) はスクリプト言語 Tcl (Ousterhout (1994)) の拡張であり、TCP/IP 接続や UNIX 特有の機能などが強化されたものである（この強化点の多くは Tcl の新しいバージョンに採り入れられている）。今回のわれわれのプログラムでは SHAZAM 自体には変更を一切加えず、全てを TclX スクリプトで記述した。この実現法は SHAZAM のようなコマンド言語による統計パッケージにおいては、もっとも容易な拡張法であると思われる。また、計算部分のプログラム変更に伴うバグの混入がないので、計算部分が複雑で信頼性の要求される統計パッケージの拡張法として望ましいものである。半面、ユーザインタフェースの根本的な改善は難しい。Tcl (およびその GUI である Tk) を統計パッケージで利用した例は、山本 他 (1995) でも見られる。

#### 4. おわりに

WWWはその使いやすいユーザインタフェースのために情報獲得および提供の手段として急速に普及しているが、分散処理を行うための簡単なプロトコルとしても有用である。われわれはその機能を利用して統計解析システム SHAZAM のための分散処理用インタフェースのプロトタイプを作成してみた。このシステムにより簡単な SHAZAM プログラムで、大規模なシミュレーションを複数の計算機で同時に実行させたり、データやプログラムをネットワーク上で共有したりすることが可能になる。

WWW 技術は現在でも急速に発展しているが、われわれのシステムでは古典的とも言える CGI 機能だけを利用した。そのため、ほとんどすべてのサーバやブラウザで利用することができる。しかし、新しい WWW 技術を利用すれば今回実現した機能をさらにいろいろな方向に拡張して、より使いやすくすることができると思われる。また、WWW におけるセキュリティは非常に重要な問題であるが、本システムではほとんど考慮していない。これらは今後の課題としたい。

われわれのシステムは <http://www.stat.hit-u.ac.jp> および <http://shazam.econ.ubc.ca> から入手できる。

#### 謝 辞

システムの作成において、インターネット上で公開されているいくつかの Tcl および TclX プログラムを利用させていただいた。その出典に関してはシステムのソースコードを参照していただきたいが、それらのプログラムの作者の方々、また Tcl および TclX の開発者の方々に深く感謝する。さらに査読者の方には、多くの有益なコメントをいただいたことを感謝したい。

#### 参 考 文 献

- Ousterhout, J. K. (1994). *Tcl and the Tk Toolkit*, Addison-Wesley, Reading, Massachusetts.
- Thioulouse, J. and Chevenet, F. (1996). NetMul, a World-Wide Web user interface for multivariate analysis software, *Comput. Statist. Data Anal.*, **21**(3), 369-372.
- White, K. J. (1997). *SHAZAM User's Reference Manual Version 8.0*, McGraw-Hill, New York.
- 山本由和, 中野純司, 田村義保 (1995). MS-Windows 版 TIMSAC72 の UNIX への移植, *統計数理*, **43**(2), 285-292.

## A WWW Interface for the Statistical Package SHAZAM

Junji Nakano

(Department of Economics, Hitotsubashi University)

Kenneth J. White

(Department of Economics, University of British Columbia)

WWW (World Wide Web) technologies greatly promote the use of Internet. They provide a well designed GUI (Graphical User Interface), which is easy to handle and beautiful to see. They also provide a simple access protocol for distributed computing. We made these merits of the WWW available for users of the SHAZAM statistical program by writing CGI (Common Gateway Interface) scripts. Our SHAZAM interface on a WWW home page extends SHAZAM commands to be able to handle URLs (Universal Resource Locator) in order to realize distributed computation without difficulty.