URL 閲覧履歴における有効な日数の解析に関する研究

97H138 野崎謙一
nozaki@h-ps011.ise.osaka-sandai.ac.jp

2001年 2月 19日

目 次

1 はじめに **1**

2 概要 **1**

2.1 個人別の Web ページ閲覧履歴 1

3 有効な n 日分と n 日前の Web ページ閲覧履歴 **5**

3.1 有効な n 日分 5

3.2 有効な n 日前 5

4 さまざまな履歴選択方法とカレンダー表示 **6**

4.1 各日毎の履歴 7

4.2 過去 30 日間の履歴 7

4.3 過去 4 週間で 1 週間づつの履歴 7

4.4 曜日毎の履歴 7

4.5 有効な数日分 (前) の履歴 7

4.6 各履歴選択の考察 7

5 考察 **8**

6 謝辞 **8**

1　はじめに

現在、ほとんどの大学の演習室のパーソナルコンピュータ (PC) は、複数の PC を特定多数が利用し、自分専用の PC が割り当てられていることは少ない。そのような環境では、ユーザが過去に訪れた Web ページ閲覧の履歴はユーザーが PC を変更するたびに見れなくなり、非常に不便である。また、ユーザーが特定の一台の PC を使っていても、他のユーザーがその PC を使うことによって、ユーザー独自の過去に訪れた Web ページの履歴を閲覧することが出来ない。

また、一般的なブラウザ (Mozilla など) では履歴が表示されるだけで、ユーザーが過去に訪れた Web ページはどれだけの頻度で訪れているのか分からず、良く行く Web ページを探すのは必ずしも容易ではない。そこで、本研究はいかにユーザーの Web 情報の検索・閲覧を支援することが出来るか検討する。

2　概要

本研究は去年の卒業研究に取り組んだ、高徳哲也の「検索機能を付けた個人別履歴の作成」と宮北哲博の「URL の履歴の自動作成」をさらに発展させ、ユーザー独自の履歴を作成して表示する。このことは、上記に述べたように、複数の PC を特定多数のユーザーが利用するような環境で有効ある。

そして、各日の履歴以外にも多彩な履歴選択方法 (有効な数日間の履歴、有効な数日前の履歴選、各日の履歴、過去 30 日間の履歴、曜日毎の履歴、過去 4 週間で 1 週間づつの履歴) を選べ、また、プロトコル、ホスト、ドメイン別に分けた簡易版を選択する事により、ユーザはより楽に履歴を見ることが出来る。

2.1　個人別の Web ページ閲覧履歴

個人別の Web ページ閲覧履歴を表示するのに Proxy サーバにあるログファイルを利用する。また、個人を識別するために、ident デーモンを使用することにより、利用者を識別して、個人別の Web ページ閲覧履歴を作成することが出来る。(クライアントと Proxy,Web サーバとの関係を図 1に示す。)

本研究はその Proxy サーバに保存されているログを使い、ユーザー毎の過去に訪れた Web ページの履歴を図 2の様にブラウザ上に表示することが出来る。また、簡易を選択することにより、図 3のような、プロトコル、ホスト、ドメイン別に表示することができる。

これらの表示されたクリッカブルに対応しているため、ユーザーは URL をクリックするだけで、その URL の Web ページを表示することが出来る。

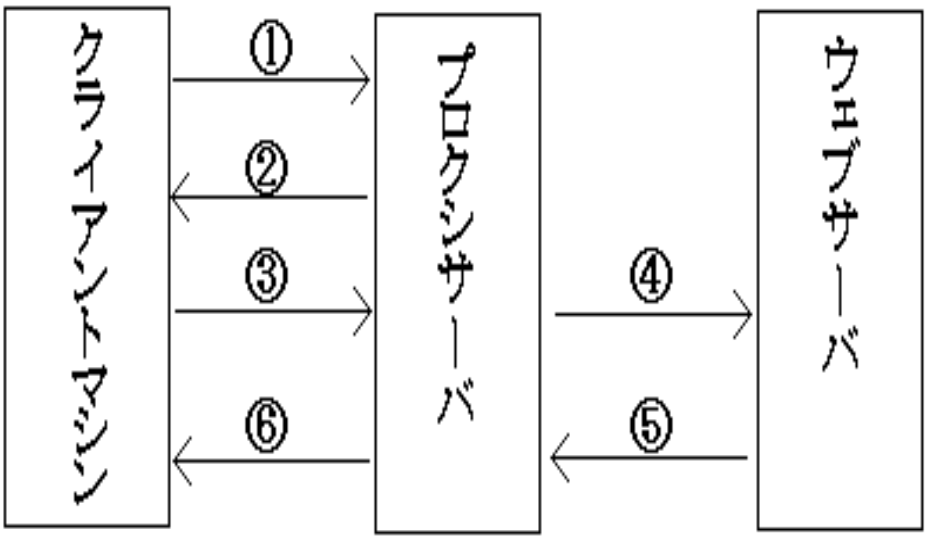


図 1: クライアントと Proxy,Web サーバーの関係

1. クライアントから Proxy サーバに対して，Web ページのデータを要求する。
2. Proxy サーバはクライアントの IDENT_DAEMON に対して、ユーザ名を問い合わせをする。
3. クライアントは Proxy サーバーに対してユーザ名を返す。
4. Proxy サーバは Web サーバに Web ページのデータを要求をする。
5. Web サーバは proxy サーバに Web ページのデータを渡す。
6. Proxy サーバはクライアントに Web ページのデータを渡す。

Netscape
ファイル(F) 編集(E) 表示(V) ジャンプ(G) Communicator(C) ヘルプ(H)
戻る
次
再読み込み
ホーム
検索
ガイド
印刷
セキュリティ
Shop
現在?の日 * 2001 / 01 / 17 (Wed) - 10 : 36 : 46s
nozaki さんの履歴 0
ABMへ戻る
簡易
• 4 点 http://www.lycos.co.jp/
• 2 点 http://www.google.com/
• 1 点 http://dir.lycos.co.jp/computer internet/
• 1 点 http://dir.lycos.co.jp/computer internet/internet/e mail/
• 1 点 http://dir.lycos.co.jp/computer internet/network/
• 1 点 http://dir.lycos.co.jp/computer internet/software/network/
• 1 点 http://dir.lycos.co.jp/computer internet/software/network/telnet/
• 1 点 http://hp.vector.co.jp/authors/VA002416/
• 1 点 http://www-fken.ise.osaka-sandai.ac.jp/fken99g/41/T/
• 1 点 http://www.geocities.co.jp/Hollywood/9752/kankyo.html
• 1 点 http://www.google.co.jp/
• 1 点 http://www.ibport.co.jp/~yui/cgiwork/mac/telnet/
• 1 点 http://www.infoseek.co.jp/
• 1 点 http://www.kcctec.co.jp/
• 1 点 http://www.kcctec.co.jp/company.html
• 1 点 http://www.kcctec.co.jp/division.html
• 1 点 http://www.kcctec.co.jp/saiyo.html
• 1 点 http://www.kcctec.co.jp/saiyofukuri.html
• 1 点 http://www.kcctec.co.jp/saiyojisseki.html
• 1 点 http://www.kcctec.co.jp/saiyotaigu.html
• 1 点 http://www.yahoo.co.jp/
ドキュメント: 完了。

図 2: ユーザー別の履歴

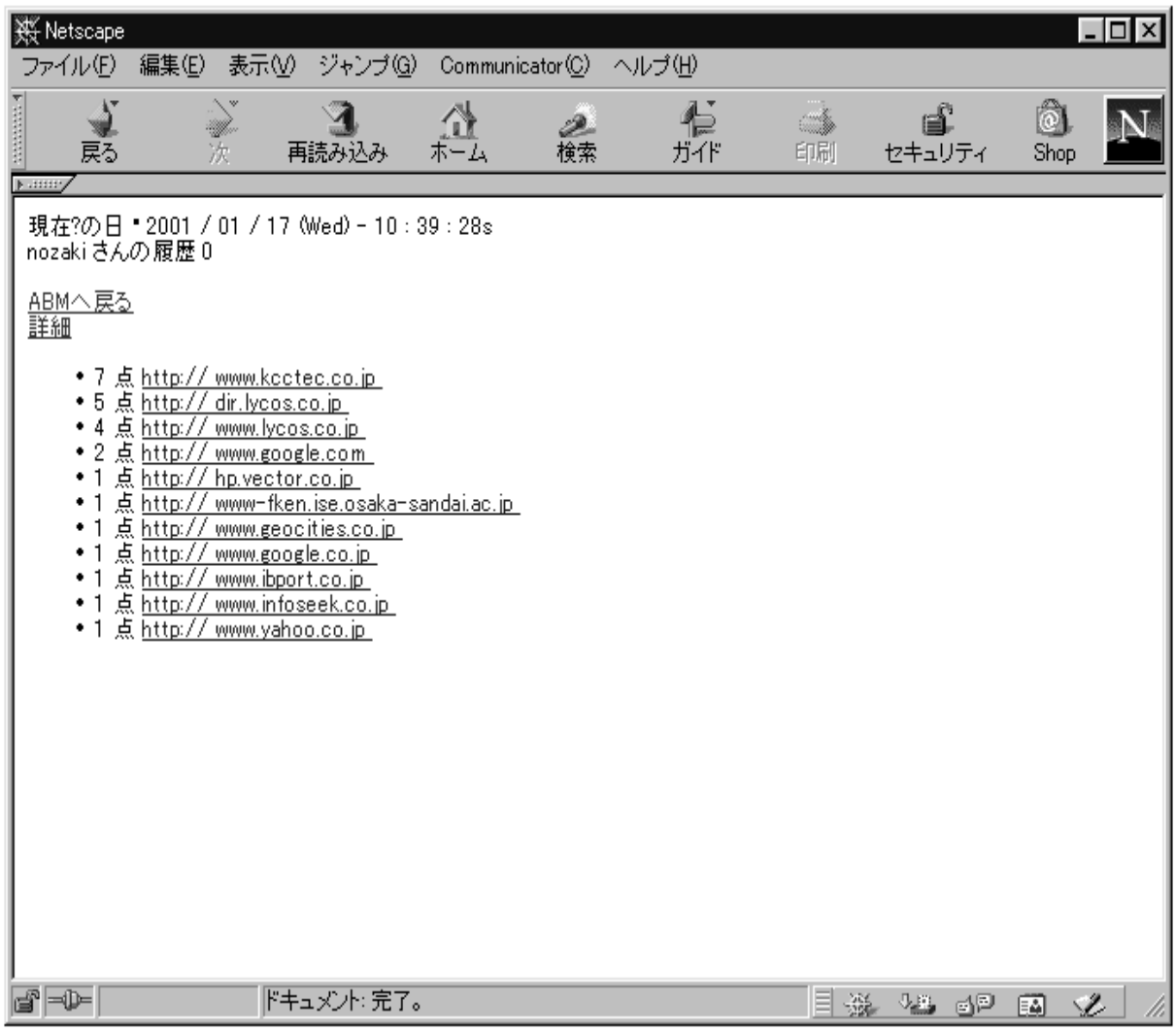

Netscape
ファイル(F) 編集(E) 表示(V) ジャンプ(G) Communicator(C) ヘルプ(H)
戻る
次
再読み込み
ホーム
検索
ガイド
印刷
セキュリティ
Shop
現在?の日 * 2001 / 01 / 17 (Wed) - 10 : 39 : 28s
nozaki さんの履歴 0
ABMへ戻る
詳細
• 7 点 http:// www.kcctec.co.jp
• 5 点 http:// dir.lycos.co.jp
• 4 点 http:// www.lycos.co.jp
• 2 点 http:// www.google.com
• 1 点 http:// hp.vector.co.jp
• 1 点 http:// www-fken.ise.osaka-sandai.ac.jp
• 1 点 http:// www.geocities.co.jp
• 1 点 http:// www.google.co.jp
• 1 点 http:// www.ibport.co.jp
• 1 点 http:// www.infoseek.co.jp
• 1 点 http:// www.yahoo.co.jp
ドキュメント: 完了。

図 3: 簡易版の履歴

3　有効な n 日分と n 日前の Web ページ閲覧履歴

3.1　有効な n 日分

従来のカレンダー通りの 3 日分などの Web ページ閲覧履歴では、数日間 Web ページを閲覧しないと Web ページ閲覧履歴が表示されなくなり、使えないときが在った。そこで、本研究の ERM(Eevey Recode Maker) は “有効な n 日間の履歴” と言うものを実装している。“有効な n 日分の履歴” とは、Web ページを閲覧したことの有る日を一回として、n 回分の Web ページ閲覧履歴を表示する。従来の 3 日分と、有効な 3 日分のとの比較を図 4に示す。(上記は ERM が使える環境が前提です。)

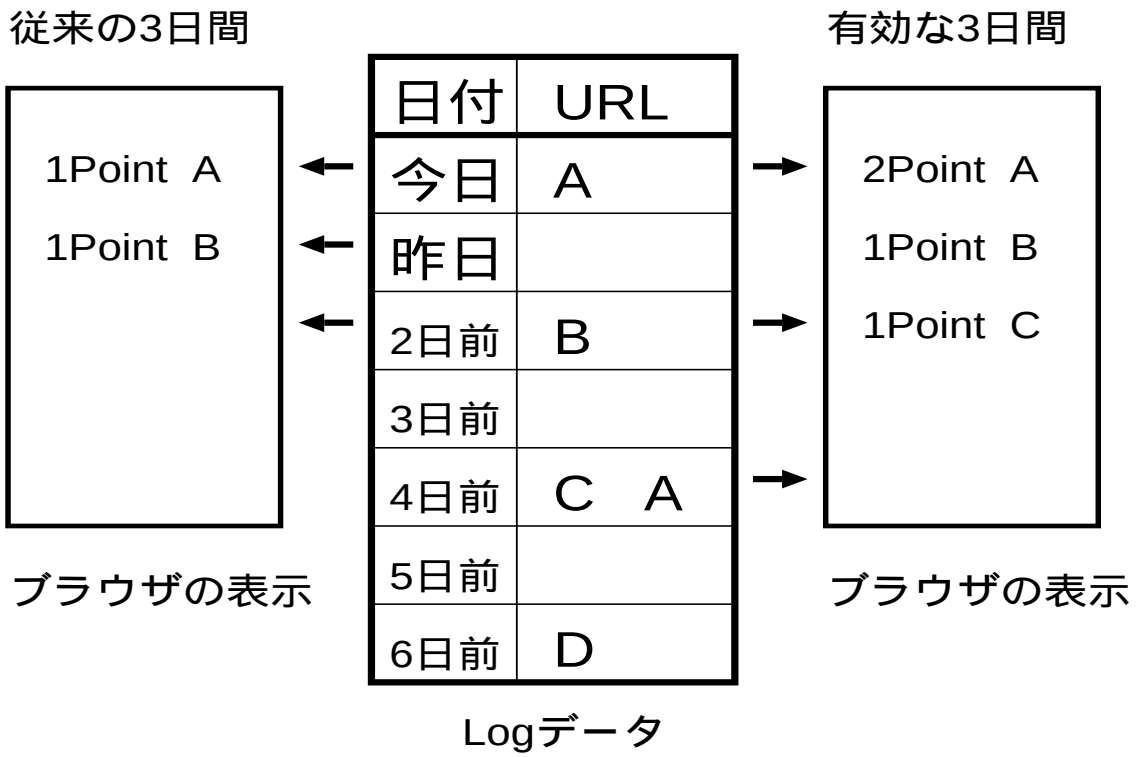


図 4: 従来の 3 日分と有効な 3 日分の履歴

3.2　有効な n 日前

従来の 1 日前、2 日前、3 日前などでは前回 Web ページを見た日の Web ページ閲覧履歴を表示するのに、何日前に閲覧していたかを覚えておく必要が有る。そこで、有効な n 日前の Web ページ閲覧履歴は、Web ページを閲覧していない日を除いた、n 日前の Web ページの閲覧履歴を表示する。

4　さまざまな履歴選択方法とカレンダー表示

本研究の ERM(Every Recode Maker) では Web 情報の検索・閲覧を支援するために、履歴選択表示に図5のようなカレンダー表示 (7x5) を採り入れ、このカレンダー表示によりユーザーは日付や曜日を視覚的に分かりやすくなっている。また、履歴選択には、有効な n 日分、有効な n 日前の履歴以外にも、各日の履歴、30 日間の履歴、曜日毎の履歴、過去 4 間で 1 週間づつの履歴、を選択することも出来る。

図 5: カレンダー表示

履歴表示はユーザが Web ページに訪れた回数を記録し、その回数をソート (回数の多い順) して表示している。これにより、ユーザーがよく行く Web ページは上位に表示される。

ERM は履歴選択にさまざまな選択方法を取り入れている。これは使うユーザや、その時々によって、記憶や考え方、使い方が違うからである。以下に各履歴選択の考察を示す。

4.1 各日毎の履歴

各日毎の履歴は、表示したい履歴の日付をあらかじめ分かっている場合に有効な履歴選択方法で、他の履歴選択方法では余分な日付けの履歴まで表示してしまう。また、履歴選択表示にはカレンダー表示を取り入れることにより、視覚的に分かりやすくなっている。

4.2 過去 30 日間の履歴

過去 30 日間の履歴とは、ユーザが過去 30 日間に閲覧した Web ページの閲覧履歴の事である。

4.3 過去 4 週間で 1 週間づつの履歴

過去 4 週間で 1 週間づつの履歴とは日曜日から土曜日までを 1 区切りとして、ユーザーの過去に訪れた Web ページの閲覧履歴を表示する。この機能は何曜日か記憶していないが、先週に見た Web ページなどを探すのに有効なプログラムである。

4.4 曜日毎の履歴

曜日毎の履歴とは、選択した曜日の過去 4 回分の Web ページ閲覧履歴を表示する。

これはログを解析した結果、同じ曜日には同じ Web ページを閲覧している事が多いことがわかった。このことはレポートなどで、有る程度決まった Web ページで調べているものと思われる。以上のことから、曜日毎によく行く Web ページはポイントが高くなり、ブラウザ上の上位によく行く Web ページが表示され、有効であることが分かる。

4.5 有効な数日分 (前) の履歴

有効な数日分 (前) の履歴とはユーザが Web ページを訪れたことの有る日を一回として、数日分 (前) の Web ページ閲覧履歴の事である。

この機能はユーザーが 2 回か 3 回前に Web 巡回した日に閲覧した Web ページを探したい時などに有効である。検索期間は 30 日間分のうちの数日分である。

4.6 各履歴選択の考察

ERM で履歴選択にさまざまな選択方法を取り入れている。これは使うユーザや、その時々によって、記憶や考え方、使い方が違うからである。

5　考察

本研究でさまざまな履歴選択方法で履歴を参照することが出来た。そして、得点化、ソートを取り入れることにより、上位によく行くページが表示されるようにもなった。しかし、その上位の全てが必ずしももう一度参照すると言うことは無い。また、ユーザを認識するために ident デーモンを利用したが、これは Windows 系の OS では標準で実装されていないのでインストールする必要があるなど、まだまだ改良の余地が有ることが分かる。

6　謝辞

本研究を進めていく上で御指導頂いた大垣 斉講師、藤井 信夫教授、中村 孝講師ならびに各研究室の方々に深く感謝の意を表します。

参考文献

[1] UNIX による計算機科学入門
共栄出版
久野 靖
1997 年
ISBN 4-621-04373-0

[2] AWK を 256 倍使うための本
株式会社アスキー
志村拓
1993 年
ISBN 4-7561-0162-3

[3] 大垣 斉 講師の ABM
http://www-fken.ise.osakasandai.ac.jp/ABM/

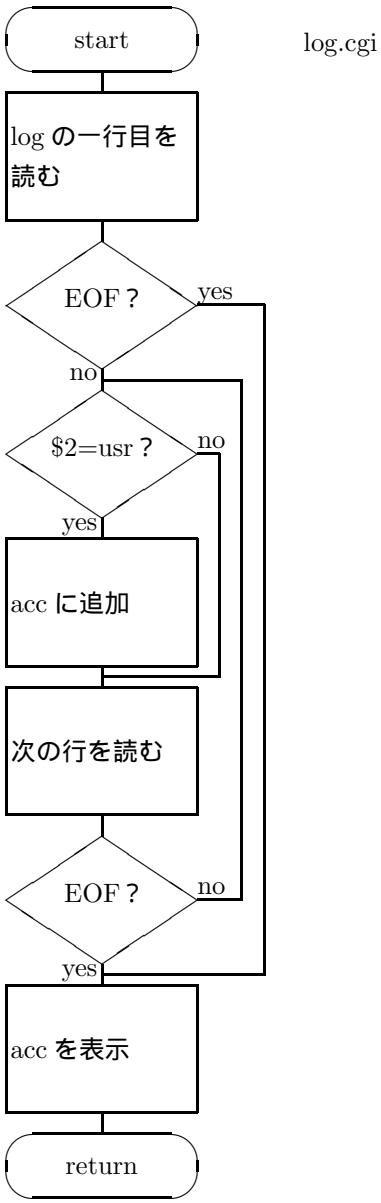
[4] 1999 年度卒業研究
高徳哲也の卒業論文
http://www-fken.ise.osakasandai.ac.jp/fken99g/42/H/

[5] 1999 年度卒業研究
宮北哲博の卒業論文
http://www-fken.ise.osakasandai.ac.jp/fken99g/41/H/

付録

log.cgi のフローチャート

機能：サーバーに有るログからユーザー別の Web ページ閲覧履歴を作成するプログラム。
環境：log には時間 (\$1)、ユーザー名 (\$2)、URL(\$3) が記載されている。
usr には環境変数の REMOTE_IDENT、acc は配列である。



log.cgi のプログラム

言語 : awk
機能 : サーバーに有るログからユーザー別の Web ページ閲覧履歴を作成するプログラム。

```
#!/bin/awk -f
BEGIN{
day=ENVIRON["QUERY_STRING"]
usr=ENVIRON["REMOTE_IDENT"]
ilog="/home/nozaki/public_html/log/log"day".dat"
print "Content-type: text/html\n\n"
print "<HTML>"
print strftime("現在?の日時-%Y / %m / %d (%a) - %H : %M : %Ss")
print "<BR>",usr,"さんの履歴",day,"<BR>"
print "<BR><A HREF=\"http://www.fken/%7Enozaki/abm/index.html\">ABM へ戻る</a><BR>"
print "<A HREF=http://www.fken/%7Enozaki/abm/lognk.cgi?"day"> 簡易</A><BR><UL>"
while(getline < ilog){
  if ($2==usr){
    acc[$3] += 1
  }
}
for (na in acc){
  printf ("<LI> %s 点 <A HREF= %s > %s </A><BR>\n",acc[na],na,na) | "sort +1nr"
}
print "</UL></HTML>"
}
```

log30n.cgi のプログラム

言語 ： awk
機能 ： 有効な n 日間の履歴。

```
#!/usr/bin/awk -f
BEGIN{
usr=ENVIRON["REMOTE_IDENT"]
day=ENVIRON["QUERY_STRING"]
print "Content-type: text/html\n"
print "<HTML>"
print strftime("%Y / %m / %d (%a) - %H : %M : %Ss")
print "<BR>",usr,"'s Recode<BR>"
b=0
for(i=0 ; i <= 30 ; i++){
  ilog="/home/nozaki/public_html/log/log"i".dat"
  if(b<=day){
    a=0
    while(getline < ilog){
      if($2==usr){
        a += 1
        break
      }
    }
    if(a>0){b += 1}
    if(b==day){
      close(ilog)
      while(getline < ilog){
        if($2==usr){
         acc[$3] += 1
        }
      }
      close (ilog)
      print i"日前,"
    }
  }
}
print "の履歴<BR>"
for (na in acc){
   printf ("%s Point<A HREF=%s > %s </A><BR>\n",acc[na],na,na) | "sort +0nr"
}
print "</HTML>"
}
```

week.cgi のプログラム

言語 ： awk
機能 ： カレンダー表示プログラム。

```
#!/bin/awk -f
BEGIN{

# 初期情報の設定
usr=ENVIRON["REMOTE_IDENT"]
year=strftime("%Y")
moon=strftime("%m")
today=strftime("%d")
week=strftime("%w")
split("31,28,31,30,31,30,31,31,30,31,30,31",M,",")
if( (year%4)==0 && ((year%100)!=0 || (year%400)==0)){ M[2]=29 }
M[0]=31
t=28+week
sun=21+week

# 曜日の設定
for(i=0;i<=6;i++){
  if(sun==21){sun=28}
  w[i]="<font size=5><a href=\"log4.cgi?"sun"\">"
  sun--
}
w[0]="<TD bgcolor=\"#ff9090\">" w[0] "日曜日</a></font></TD>"
w[1]="<TD bgcolor=\"#ffff90\">" w[1] "月曜日</a></font></TD>"
w[2]="<TD bgcolor=\"#ffff90\">" w[2] "火曜日</a></font></TD>"
w[3]="<TD bgcolor=\"#ffff90\">" w[3] "水曜日</a></font></TD>"
w[4]="<TD bgcolor=\"#ffff90\">" w[4] "木曜日</a></font></TD>"
w[5]="<TD bgcolor=\"#ffff90\">" w[5] "金曜日</a></font></TD>"
w[6]="<TD bgcolor=\"#90ffff\">" w[6] "土曜日</a></font></TD>"
color="<TD BGCOLOR=\"#90ff90\"><FONT SIZE=5>"
for(i=0 ; i <= week+28 ; i++){
  if(today==0){
    today = M[moon-1]
    color="<TD BGCOLOR=\"#ff90ff\"><FONT SIZE=5>"
  }
  D[i] = today
  if(today==1){D[i] = moon*1 "月" today"日"}
  CF[i] = color
  today--
}
```

```
for(i=0 ; i<=week ; i++){ D[i+29]=100 } # No Data
for(i=-1 ; i>=-6 ; i--){ D[i]=100 } # No Data
D[0]="今日"

# html 表示部分
print "Content-type: text/html\n"
print "<HTML><head><TITLE>カレンダー表示 (7x4) テスト</TITLE>"
print "<meta http-equiv=\"Content-Type\" content=\"text/html\">"
print "<META HTTP-EQUIV=\"Pragma\" CONTENT=\"no-cache\">"
print "<SCRIPT LANGUAGE=\"JavaScript\">"
print "<!-- "
print "function jump(sel) {
         if (sel.options[sel.selectedIndex].value) {
           location.href = sel.options[sel.selectedIndex].value;}} //-->"
print "</SCRIPT></head>"
print "<body bgcolor=\"white\">"
print strftime("現在?の日時-%Y / %m / %d (%a) - %H : %M : %Ss")
print "<BR><font size=3><A HREF=\"index.html\">戻る</a></font>"
print "<TABLE BORDER><TR>"
for(i=0 ; i<=6 ; i++){ print w[i] } # 曜日表示
print "<TD><font size=6><A HREF=\"logall.cgi\"> ALL(30day) </a></font></TD></TR>"

# カレンダー部分表示部分
for(i=1 ; i<=5 ; i++){
  print "<TR ALIGN=center>"
  for(j=1 ; j<=7 ; j++){
    if(D[t]==100){ print "<TD> No Data </TD>" }
    else{
      print CF[t]"<A HREF=\"logn.cgi?"t"\">"D[t]"</a></font></TD>"
    }
    --t
  }
  print "<TD><font size=6><A HREF=\"log7.cgi?"t+1"\"> ← </a></font></TD></TR>"
}
print "</TABLE>"

print "<BR><FORM><SELECT name=\"select\" onChange=\"jump(this);\">"
print "<OPTION SELECTED> 有効な n 日分の履歴 </OPTION>"
for(i=1 ; i<=27 ; i++){
  print "<OPTION value=\"log30x.cgi?"i"\">有効な"i"日分の履歴</OPTION>"
}
print "</SELECT><INPUT type=\"submit\" value=\"GO\"><FORM>"
```

```
print "<BR><A HREF=\"nday.html\">有効な n 日分の履歴の説明 </A><BR>"

print "<BR><FORM><SELECT name=\"select\" onChange=\"jump(this);\">"
print "<OPTION SELECTED> 有効な n 日前の履歴 </OPTION>"
for(i=1 ; i<=27 ; i++){
  print "<OPTION value=\"log30n.cgi?"i"\">有効な"i"日前の履歴</OPTION>"
}
print "</SELECT><INPUT type=\"submit\" value=\"GO\"><FORM>"
print "<BR><A HREF=\"nday.html\">有効な n 日前の履歴の説明 </A><BR>"

print "</body></HTML>"
}
```