Consumer Behavior Prediction through Web log Analysis

By Maeda Hanafi

hanafim1@owls.southernct.edu

Abstract

With the internet, businesses, such as Amazon, can predict the behavior of customers using log files of the web server. Log files contain web activity information such as: time, number of page views, number of server requests, number of repeated visits, access time on the page, IP address, etc. This information is passed back and forth from client (web browser) to a web server through http requests. By doing web analytics on this information, the marketer can predict the consumer behavior. By predicting consumer behavior, online businesses can remain competitive.

Introduction

Determining the consumer's behavior is essential to enhance an e-commerce website—to make more productive and profitable. Businesses must make decisions such as advertising to fulfill consumer demand. There are many ways to determine consumer behavior. One way is to use server log files. In this paper, I will demonstrate how to utilize a weblog server, based on Microsoft windows web server (Internet Information Server, IIS), and to analyze the behavior of the visitors or customers; for this purpose, I will use the weblog file from a real e-commerce website, thanks to Algosmith Computing allowing me to use the facility for this research.

Weblog files

This file is a text file containing default properties: Access date, time, client IP address, username, server IP address, server port, method, URL stem, URL query, protocol status, and user agent. This can be expanded by adding more properties such as service name, server name, bytes sent, bytes received, time taken, protocol version, post cookie, and referral; see figure 1 in Appendix.

Those properties are actually click streams generated by a web server that are simply series of pages requested by remote client computers.[1] Reading the click streams from the log file directly can be cumbersome; this may be extracted or imported into a database so that it is easier to analyze.

Here is one of the methods I used to extract the click streams information into the database; first, all of the delimiters in the log file are replaced by tabs. Then, create an empty table into which the data in the log file will be imported I used Microsoft SQL Server 8.0 in this case. This job is done through the import wizard in MS SQL Server; the completed imported log file can be seen in figure 2 in Appendix.

One important thing that has to be performed before importing is to keep the order or sequence of how every single row in the log file recorded, as this is considered as click streams. To do this, I create a column with auto-number data type.

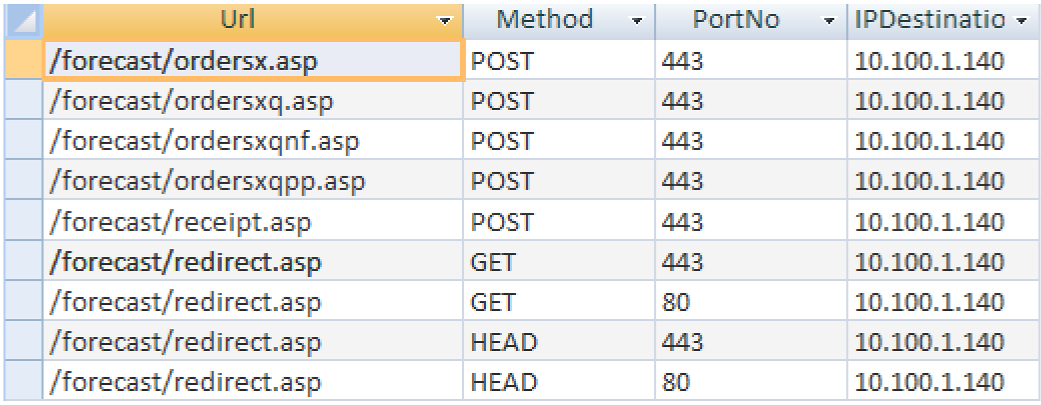
Once all the data is imported, clean up unnecessary information, such as in this case the User-Agent, and rename all the columns with appropriate names according to its property. In addition, to make the query faster, some fields are indexed. See figure 3 in Appendix.

Selected Case

As mentioned above, this paper will be based on a real ecommerce website. The real domain name will remain undisclosed for company privacy. Instead, throughout the paper, I will use www.xyz.com, which is not associated with any organization.

xyz.com has several different pages; the paths towards offering products for customers are only 6 files or paths, i.e.: 1) redirect.asp, 2) ordersx.asp, 3) ordersxq.asp, 4) ordersxqnf.asp,

5) ordersxqpp.asp, and 6) receipt.asp. See figure 1 below; for more list see Appendix figure 4.



Url	Method	PortNo	IPDestinatio
/forecast/ordersx.asp	POST	443	10.100.1.140
/forecast/ordersxq.asp	POST	443	10.100.1.140
/forecast/ordersxqnf.asp	POST	443	10.100.1.140
/forecast/ordersxqpp.asp	POST	443	10.100.1.140
/forecast/receipt.asp	POST	443	10.100.1.140
/forecast/redirect.asp	GET	443	10.100.1.140
/forecast/redirect.asp	GET	80	10.100.1.140
/forecast/redirect.asp	HEAD	443	10.100.1.140
/forecast/redirect.asp	HEAD	80	10.100.1.140

Figure 1. Paths which offer products

There are so many advertising pages preceding those six paths, excluded from the list in figure 1 above, which contain product offers. When certain products are clicked, it always hits path (1). The following possible simple paths of the customer clicks may be: 1) {1}; 2) {1, 2}; 3) {1, 3}; 4) {1, 4}; 5) {1, 5}; 6) {1, 2, 6}; 7) {1, 3, 6}; 8) {1, 4, 6}; and 9) {1, 5, 6}. Customers may generate more complex paths that satisfy the regular expression: 1[2345]* or 1[2345]$^{+}$6, which can be graphed in a finite automaton, as seen in figure 2 below. Any path that ends up with path (6) means the customer buys the product (s); others mean the customers don't buy.

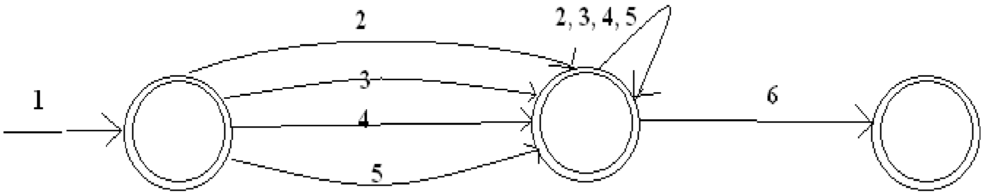
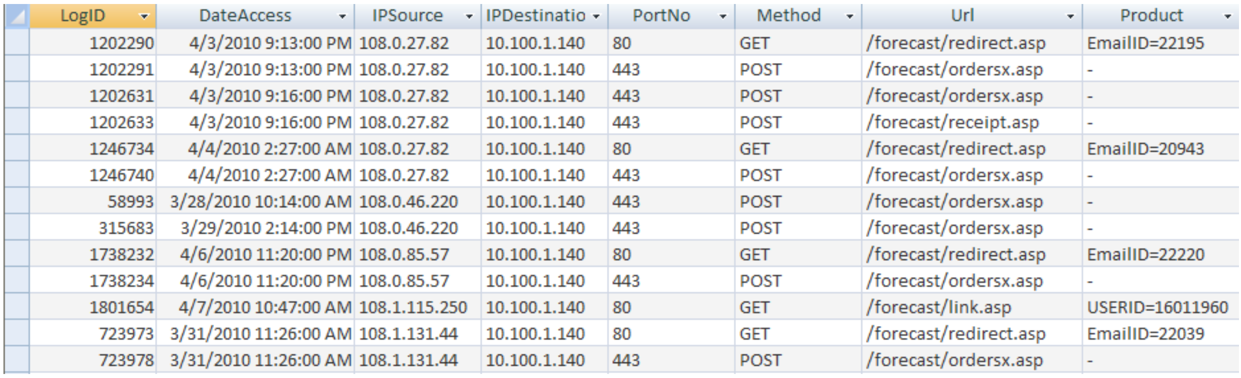


Figure 2. Paths Possibility

As seen in figure 3 below, any page uses POST method, the product is printed with sign '–' (dash). This is because the URL query is not recorded in the weblog file, but GET and HEAD methods.[2] By the absence of the correct values in the product field, it will be difficult or almost impossible to do query that requires to join on the product field. Therefore, the product fields have to be filled with the correct values.



LogID	DateAccess	IPSource	IPDestinatio	PortNo	Method	Url	Product
1202290	4/3/2010 9:13:00 PM	108.0.27.82	10.100.1.140	80	GET	/forecast/redirect.asp	EmailID=22195
1202291	4/3/2010 9:13:00 PM	108.0.27.82	10.100.1.140	443	POST	/forecast/ordersx.asp	-
1202631	4/3/2010 9:16:00 PM	108.0.27.82	10.100.1.140	443	POST	/forecast/ordersx.asp	-
1202633	4/3/2010 9:16:00 PM	108.0.27.82	10.100.1.140	443	POST	/forecast/receipt.asp	-
1246734	4/4/2010 2:27:00 AM	108.0.27.82	10.100.1.140	80	GET	/forecast/redirect.asp	EmailID=20943
1246740	4/4/2010 2:27:00 AM	108.0.27.82	10.100.1.140	443	POST	/forecast/ordersx.asp	-
58993	3/28/2010 10:14:00 AM	108.0.46.220	10.100.1.140	443	POST	/forecast/ordersx.asp	-
315683	3/29/2010 2:14:00 PM	108.0.46.220	10.100.1.140	443	POST	/forecast/ordersx.asp	-
1738232	4/6/2010 11:20:00 PM	108.0.85.57	10.100.1.140	80	GET	/forecast/redirect.asp	EmailID=22220
1738234	4/6/2010 11:20:00 PM	108.0.85.57	10.100.1.140	443	POST	/forecast/ordersx.asp	-
1801654	4/7/2010 10:47:00 AM	108.1.115.250	10.100.1.140	80	GET	/forecast/link.asp	USERID=16011960
723973	3/31/2010 11:26:00 AM	108.1.131.44	10.100.1.140	80	GET	/forecast/redirect.asp	EmailID=22039
723978	3/31/2010 11:26:00 AM	108.1.131.44	10.100.1.140	443	POST	/forecast/ordersx.asp	-

Figure 3. Imported Weblog file with tabs in Product field

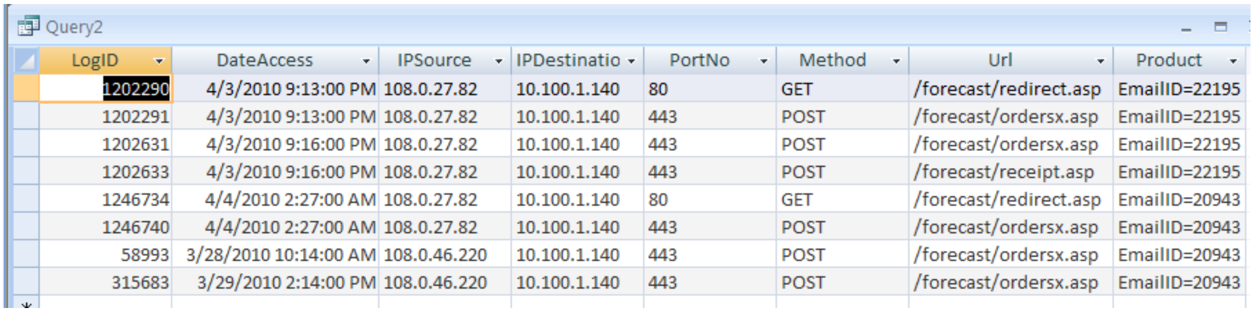
Now, the question might come is what value has to be set to the product field currently has the dash sign. By doing "SELECT" statement and "SORT" by IPSource and DateAccess, it will give me a clear clue that the correct value will be the same as the one when the given visitor hits the redirect path (path 1). This algorithm can be then translated into the SQL server script below[6]:

```
DECLARE @colA varchar(200)
DECLARE @colB varchar(200)
DECLARE @colC int
DECLARE @tmpCol varchar(200)
DECLARE @MyCursor CURSOR
SET @MyCursor = CURSOR FAST_FORWARD
FOR
        SELECT IPSource, Product, LogID From CleanWebLog ORDER BY IPSource, LogID
OPEN @MyCursor
FETCH NEXT FROM @MyCursor
INTO @ColA, @ColB, @colC
```

```
WHILE @@FETCH_STATUS = 0
BEGIN
        PRINT @ColA
        IF LEFT(@ColB,7) = 'EmailID'
        BEGIN
                SET @tmpCol = @ColB
                PRINT @tmpCol
        END
        ELSE IF @ColB = '-'
        BEGIN
                PRINT 'update with ' + @tmpCol
                UPDATE CleanWebLog SET Product=@tmpCol WHERE LogID=@colC
        END
        FETCH NEXT FROM @MyCursor
        INTO @ColA, @ColB, @colC
END
CLOSE @MyCursor
DEALLOCATE @MyCursor
```

Figure 4. SQL Server Script to Update Product Field

In order to run the script, the log file must be exported into MS SQL Server. Once the script has run completely, which basically is to replace dash signs in product field with the correct product. The table should look like the following:



LogID	DateAccess	IPSource	IPDestinatio	PortNo	Method	Url	Product
1202290	4/3/2010 9:13:00 PM	108.0.27.82	10.100.1.140	80	GET	/forecast/redirect.asp	EmailID=22195
1202291	4/3/2010 9:13:00 PM	108.0.27.82	10.100.1.140	443	POST	/forecast/ordersx.asp	EmailID=22195
1202631	4/3/2010 9:16:00 PM	108.0.27.82	10.100.1.140	443	POST	/forecast/ordersx.asp	EmailID=22195
1202633	4/3/2010 9:16:00 PM	108.0.27.82	10.100.1.140	443	POST	/forecast/receipt.asp	EmailID=22195
1246734	4/4/2010 2:27:00 AM	108.0.27.82	10.100.1.140	80	GET	/forecast/redirect.asp	EmailID=20943
1246740	4/4/2010 2:27:00 AM	108.0.27.82	10.100.1.140	443	POST	/forecast/ordersx.asp	EmailID=20943
58993	3/28/2010 10:14:00 AM	108.0.46.220	10.100.1.140	443	POST	/forecast/ordersx.asp	EmailID=20943
315683	3/29/2010 2:14:00 PM	108.0.46.220	10.100.1.140	443	POST	/forecast/ordersx.asp	EmailID=20943

Figure 5. Partial result of Updated Imported Weblog file

Port number 443 is commonly used for the page that comes with SSL (Secure Socket Layer) certificate to encrypt web communications;[3] port 80 is for any page without SSL.[4] So far, this information does not play important roles in this paper.

The IPSource field from the above figure is the IP address of the customer’s computer. Since this value is unique for every single computer[7], this can be used to represent a visitor or customer identification.

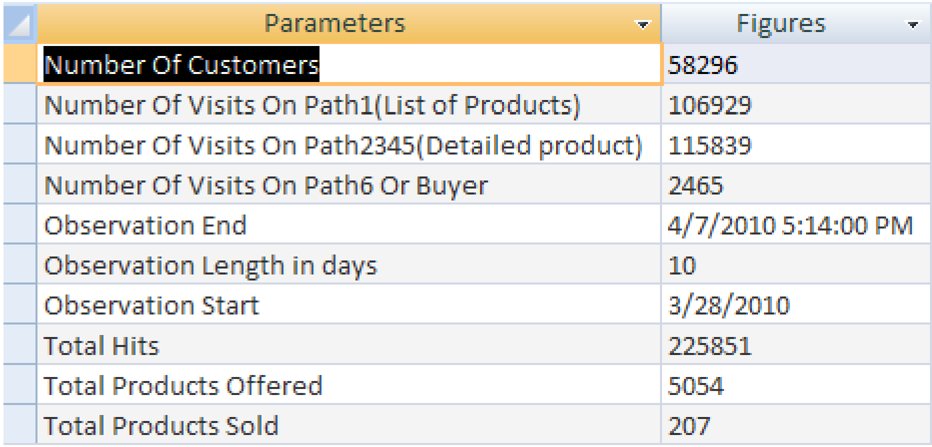
Analysis

In this section, a list of question and answer (Q/A list) will be presented to know how the visitors or consumers behave during visiting xyz.com. Before, here is the basic information which can be drawn from the log file:

```
SELECT "Number Of Customers" As [Parameters], COUNT(IPSource) AS [Figures]
FROM (SELECT DISTINCT IPSource
        FROM WebLogReorder);
UNION
SELECT "Total Hits" As [Parameters], COUNT(*) AS [Figures]
FROM WebLogReorder
WHERE Url LIKE '*redirect*' OR Url LIKE '*orders*' OR Url LIKE '*receipt*';
UNION
SELECT "Number Of Visits On Path1" As [Parameters], COUNT(Url) AS [Figures]
FROM WebLogReorder
WHERE Url LIKE '*redirect*';
UNION
SELECT "Number Of Visits On Path2345" As [Parameters], COUNT(Url) AS [Figures]
FROM WebLogReorder
WHERE Url LIKE '*ordersx*'
UNION
SELECT "Number Of Visits On Path6 Or Buyer" As [Parameters], COUNT(*) AS [Figures]
FROM (SELECT DISTINCT IPSource
        FROM WebLogReorder
        WHERE Url LIKE '*receipt*')
UNION
SELECT "Total Products Offered" As [Parameters], COUNT(*) AS [Figures]
FROM (SELECT DISTINCT Product
        FROM WebLogReorder)
UNION
SELECT "Total Products Sold" As [Parameters], COUNT(*) AS [Figures]
FROM (SELECT DISTINCT Product
        FROM WebLogReorder
        WHERE Url LIKE '*receipt*' )
UNION
SELECT "Observation Start" As [Parameters], MIN(DateAccess) AS   [Figures]
FROM WebLogReorder;
```

```
UNION
SELECT "Observation End" As [Parameters], MAX(DateAccess) AS   [Figures]
FROM WebLogReorder;
UNION
SELECT "Observation Length in days" As [Parameters],
DateDiff('d',MIN(DateAccess),MAX(DateAccess)) AS   [Figures]
FROM WebLogReorder;
```

The result of the query is shown below:



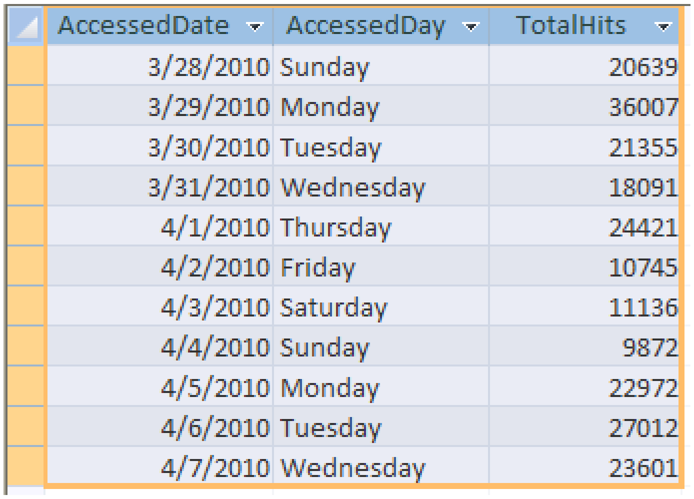
Parameters	Figures
Number Of Customers	58296
Number Of Visits On Path1(List of Products)	106929
Number Of Visits On Path2345(Detailed product)	115839
Number Of Visits On Path6 Or Buyer	2465
Observation End	4/7/2010 5:14:00 PM
Observation Length in days	10
Observation Start	3/28/2010
Total Hits	225851
Total Products Offered	5054
Total Products Sold	207

Figure 6. Basic Information

When we look at the number of visits at path 1, row 2 from figure above, is approximately twice of the number of customers, in row 1 above. This means, in average, the customer reviews the product twice. From row 3 and 2, path2345 and path1 respectively, it tells us that customers are interested to see details of the product. In row 3 and row 4, path2345 and path6 respectively, the number of visitors on path2345 is much greater than those who actually buy, which means that most visitors are simply interested in reviewing the detailed products but not buying them. Another proof of this is from looking at row 9 and row 10 where the number of products sold is a small percentage of the number of products offered.

The following query and the result set are to try to find the daily distribution trend of the number of customers.

```
SELECT DateValue(DateAccess) AS AccessedDate, WeekdayName(Weekday(DateAccess))
AS AccessedDay, COUNT(*) AS TotalHits
FROM WebLogReorder
WHERE Url LIKE '*redirect*' OR Url LIKE '*ordersx*' OR Url LIKE '*receipt*'
GROUP BY DateValue(DateAccess), WeekdayName(Weekday(DateAccess))
ORDER BY DateValue(DateAccess), COUNT(*) DESC;
```



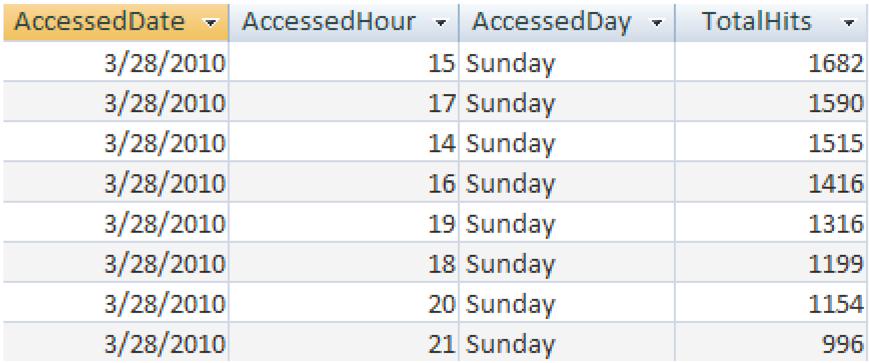
AccessedDate	AccessedDay	TotalHits
3/28/2010	Sunday	20639
3/29/2010	Monday	36007
3/30/2010	Tuesday	21355
3/31/2010	Wednesday	18091
4/1/2010	Thursday	24421
4/2/2010	Friday	10745
4/3/2010	Saturday	11136
4/4/2010	Sunday	9872
4/5/2010	Monday	22972
4/6/2010	Tuesday	27012
4/7/2010	Wednesday	23601

Figure 7. Number of Daily Hits Throughout Observation

From the figure, Monday on the first week of the observation results shows the highest hits, but on the second week it occurs on Tuesday. It seems very difficult to predict when customers are most likely to visit, at least we need more data, more than ten days, to observe.

For this reason, the analysis is furthered to the hourly level. The query and partial result set, including the chart with the whole result set, are shown in the following:

```
SELECT DateValue(DateAccess) AS AccessedDate, HOUR(DateAccess) AS AccessedHour,
WeekdayName(Weekday(DateAccess)) AS AccessedDay, COUNT(*) AS TotalHits
FROM WebLogReorder
WHERE Url LIKE '*redirect*' OR Url LIKE '*ordersx*' OR Url LIKE '*receipt*'
GROUP BY DateValue(DateAccess), HOUR(DateAccess),
WeekdayName(Weekday(DateAccess))
ORDER BY DateValue(DateAccess), COUNT(*) DESC;
```



AccessedDate	AccessedHour	AccessedDay	TotalHits
3/28/2010	15	Sunday	1682
3/28/2010	17	Sunday	1590
3/28/2010	14	Sunday	1515
3/28/2010	16	Sunday	1416
3/28/2010	19	Sunday	1316
3/28/2010	18	Sunday	1199
3/28/2010	20	Sunday	1154
3/28/2010	21	Sunday	996

Figure 8. Partial Result of Hourly Hit

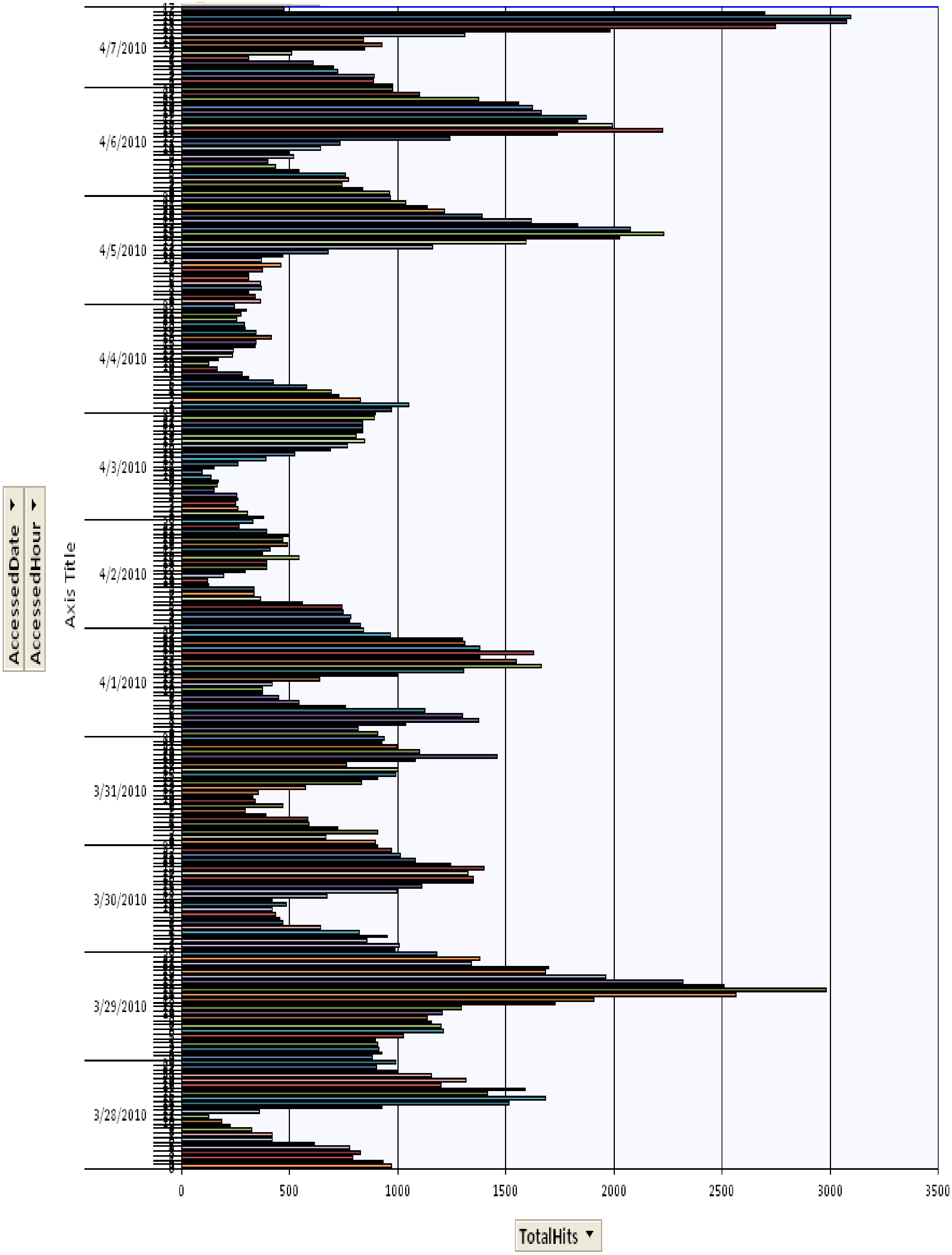


Figure 9. Chart of Hourly Hits

The x-axis is the total hits per hour, and y-axis shows the hourly access time. The figure shows that most visitors shops around late afternoon during the weekdays. But during the weekends, starting from Friday night, it occurs around midnight.

Following queries are to find the number of visitors who accessed paths 1 (viewing the product), {2, 3, 4, 5} (viewing the product in detail), and 6 (buying the product) -- to determine the visitor's behavior on each product.

On path 1: mngr_ProductViewOnPath1

```
SELECT Product, COUNT(*) AS TotalHitsOnPath1 FROM WebLogReorder
WHERE Url LIKE '*redirect*' GROUP BY Product;
```

On paths {2, 3, 4, 5}: mngr_ProductViewOnPath2345

```
SELECT Product, COUNT(*) AS TotalHitsOnPath2345 FROM WebLogReorder
WHERE Url LIKE '*ordersx*' GROUP BY Product;
```

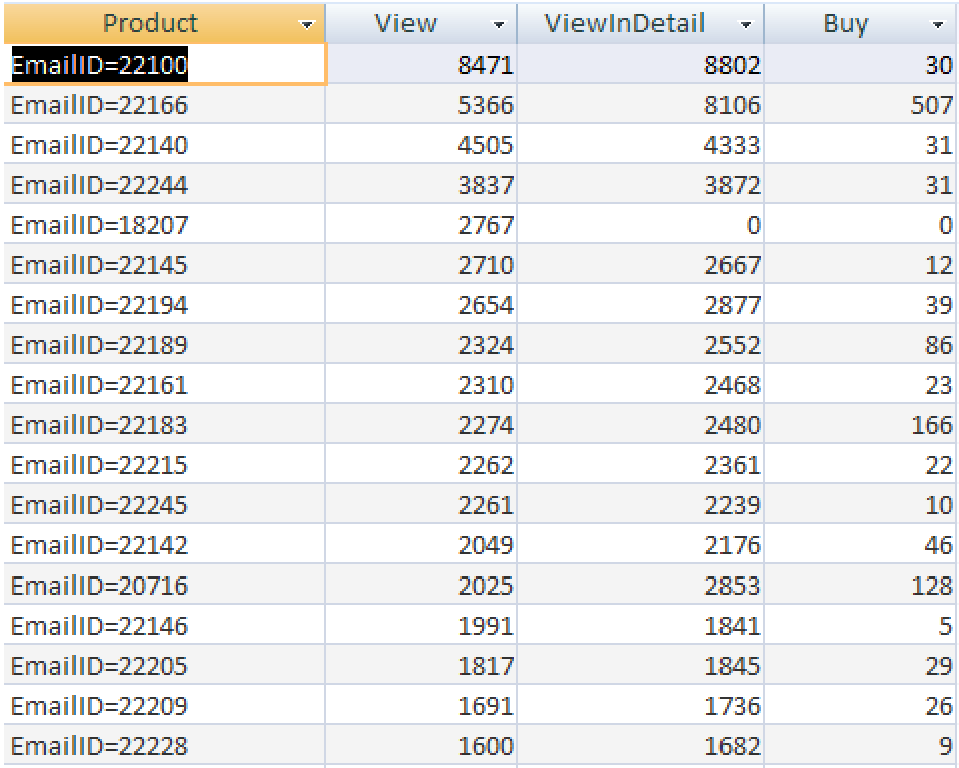
On path 6: mngr_ProductViewOnPath6

```
SELECT Product, COUNT(*) AS TotalHitsOnPath6 FROM WebLogReorder
WHERE Url LIKE '*receipt*' GROUP BY Product;
```

With those three queries from above, the behavior of the visitors against certain products can be determined by joining them together. There are three different queries that are basically all the same, but sorted differently to draw conclusions.

Version 1: mngr_PerformanceOfProductOffer1, sorted by View

```
SELECT A.Product, IIF(TotalHitsOnPath1>=0,TotalHitsOnPath1,0) AS [View],
IIF(TotalHitsOnPath2>=0,TotalHitsOnPath2,0) AS ViewInDetail,
IIF(TotalHitsOnPath3>=0,TotalHitsOnPath3,0) AS Buy
FROM (mngr_ProductViewOnPath1 AS A
        LEFT JOIN mngr_ProductViewOnPath2345 AS B
        ON A.Product = B.Product)
LEFT JOIN mngr_ProductViewOnPath6 AS C ON B.Product = C.Product
ORDER BY A.TotalHitsOnPath1 DESC;
```



Product	View	ViewInDetail	Buy
EmailID=22100	8471	8802	30
EmailID=22166	5366	8106	507
EmailID=22140	4505	4333	31
EmailID=22244	3837	3872	31
EmailID=18207	2767	0	0
EmailID=22145	2710	2667	12
EmailID=22194	2654	2877	39
EmailID=22189	2324	2552	86
EmailID=22161	2310	2468	23
EmailID=22183	2274	2480	166
EmailID=22215	2262	2361	22
EmailID=22245	2261	2239	10
EmailID=22142	2049	2176	46
EmailID=20716	2025	2853	128
EmailID=22146	1991	1841	5
EmailID=22205	1817	1845	29
EmailID=22209	1691	1736	26
EmailID=22228	1600	1682	9

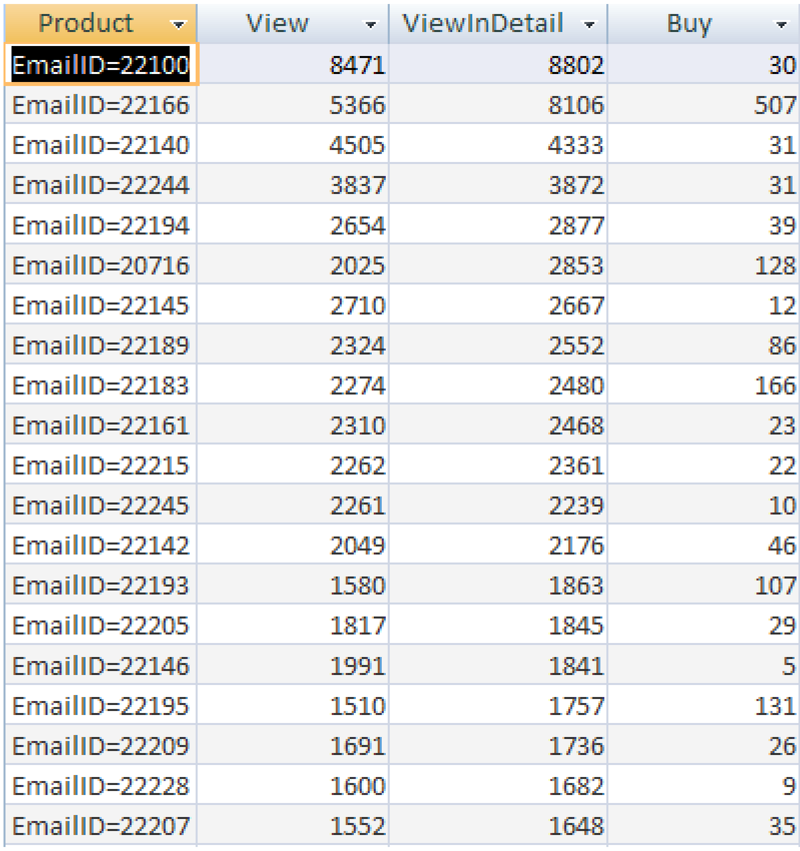
Figure 12. Partial Result of Product Views Sorted by the View Field

From the figure above, products with so many views means that they are highly encouraged to promote and have more links to it like product with EmailID = 22100. However, this certain product is bought by a small percentage of the viewers; these products should be replaced with other products with many views and a high percentage of viewers who actually buys it.

Version 2: mngr_PerformanceOfProductOffer2345, sorted by ViewInDetail

```
SELECT A.Product, IIF (TotalHitsOnPath1>=0, TotalHitsOnPath1,0) AS [View],
IIF (TotalHitsOnPath2345>=0, TotalHitsOnPath2345,0) AS ViewInDetail,
IIF (TotalHitsOnPath3>=0, TotalHitsOnPath3,0) AS Buy
FROM (mngr_ProductViewOnPath1 AS A
        LEFT JOIN mngr_ProductViewOnPath2345 AS B ON A.Product=B.Product)
LEFT JOIN mngr_ProductViewOnPath6 AS C ON C.Product=B.Product
ORDER BY TotalHitsOnPath2345 DESC;
```

The resulting table shows the products that attracted customers. The partial result is shown below.



Product	View	ViewInDetail	Buy
EmailID=22100	8471	8802	30
EmailID=22166	5366	8106	507
EmailID=22140	4505	4333	31
EmailID=22244	3837	3872	31
EmailID=22194	2654	2877	39
EmailID=20716	2025	2853	128
EmailID=22145	2710	2667	12
EmailID=22189	2324	2552	86
EmailID=22183	2274	2480	166
EmailID=22161	2310	2468	23
EmailID=22215	2262	2361	22
EmailID=22245	2261	2239	10
EmailID=22142	2049	2176	46
EmailID=22193	1580	1863	107
EmailID=22205	1817	1845	29
EmailID=22146	1991	1841	5
EmailID=22195	1510	1757	131
EmailID=22209	1691	1736	26
EmailID=22228	1600	1682	9
EmailID=22207	1552	1648	35

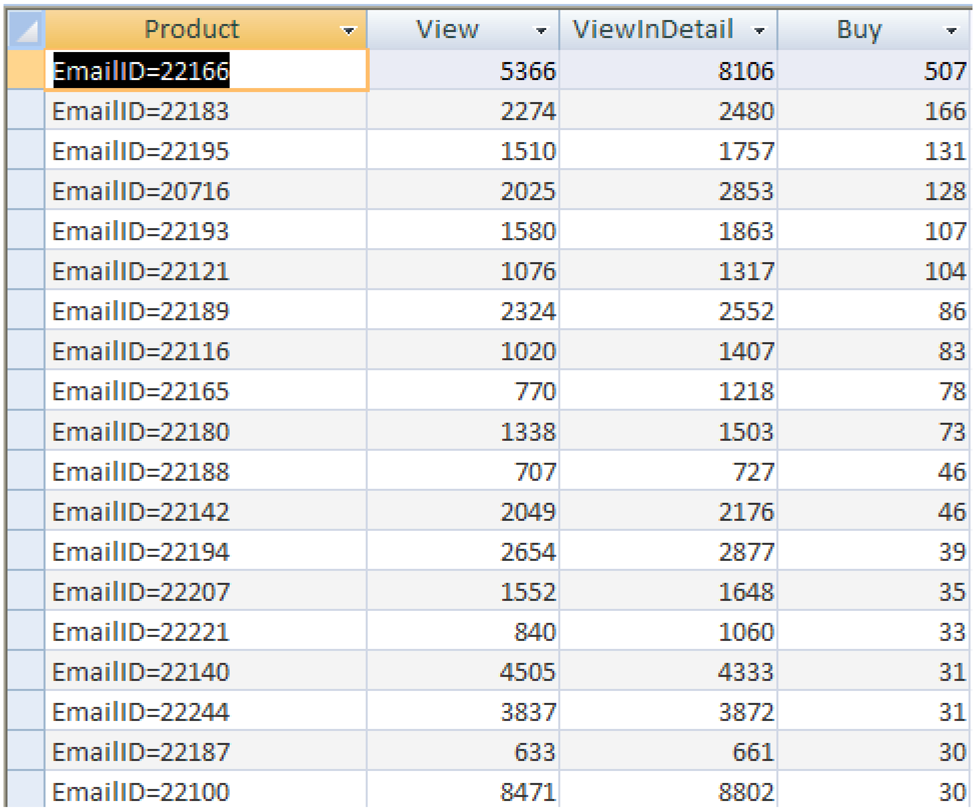
Figure 13. Partial Result of Product Views Sorted by the ViewInDetail Field

From the above figure, some of the products have a high number of visitors and reviewers, but these products don't have a many customers such as product with EmailID=22100 or EmailID=22140. One way to increase visitor's interests is to change the details of the products to get the likelihood of the visitors to buy it.

Version 3: mngr_PerformanceOfProductOffer6, sorted by Buy

```
SELECT A.Product, IIF (TotalHitsOnPath1>=0, TotalHitsOnPath1,0) AS [View],
IIF (TotalHitsOnPath2345>=0, TotalHitsOnPath2345, 0) AS ViewInDetail,
IIF (TotalHitsOnPath6>=0, TotalHitsOnPath6, 0) AS Buy
FROM (mngr_ProductViewOnPath1 AS A
        LEFT JOIN mngr_ProductViewOnPath2345 AS B ON A.Product=B.Product)
LEFT JOIN mngr_ProductViewOnPath6 AS C ON C.Product=B.Product
ORDER BY TotalHitsOnPath6 DESC;
```

The resulting table shows the number of customers buying the product. The partial result is shown below.



Product	View	ViewInDetail	Buy
EmailID=22166	5366	8106	507
EmailID=22183	2274	2480	166
EmailID=22195	1510	1757	131
EmailID=20716	2025	2853	128
EmailID=22193	1580	1863	107
EmailID=22121	1076	1317	104
EmailID=22189	2324	2552	86
EmailID=22116	1020	1407	83
EmailID=22165	770	1218	78
EmailID=22180	1338	1503	73
EmailID=22188	707	727	46
EmailID=22142	2049	2176	46
EmailID=22194	2654	2877	39
EmailID=22207	1552	1648	35
EmailID=22221	840	1060	33
EmailID=22140	4505	4333	31
EmailID=22244	3837	3872	31
EmailID=22187	633	661	30
EmailID=22100	8471	8802	30

Figure 14. Partial Result of Product Views Sorted by the Buy Field

From the figure above, the customer behavior is very interesting. One thing, the many customers view certain product, for example EmailID=22100, but it generates low sales. On the other thing, such as EmailID=22166, it lower number of hits but high sales. One of the interpretations that may be drawn here is the visitor at first, path 1, is very attractive. Once he or she reaches the details, path {2, 3, 4, 5}, the product detail information is poor.

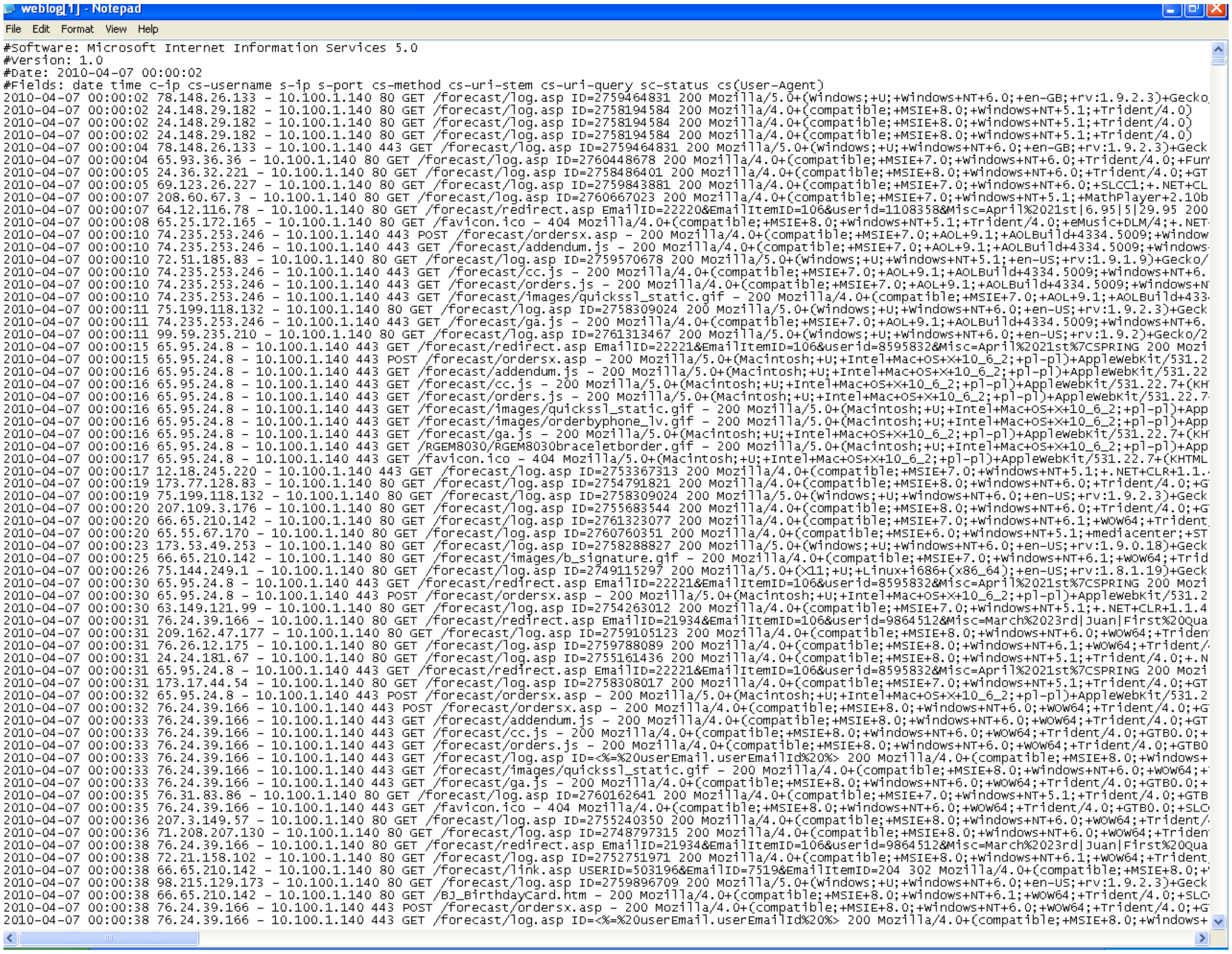
Summary

The consumer behavior mentioned in the analysis section above produces some very important clues that can help marketing people. This research can be expanded more and generate more comprehensive conclusion with longer time span of observation—the longer the observation time, the more data in the log file we get.

Reference

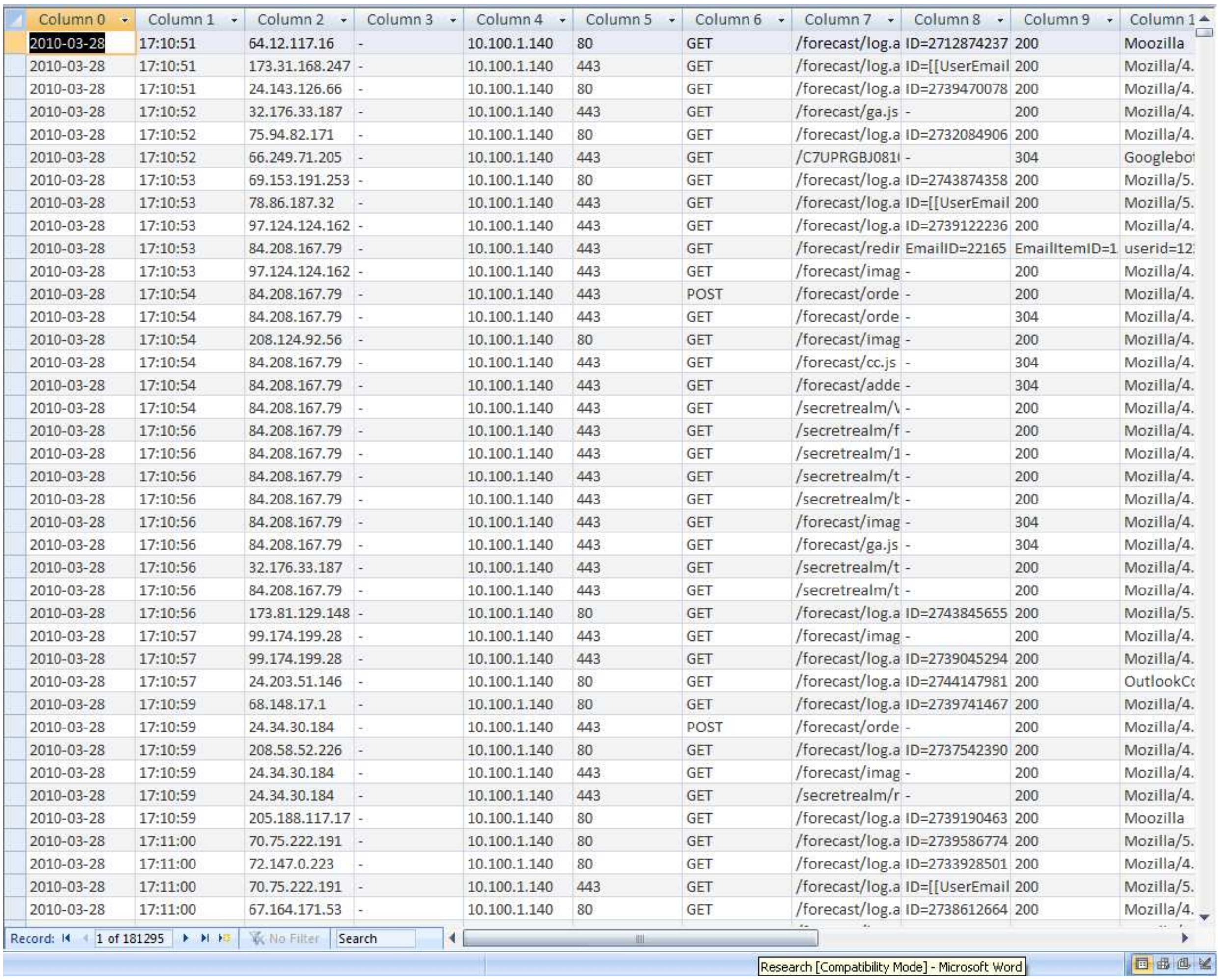
[1] http://www.rhsmith.umd.edu/faculty/wmoe/MoeFader Evolving Visits JIM 2004.pdf
[2] http://www.w3.org/Protocols/rfc2616/rfc2616-sec9.html
[3] http://tools.ietf.org/html/rfc2660
[4] http://tools.ietf.org/html/rfc2616
[5] Introduction to Computer Theory, by Daniel L.A. Cohen
[6] Coding Tips: Programmer's Quick References, by Imam Hanafi
[7] http://tools.ietf.org/html/rfc760

Appendix



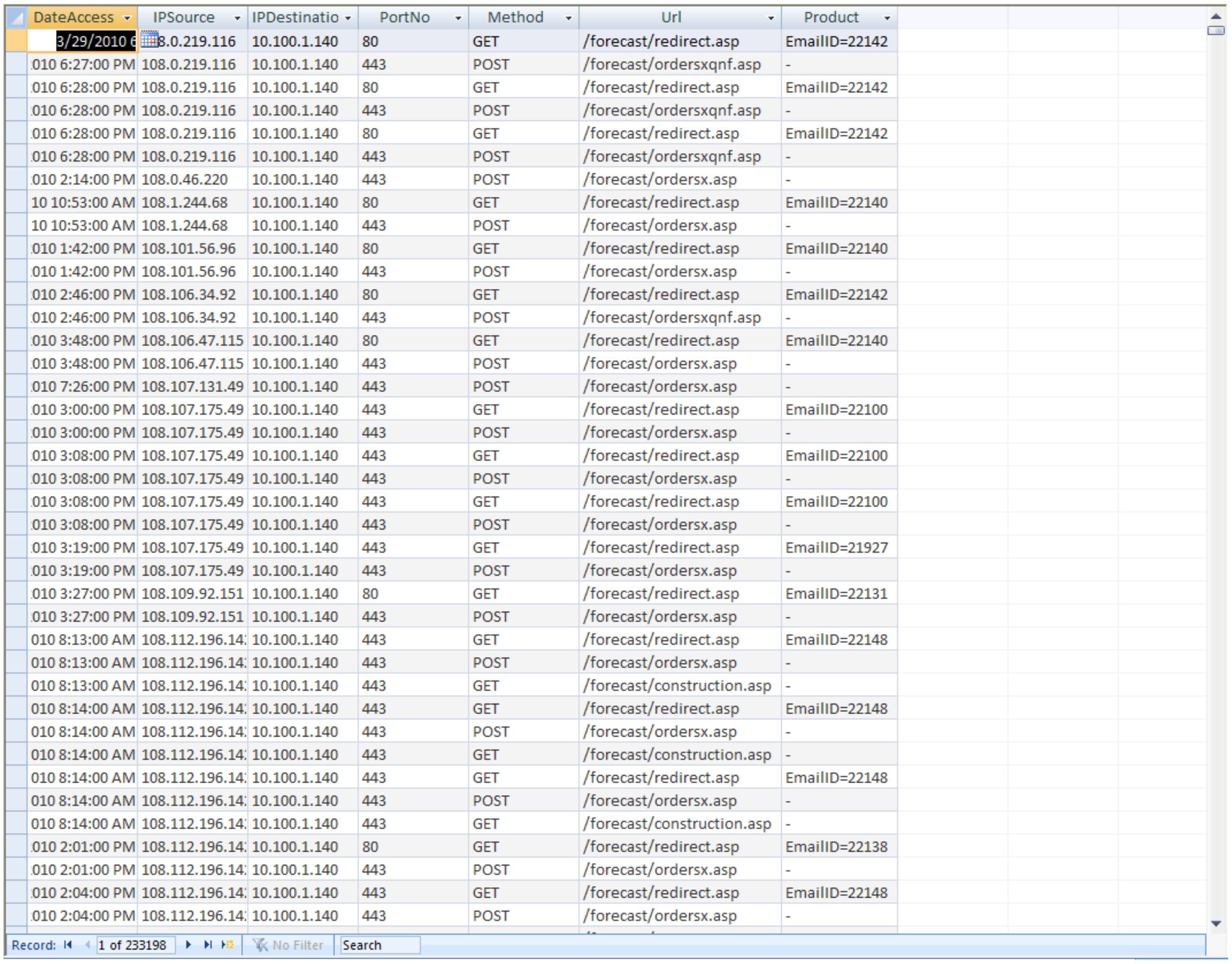
```
#Software: Microsoft Internet Information Services 5.0
#Version: 1.0
#Date: 2010-04-07 00:00:02
#Fields: date time c-ip cs-username s-ip s-port cs-method cs-uri-stem cs-uri-query sc-status cs(User-Agent)
2010-04-07 00:00:02 78.148.26.133 - 10.100.1.140 80 GET /forecast/log.asp ID=2759464831 200 Mozilla/5.0+(Windows;+U;+Windows+NT+6.0;+en-GB;+rv:1.9.2.3)+Gecko
2010-04-07 00:00:02 24.148.29.182 - 10.100.1.140 80 GET /forecast/log.asp ID=2758194584 200 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+5.1;+Trident/4.0)
2010-04-07 00:00:02 24.148.29.182 - 10.100.1.140 80 GET /forecast/log.asp ID=2758194584 200 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+5.1;+Trident/4.0)
2010-04-07 00:00:02 24.148.29.182 - 10.100.1.140 80 GET /forecast/log.asp ID=2758194584 200 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+5.1;+Trident/4.0)
2010-04-07 00:00:04 78.148.26.133 - 10.100.1.140 443 GET /forecast/log.asp ID=2759464831 200 Mozilla/5.0+(Windows;+U;+Windows+NT+6.0;+en-GB;+rv:1.9.2.3)+Geck
2010-04-07 00:00:04 65.93.36.36 - 10.100.1.140 80 GET /forecast/log.asp ID=2760448678 200 Mozilla/4.0+(compatible;+MSIE+7.0;+Windows+NT+6.0;+Trident/4.0;+Fun
2010-04-07 00:00:05 24.36.32.221 - 10.100.1.140 80 GET /forecast/log.asp ID=2758486401 200 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+6.0;+Trident/4.0;+GT
2010-04-07 00:00:05 69.123.26.227 - 10.100.1.140 80 GET /forecast/log.asp ID=2759843881 200 Mozilla/4.0+(compatible;+MSIE+7.0;+Windows+NT+6.0;+SLCC1;+.NET+CL
2010-04-07 00:00:07 208.60.67.3 - 10.100.1.140 80 GET /forecast/log.asp ID=2760667023 200 Mozilla/4.0+(compatible;+MSIE+7.0;+Windows+NT+5.1;+MathPlayer+2.10b
2010-04-07 00:00:07 64.12.116.78 - 10.100.1.140 80 GET /forecast/redirect.asp EmailID=22220&EmailItemID=106&userid=1108358&Misc=April%2021st|6.95|5|29.95 200
2010-04-07 00:00:08 65.25.172.165 - 10.100.1.140 80 GET /favicon.ico - 404 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+5.1;+Trident/4.0;+eMusic+DLM/4;+.NET
2010-04-07 00:00:10 74.235.253.246 - 10.100.1.140 443 POST /forecast/ordersx.asp - 200 Mozilla/4.0+(compatible;+MSIE+7.0;+AOL+9.1;+AOLBuild+4334.5009;+Window
2010-04-07 00:00:10 74.235.253.246 - 10.100.1.140 443 GET /forecast/addendum.js - 200 Mozilla/4.0+(compatible;+MSIE+7.0;+AOL+9.1;+AOLBuild+4334.5009;+Windows
2010-04-07 00:00:10 72.51.185.83 - 10.100.1.140 80 GET /forecast/log.asp ID=2759570678 200 Mozilla/5.0+(Windows;+U;+Windows+NT+5.1;+en-US;+rv:1.9.1.9)+Gecko/
2010-04-07 00:00:10 74.235.253.246 - 10.100.1.140 443 GET /forecast/cc.js - 200 Mozilla/4.0+(compatible;+MSIE+7.0;+AOL+9.1;+AOLBuild+4334.5009;+Windows+NT+6.
2010-04-07 00:00:10 74.235.253.246 - 10.100.1.140 443 GET /forecast/orders.js - 200 Mozilla/4.0+(compatible;+MSIE+7.0;+AOL+9.1;+AOLBuild+4334.5009;+Windows+N
2010-04-07 00:00:10 74.235.253.246 - 10.100.1.140 443 GET /forecast/images/quickssl_static.gif - 200 Mozilla/4.0+(compatible;+MSIE+7.0;+AOL+9.1;+AOLBuild+433
2010-04-07 00:00:11 75.199.118.132 - 10.100.1.140 80 GET /forecast/log.asp ID=2758309024 200 Mozilla/5.0+(Windows;+U;+Windows+NT+6.0;+en-US;+rv:1.9.2.3)+Geck
2010-04-07 00:00:11 74.235.253.246 - 10.100.1.140 443 GET /forecast/ga.js - 200 Mozilla/4.0+(compatible;+MSIE+7.0;+AOL+9.1;+AOLBuild+4334.5009;+Windows+NT+6.
2010-04-07 00:00:11 99.59.235.210 - 10.100.1.140 80 GET /forecast/log.asp ID=2761313467 200 Mozilla/5.0+(Windows;+U;+Windows+NT+6.0;+en-US;+rv:1.9.2)+Gecko/2
2010-04-07 00:00:15 65.95.24.8 - 10.100.1.140 443 GET /forecast/redirect.asp EmailID=22221&EmailItemID=106&userid=8595832&Misc=April%2021st%7CSPRING 200 Mozi
2010-04-07 00:00:15 65.95.24.8 - 10.100.1.140 443 POST /forecast/ordersx.asp - 200 Mozilla/5.0+(Macintosh;+U;+Intel+Mac+OS+X+10_6_2;+pl-pl)+AppleWebKit/531.2
2010-04-07 00:00:16 65.95.24.8 - 10.100.1.140 443 GET /forecast/addendum.js - 200 Mozilla/5.0+(Macintosh;+U;+Intel+Mac+OS+X+10_6_2;+pl-pl)+AppleWebKit/531.22
2010-04-07 00:00:16 65.95.24.8 - 10.100.1.140 443 GET /forecast/cc.js - 200 Mozilla/5.0+(Macintosh;+U;+Intel+Mac+OS+X+10_6_2;+pl-pl)+AppleWebKit/531.22.7+(KH
2010-04-07 00:00:16 65.95.24.8 - 10.100.1.140 443 GET /forecast/orders.js - 200 Mozilla/5.0+(Macintosh;+U;+Intel+Mac+OS+X+10_6_2;+pl-pl)+AppleWebKit/531.22.7
2010-04-07 00:00:16 65.95.24.8 - 10.100.1.140 443 GET /forecast/images/quickssl_static.gif - 200 Mozilla/5.0+(Macintosh;+U;+Intel+Mac+OS+X+10_6_2;+pl-pl)+App
2010-04-07 00:00:16 65.95.24.8 - 10.100.1.140 443 GET /forecast/images/orderbyphone_lv.gif - 200 Mozilla/5.0+(Macintosh;+U;+Intel+Mac+OS+X+10_6_2;+pl-pl)+App
2010-04-07 00:00:16 65.95.24.8 - 10.100.1.140 443 GET /forecast/ga.js - 200 Mozilla/5.0+(Macintosh;+U;+Intel+Mac+OS+X+10_6_2;+pl-pl)+AppleWebKit/531.22.7+(KH
2010-04-07 00:00:16 65.95.24.8 - 10.100.1.140 443 GET /RGEM8030/RGEM8030braceletborder.gif - 200 Mozilla/5.0+(Macintosh;+U;+Intel+Mac+OS+X+10_6_2;+pl-pl)+App
2010-04-07 00:00:17 65.95.24.8 - 10.100.1.140 443 GET /favicon.ico - 404 Mozilla/5.0+(Macintosh;+U;+Intel+Mac+OS+X+10_6_2;+pl-pl)+AppleWebKit/531.22.7+(KHTML
2010-04-07 00:00:17 12.18.245.220 - 10.100.1.140 443 GET /forecast/log.asp ID=2753367313 200 Mozilla/4.0+(compatible;+MSIE+7.0;+Windows+NT+5.1;+.NET+CLR+1.1.
2010-04-07 00:00:19 173.77.128.83 - 10.100.1.140 80 GET /forecast/log.asp ID=2754791821 200 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+6.0;+Trident/4.0;+G
2010-04-07 00:00:19 75.199.118.132 - 10.100.1.140 80 GET /forecast/log.asp ID=2758309024 200 Mozilla/5.0+(Windows;+U;+Windows+NT+6.0;+en-US;+rv:1.9.2.3)+Geck
2010-04-07 00:00:20 207.109.3.176 - 10.100.1.140 80 GET /forecast/log.asp ID=2755683544 200 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+6.0;+Trident/4.0;+G
2010-04-07 00:00:20 66.65.210.142 - 10.100.1.140 80 GET /forecast/log.asp ID=2761323077 200 Mozilla/4.0+(compatible;+MSIE+7.0;+Windows+NT+6.1;+WOW64;+Trident
2010-04-07 00:00:20 65.55.67.170 - 10.100.1.140 80 GET /forecast/log.asp ID=2760760351 200 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1;+mediacenter;+ST
2010-04-07 00:00:23 173.53.49.253 - 10.100.1.140 80 GET /forecast/log.asp ID=2758288827 200 Mozilla/5.0+(Windows;+U;+Windows+NT+6.0;+en-US;+rv:1.9.0.18)+Geck
2010-04-07 00:00:25 66.65.210.142 - 10.100.1.140 80 GET /forecast/images/b_signature.gif - 200 Mozilla/4.0+(compatible;+MSIE+7.0;+Windows+NT+6.1;+WOW64;+Trid
2010-04-07 00:00:26 75.144.249.1 - 10.100.1.140 80 GET /forecast/log.asp ID=249115297 200 Mozilla/5.0+(X11;+U;+Linux+i686+(x86_64);+en-US;+rv:1.8.1.19)+Geck
2010-04-07 00:00:30 65.95.24.8 - 10.100.1.140 443 GET /forecast/redirect.asp EmailID=22221&EmailItemID=106&userid=8595832&Misc=April%2021st%7CSPRING 200 Mozi
2010-04-07 00:00:30 65.95.24.8 - 10.100.1.140 443 POST /forecast/ordersx.asp - 200 Mozilla/5.0+(Macintosh;+U;+Intel+Mac+OS+X+10_6_2;+pl-pl)+AppleWebKit/531.2
2010-04-07 00:00:30 63.149.121.99 - 10.100.1.140 80 GET /forecast/log.asp ID=2754263012 200 Mozilla/4.0+(compatible;+MSIE+7.0;+Windows+NT+5.1;+.NET+CLR+1.1.4
2010-04-07 00:00:31 76.24.39.166 - 10.100.1.140 80 GET /forecast/redirect.asp EmailID=21934&EmailItemID=106&userid=9864512&Misc=March%2023rd|Juan|First%20Qua
2010-04-07 00:00:31 209.162.47.177 - 10.100.1.140 80 GET /forecast/log.asp ID=2759105123 200 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+6.0;+WOW64;+Triden
2010-04-07 00:00:31 76.26.12.175 - 10.100.1.140 80 GET /forecast/log.asp ID=2759788089 200 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+6.1;+WOW64;+Trident/
2010-04-07 00:00:31 24.24.181.67 - 10.100.1.140 80 GET /forecast/log.asp ID=2755161436 200 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+5.1;+Trident/4.0;+.N
2010-04-07 00:00:31 65.95.24.8 - 10.100.1.140 443 GET /forecast/redirect.asp EmailID=22221&EmailItemID=106&userid=8595832&Misc=April%2021st%7CSPRING 200 Mozi
2010-04-07 00:00:31 173.17.44.54 - 10.100.1.140 80 GET /forecast/log.asp ID=2758308017 200 Mozilla/4.0+(compatible;+MSIE+7.0;+Windows+NT+5.1;+Trident/4.0;+GT
2010-04-07 00:00:32 65.95.24.8 - 10.100.1.140 443 POST /forecast/ordersx.asp - 200 Mozilla/5.0+(Macintosh;+U;+Intel+Mac+OS+X+10_6_2;+pl-pl)+AppleWebKit/531.2
2010-04-07 00:00:32 76.24.39.166 - 10.100.1.140 443 POST /forecast/ordersx.asp - 200 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+6.0;+WOW64;+Trident/4.0;+G
2010-04-07 00:00:33 76.24.39.166 - 10.100.1.140 443 GET /forecast/addendum.js - 200 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+6.0;+WOW64;+Trident/4.0;+GT
2010-04-07 00:00:33 76.24.39.166 - 10.100.1.140 443 GET /forecast/cc.js - 200 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+6.0;+WOW64;+Trident/4.0;+GTB0.0;+
2010-04-07 00:00:33 76.24.39.166 - 10.100.1.140 443 GET /forecast/orders.js - 200 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+6.0;+WOW64;+Trident/4.0;+GTB0
2010-04-07 00:00:33 76.24.39.166 - 10.100.1.140 443 GET /forecast/log.asp ID=<%=%20userEmail.userEmailId%20%> 200 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+
2010-04-07 00:00:33 76.24.39.166 - 10.100.1.140 443 GET /forecast/images/quickssl_static.gif - 200 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+6.0;+WOW64;+
2010-04-07 00:00:33 76.24.39.166 - 10.100.1.140 443 GET /forecast/ga.js - 200 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+6.0;+WOW64;+Trident/4.0;+GTB0.0;+
2010-04-07 00:00:35 76.31.83.86 - 10.100.1.140 80 GET /forecast/log.asp ID=2760162641 200 Mozilla/4.0+(compatible;+MSIE+7.0;+Windows+NT+5.1;+Trident/4.0;+GTB
2010-04-07 00:00:35 76.24.39.166 - 10.100.1.140 443 GET /favicon.ico - 404 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+6.0;+WOW64;+Trident/4.0;+GTB0.0;+SLC
2010-04-07 00:00:36 207.3.149.57 - 10.100.1.140 80 GET /forecast/log.asp ID=2755240350 200 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+6.0;+WOW64;+Trident/
2010-04-07 00:00:36 71.208.207.130 - 10.100.1.140 80 GET /forecast/log.asp ID=2748797315 200 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+6.0;+WOW64;+Triden
2010-04-07 00:00:38 76.24.39.166 - 10.100.1.140 80 GET /forecast/redirect.asp EmailID=21934&EmailItemID=106&userid=9864512&Misc=March%2023rd|Juan|First%20Qua
2010-04-07 00:00:38 72.21.158.102 - 10.100.1.140 80 GET /forecast/log.asp ID=2752751971 200 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+6.1;+WOW64;+Trident
2010-04-07 00:00:38 66.65.210.142 - 10.100.1.140 80 GET /forecast/link.asp USERID=503196&EmailID=7519&EmailItemID=204 302 Mozilla/4.0+(compatible;+MSIE+8.0;+
2010-04-07 00:00:38 98.215.129.173 - 10.100.1.140 80 GET /forecast/log.asp ID=2759896709 200 Mozilla/5.0+(Windows;+U;+Windows+NT+6.0;+en-US;+rv:1.9.2.3)+Geck
2010-04-07 00:00:38 66.65.210.142 - 10.100.1.140 80 GET /BJ_BirthdayCard.htm - 200 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+6.1;+WOW64;+Trident/4.0;+SLC
2010-04-07 00:00:38 76.24.39.166 - 10.100.1.140 443 POST /forecast/ordersx.asp - 200 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+NT+6.0;+WOW64;+Trident/4.0;+G
2010-04-07 00:00:38 76.24.39.166 - 10.100.1.140 443 GET /forecast/log.asp ID=<%=%20userEmail.userEmailId%20%> 200 Mozilla/4.0+(compatible;+MSIE+8.0;+Windows+
```

Figure 1. Actual weblog file



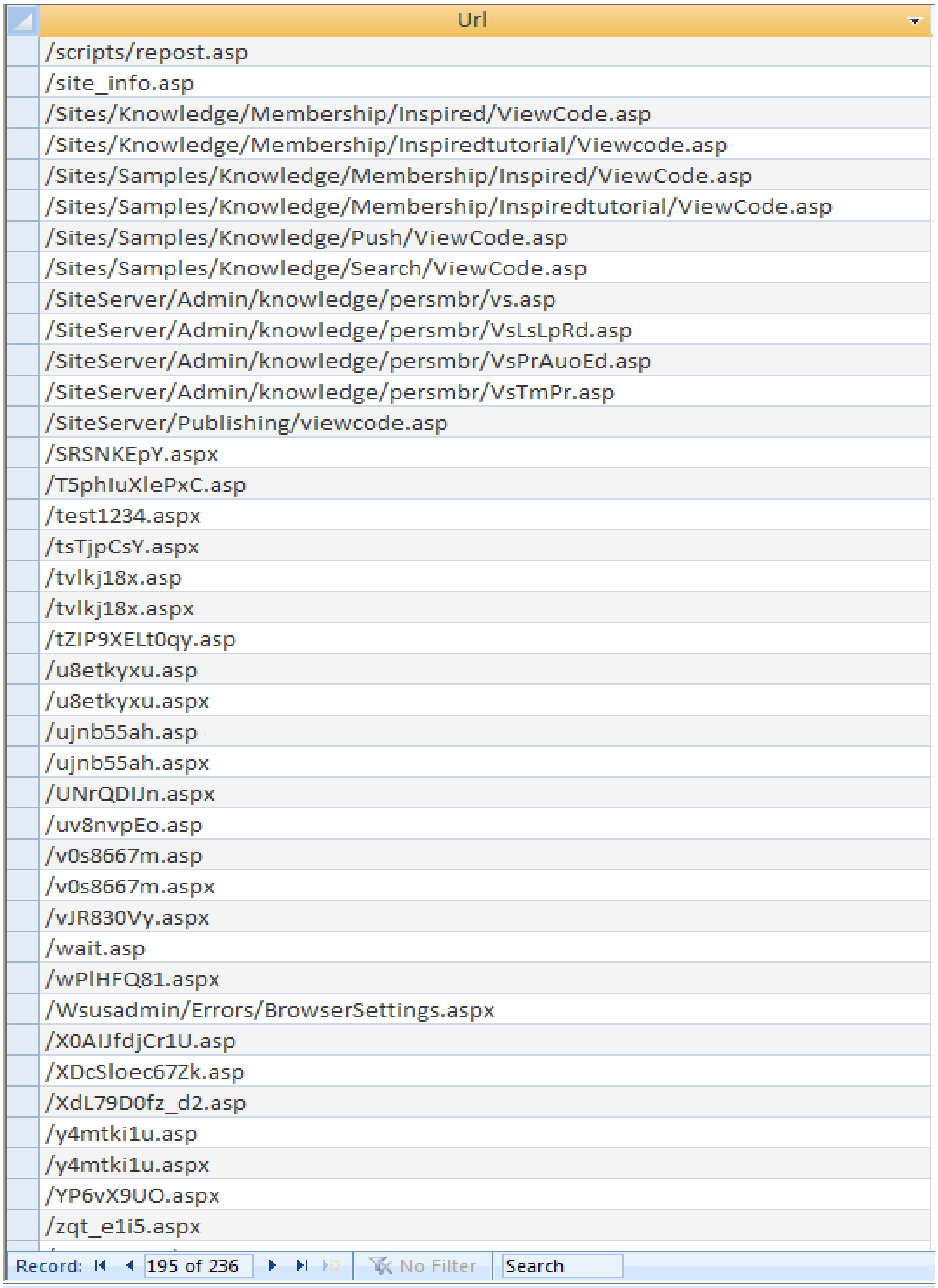
Column 0	Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9	Column 1
2010-03-28	17:10:51	64.12.117.16	-	10.100.1.140	80	GET	/forecast/log.a	ID=2712874237	200	Moozilla
2010-03-28	17:10:51	173.31.168.247	-	10.100.1.140	443	GET	/forecast/log.a	ID=[[UserEmail	200	Mozilla/4.
2010-03-28	17:10:51	24.143.126.66	-	10.100.1.140	80	GET	/forecast/log.a	ID=2739470078	200	Mozilla/4.
2010-03-28	17:10:52	32.176.33.187	-	10.100.1.140	443	GET	/forecast/ga.js	-	200	Mozilla/4.
2010-03-28	17:10:52	75.94.82.171	-	10.100.1.140	80	GET	/forecast/log.a	ID=2732084906	200	Mozilla/4.
2010-03-28	17:10:52	66.249.71.205	-	10.100.1.140	443	GET	/C7UPRGBJ081(	-	304	Googlebot
2010-03-28	17:10:53	69.153.191.253	-	10.100.1.140	80	GET	/forecast/log.a	ID=2743874358	200	Mozilla/5.
2010-03-28	17:10:53	78.86.187.32	-	10.100.1.140	443	GET	/forecast/log.a	ID=[[UserEmail	200	Mozilla/5.
2010-03-28	17:10:53	97.124.124.162	-	10.100.1.140	443	GET	/forecast/log.a	ID=2739122236	200	Mozilla/4.
2010-03-28	17:10:53	84.208.167.79	-	10.100.1.140	443	GET	/forecast/redir	EmailID=22165	EmailItemID=1	userid=12
2010-03-28	17:10:53	97.124.124.162	-	10.100.1.140	443	GET	/forecast/imag	-	200	Mozilla/4.
2010-03-28	17:10:54	84.208.167.79	-	10.100.1.140	443	POST	/forecast/orde	-	200	Mozilla/4.
2010-03-28	17:10:54	84.208.167.79	-	10.100.1.140	443	GET	/forecast/orde	-	304	Mozilla/4.
2010-03-28	17:10:54	208.124.92.56	-	10.100.1.140	80	GET	/forecast/imag	-	200	Mozilla/4.
2010-03-28	17:10:54	84.208.167.79	-	10.100.1.140	443	GET	/forecast/cc.js	-	304	Mozilla/4.
2010-03-28	17:10:54	84.208.167.79	-	10.100.1.140	443	GET	/forecast/adde	-	304	Mozilla/4.
2010-03-28	17:10:54	84.208.167.79	-	10.100.1.140	443	GET	/secretrealm/\	-	200	Mozilla/4.
2010-03-28	17:10:56	84.208.167.79	-	10.100.1.140	443	GET	/secretrealm/f	-	200	Mozilla/4.
2010-03-28	17:10:56	84.208.167.79	-	10.100.1.140	443	GET	/secretrealm/1	-	200	Mozilla/4.
2010-03-28	17:10:56	84.208.167.79	-	10.100.1.140	443	GET	/secretrealm/t	-	200	Mozilla/4.
2010-03-28	17:10:56	84.208.167.79	-	10.100.1.140	443	GET	/secretrealm/t	-	200	Mozilla/4.
2010-03-28	17:10:56	84.208.167.79	-	10.100.1.140	443	GET	/forecast/imag	-	304	Mozilla/4.
2010-03-28	17:10:56	84.208.167.79	-	10.100.1.140	443	GET	/forecast/ga.js	-	304	Mozilla/4.
2010-03-28	17:10:56	32.176.33.187	-	10.100.1.140	443	GET	/secretrealm/t	-	200	Mozilla/4.
2010-03-28	17:10:56	84.208.167.79	-	10.100.1.140	443	GET	/secretrealm/t	-	200	Mozilla/4.
2010-03-28	17:10:56	173.81.129.148	-	10.100.1.140	80	GET	/forecast/log.a	ID=2743845655	200	Mozilla/5.
2010-03-28	17:10:57	99.174.199.28	-	10.100.1.140	443	GET	/forecast/imag	-	200	Mozilla/4.
2010-03-28	17:10:57	99.174.199.28	-	10.100.1.140	443	GET	/forecast/log.a	ID=2739045294	200	Mozilla/4.
2010-03-28	17:10:57	24.203.51.146	-	10.100.1.140	80	GET	/forecast/log.a	ID=2744147981	200	OutlookCc
2010-03-28	17:10:59	68.148.17.1	-	10.100.1.140	80	GET	/forecast/log.a	ID=2739741467	200	Mozilla/4.
2010-03-28	17:10:59	24.34.30.184	-	10.100.1.140	443	POST	/forecast/orde	-	200	Mozilla/4.
2010-03-28	17:10:59	208.58.52.226	-	10.100.1.140	80	GET	/forecast/log.a	ID=2737542390	200	Mozilla/4.
2010-03-28	17:10:59	24.34.30.184	-	10.100.1.140	443	GET	/forecast/imag	-	200	Mozilla/4.
2010-03-28	17:10:59	24.34.30.184	-	10.100.1.140	443	GET	/secretrealm/r	-	200	Mozilla/4.
2010-03-28	17:10:59	205.188.117.17	-	10.100.1.140	80	GET	/forecast/log.a	ID=2739190463	200	Moozilla
2010-03-28	17:11:00	70.75.222.191	-	10.100.1.140	80	GET	/forecast/log.a	ID=2739586774	200	Mozilla/5.
2010-03-28	17:11:00	72.147.0.223	-	10.100.1.140	80	GET	/forecast/log.a	ID=2733928501	200	Mozilla/4.
2010-03-28	17:11:00	70.75.222.191	-	10.100.1.140	443	GET	/forecast/log.a	ID=[[UserEmail	200	Mozilla/5.
2010-03-28	17:11:00	67.164.171.53	-	10.100.1.140	80	GET	/forecast/log.a	ID=2738612664	200	Mozilla/4.

Figure 2. Imported log file



DateAccess	IPSource	IPDestinatio	PortNo	Method	Url	Product
3/29/2010 6	8.0.219.116	10.100.1.140	80	GET	/forecast/redirect.asp	EmailID=22142
010 6:27:00 PM	108.0.219.116	10.100.1.140	443	POST	/forecast/ordersxqnf.asp	-
010 6:28:00 PM	108.0.219.116	10.100.1.140	80	GET	/forecast/redirect.asp	EmailID=22142
010 6:28:00 PM	108.0.219.116	10.100.1.140	443	POST	/forecast/ordersxqnf.asp	-
010 6:28:00 PM	108.0.219.116	10.100.1.140	80	GET	/forecast/redirect.asp	EmailID=22142
010 6:28:00 PM	108.0.219.116	10.100.1.140	443	POST	/forecast/ordersxqnf.asp	-
010 2:14:00 PM	108.0.46.220	10.100.1.140	443	POST	/forecast/ordersx.asp	-
10 10:53:00 AM	108.1.244.68	10.100.1.140	80	GET	/forecast/redirect.asp	EmailID=22140
10 10:53:00 AM	108.1.244.68	10.100.1.140	443	POST	/forecast/ordersx.asp	-
010 1:42:00 PM	108.101.56.96	10.100.1.140	80	GET	/forecast/redirect.asp	EmailID=22140
010 1:42:00 PM	108.101.56.96	10.100.1.140	443	POST	/forecast/ordersx.asp	-
010 2:46:00 PM	108.106.34.92	10.100.1.140	80	GET	/forecast/redirect.asp	EmailID=22142
010 2:46:00 PM	108.106.34.92	10.100.1.140	443	POST	/forecast/ordersxqnf.asp	-
010 3:48:00 PM	108.106.47.115	10.100.1.140	80	GET	/forecast/redirect.asp	EmailID=22140
010 3:48:00 PM	108.106.47.115	10.100.1.140	443	POST	/forecast/ordersx.asp	-
010 7:26:00 PM	108.107.131.49	10.100.1.140	443	POST	/forecast/ordersx.asp	-
010 3:00:00 PM	108.107.175.49	10.100.1.140	443	GET	/forecast/redirect.asp	EmailID=22100
010 3:00:00 PM	108.107.175.49	10.100.1.140	443	POST	/forecast/ordersx.asp	-
010 3:08:00 PM	108.107.175.49	10.100.1.140	443	GET	/forecast/redirect.asp	EmailID=22100
010 3:08:00 PM	108.107.175.49	10.100.1.140	443	POST	/forecast/ordersx.asp	-
010 3:08:00 PM	108.107.175.49	10.100.1.140	443	GET	/forecast/redirect.asp	EmailID=22100
010 3:08:00 PM	108.107.175.49	10.100.1.140	443	POST	/forecast/ordersx.asp	-
010 3:19:00 PM	108.107.175.49	10.100.1.140	443	GET	/forecast/redirect.asp	EmailID=21927
010 3:19:00 PM	108.107.175.49	10.100.1.140	443	POST	/forecast/ordersx.asp	-
010 3:27:00 PM	108.109.92.151	10.100.1.140	80	GET	/forecast/redirect.asp	EmailID=22131
010 3:27:00 PM	108.109.92.151	10.100.1.140	443	POST	/forecast/ordersx.asp	-
010 8:13:00 AM	108.112.196.14	10.100.1.140	443	GET	/forecast/redirect.asp	EmailID=22148
010 8:13:00 AM	108.112.196.14	10.100.1.140	443	POST	/forecast/ordersx.asp	-
010 8:13:00 AM	108.112.196.14	10.100.1.140	443	GET	/forecast/construction.asp	-
010 8:14:00 AM	108.112.196.14	10.100.1.140	443	GET	/forecast/redirect.asp	EmailID=22148
010 8:14:00 AM	108.112.196.14	10.100.1.140	443	POST	/forecast/ordersx.asp	-
010 8:14:00 AM	108.112.196.14	10.100.1.140	443	GET	/forecast/construction.asp	-
010 8:14:00 AM	108.112.196.14	10.100.1.140	443	GET	/forecast/redirect.asp	EmailID=22148
010 8:14:00 AM	108.112.196.14	10.100.1.140	443	POST	/forecast/ordersx.asp	-
010 8:14:00 AM	108.112.196.14	10.100.1.140	443	GET	/forecast/construction.asp	-
010 2:01:00 PM	108.112.196.14	10.100.1.140	80	GET	/forecast/redirect.asp	EmailID=22138
010 2:01:00 PM	108.112.196.14	10.100.1.140	443	POST	/forecast/ordersx.asp	-
010 2:04:00 PM	108.112.196.14	10.100.1.140	443	GET	/forecast/redirect.asp	EmailID=22148
010 2:04:00 PM	108.112.196.14	10.100.1.140	443	POST	/forecast/ordersx.asp	-

Figure 3. Clean version of the imported log file



Url
/scripts/repost.asp
/site_info.asp
/Sites/Knowledge/Membership/Inspired/ViewCode.asp
/Sites/Knowledge/Membership/Inspiredtutorial/Viewcode.asp
/Sites/Samples/Knowledge/Membership/Inspired/ViewCode.asp
/Sites/Samples/Knowledge/Membership/Inspiredtutorial/ViewCode.asp
/Sites/Samples/Knowledge/Push/ViewCode.asp
/Sites/Samples/Knowledge/Search/ViewCode.asp
/SiteServer/Admin/knowledge/persmbr/vs.asp
/SiteServer/Admin/knowledge/persmbr/VsLsLpRd.asp
/SiteServer/Admin/knowledge/persmbr/VsPrAuoEd.asp
/SiteServer/Admin/knowledge/persmbr/VsTmPr.asp
/SiteServer/Publishing/viewcode.asp
/SRSNKEpY.aspx
/T5phIuXlePxC.asp
/test1234.aspx
/tsTjpCsY.aspx
/tvlkj18x.asp
/tvlkj18x.aspx
/tZIP9XELt0qy.asp
/u8etkyxu.asp
/u8etkyxu.aspx
/ujnb55ah.asp
/ujnb55ah.aspx
/UNrQDIJn.aspx
/uv8nvpEo.asp
/v0s8667m.asp
/v0s8667m.aspx
/vJR830Vy.aspx
/wait.asp
/wPlHFQ81.aspx
/Wsusadmin/Errors/BrowserSettings.aspx
/X0AIJfdjCr1U.asp
/XDcSloec67Zk.asp
/XdL79D0fz_d2.asp
/y4mtki1u.asp
/y4mtki1u.aspx
/YP6vX9UO.aspx
/zqt_e1i5.aspx

Record: 195 of 236

Figure 4 Partial List of URL