Chapter 6

IMPLEMENTATION

6.1. Case Study

Current marketing developments must outline a way to perk up or enhance the current marketing plans. Analyze the current marketing strategies, research the company and the history of the project to gain a thorough understanding of the situation and strategies that have been employed.

Analyze the audience. Which company is trying to reach audience? What are the ages, occupations, geographic locations, social status, economic status and perceptions of their company audience? You can come to these conclusions by studying the company's current marketing materials and events, or by asking a company representative directly.

This Case study is to present the user behaviors on popular websites to the hardware manufacturing company.

ABC Hardware Company planning to increases the hardware production as a part of Quarter 3 planning. Analysis community started mining the web with various vendors; all the vendors/web hosting companies are providing the dump of data which contains information related to all the transactions (i.e. out of 100% of dump data, 20% belongs to Hardware Company and in that 10% of data matches to the ABC hardware company related information). Analysis community efforts are ineffectively utilized in finding 10% of data information from 100% of dump data.
Analysis community is a 20 member’s team and on average the payment of each team member is $1000 so that the total cost of the team per a business day is $160,000.

Team is spending 24 hours of efforts for analyzing 10% of data (not for finding the information) out of 100% dump. Team is spending 240 hours for analyzing 100% of dump.

As per the previous description and past history out of 240 hours, only 24 hours are effectively used because 100% dump contains only 10% of data, which is valuable to ABC Company.

Since log dump contains unrelated and unwanted data along with the relevant information, to find out the relevant information team has to dig into entire dump received from external vendors/web hosting companies. While analyzing data, 80% of data found to be unrelated, 10% of data relative data and 10% is useful information. Company is loosing 80% of efforts to filter the unwanted data from the dump. But 80% is equal to 200 person hours and in terms of money, the wastage is $2,00,000.

Company is looking to improve the performance of the Analysis community by minimizing the efforts wastage.

To minimize the efforts wastage company require a dump which give more relevant data for further analysis to plan for the production. Since the company is planning for a time period with a Quarter (3months) frequency, dump receiving time will also become a base factor.
Online Miner striking features address the challenges of ABC Company. Since Online Miner provides filters on the screens to pull out the data by applying the keyword transformation and moving on to the repository according to the predefined category. In a short time, the tool will fetch the required information from specified source and moves on to the predefined category. As the data is deposited on category basis; it enables the feature of on demand analysis. Relevant category of information can be taken by the analysis community at any time. Online Miner data statistics will be readily available at any time so that analysis community can view the statistics of the required category of information in online.

Online miner can provide the data on the category basis to the ABC Company. So it avoids 80% of effort for separating unwanted information from the dump, since the statistics are provided online (statistics provided along with live data) analysis community can fetch the information from online miner very closer to the quarter end.

ABC Company provided its own relevant key facts to capture information from the web. Below is the list of word describers (key words) provided by ABC Company.

- Architectures
- Diagnosis
- documentation
- failure and recovery
- Failures
- identification
- integration
- integrity
- matrices
- models
- noise
- order
- reduction
- reliability
- security
- sensitivity
- state estimation
- synthesis
- theory
- transfer functions

Below is the Fig 6.1 allows the tool admin to configure the ABC provided business key words related to its own company business analysis. Admin of the Online Miner can configure the key words using "CREATE NEW WORD CATEGORY".
Fig 6.1. Screen for Key words configuration

To generate the statistics and capture the user behavior, two demo sites with search engine simulator and an e-marketing web site are created. Online Miner is integrated into two websites to monitor the user behaviors.

Fig 7.2 is the sample report on user behaviors, since the Online Miner can capture the logged in user information so reports can be provided basing on user ids.
Above graph explains “Suresh” area of interest is Studies and he worked towards Hardware and Socio-technical related stuff. With the analysis Suresh behavior is mapped towards technologies.

Next comes “Prasad” area of interest is Studies and he worked towards Hardware, Bio-Technologies related stuff and he is also worked towards transport. With the analysis, Prasad behavior is mapped towards technologies, Bio-Technologies and Hardware.

With the generated data, we can have a look on the data in various angles. Fig 7.3 is the screen shot with the most interested item for the web users.
The above chart represents the most interested element among the visited web users is "studies Hardware".

Below is the screenshot (Fig 7.4) for Online Miner with various category statistics. All the data is grouped into the various categories and each category contains related activity codes.
Fig 6.4. Online Miner Category Statistics

As ABC Company is interested in hardware information. Below is the screen shot (Fig 7.4) with online selection of hardware category
Fig 6.5. Chart for Hardware category Information
6.2. Comparison of Online Miner with Web Usage Mining tools

6.2.1. Pictorial differentiation

6.2.1.1. Current Web Usage Mining Tools Presentation

Stage 1

Below picture represents the user behavior and information available at user computer. It is considered as “Base data”.

![Base Data](image)

<table>
<thead>
<tr>
<th>Hyper link hits</th>
<th>User behavior 1</th>
<th>User behavior 2</th>
<th>User behavior 3</th>
<th>Error Data</th>
<th>Security Info</th>
<th>Firewall</th>
</tr>
</thead>
</table>

Fig 6.6: Base Data

Stage 2

Base data collected from transactions and stored into the log file by the web servers,

![Log file data](image)

Fig 6.7: Log file data
Why data looks mixed and scattered in the log file?

Most of the transaction populated with security information, user behavior, sessions (if exist), firewall related data, error data (if exist) and user hits on the hyper links etc., all the data collected from the transaction will be stored in the log file \textit{as-is}, this is the reason why data in the log file look scattered or disordered.

Stage 3

Tools come into action and apply the complicated quires with the criteria along with hard-coded strings to pull out the data from log file and move on to the WUM repository.

Example

The tool has to gather the user \textbf{behavior 1}, \textbf{user behavior 2} and \textbf{user behavior 3} and rest of the data is ignored from the log file.

According to the above criteria the data visible in the WUM repository is shown below (Fig: Data taken out from bulk log file).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig6.8.png}
\caption{Data taken out from bulk log file}
\end{figure}

Stage 4

Transformation of the data is to adjust the data according to business requirements.
Example:

The tool configured to transform the data from "user behavior 3" to "User behavior 2"

After applying the transformation on the WUM Repository, updated view of data given below (Fig: Data after applying transformation)

\[ \text{Fig 6.9: Data after applying transformation} \]

Stage 4

Data is ready in WUM repository for submitting to report.

Reports generation will depend on the business specific design.

6.2.1.2. Online Miner Tool Presentation

Stage 1

Below picture represents the user behavior and information generated from the user computer. It is considered as "Base data".

\[ \text{Fig 6.10: Base Data} \]
Stage 2

Online Miner is into the action, it first applies the “Rule set” on the “Base data” along with transformations through keywords as per our previous discussion transformation of the data is to adjust the data according to business requirements.

Example

Lets take a scenario where the Rule set is defined to fetch the “User behavior 1”, “User behavior 2” and “User behavior 3” data and rest of the data has to be ignored from the Base data with the transformation of data from “User behavior 3” to “User behavior 2”

After applying the “Rule set” and transformation on the Base data the data stored in the WUM will be: Fig: Data after applying Rule set.

```
AAAAAA 😊😊😊😊
AAAA 😊😊😊
```

Fig 6.11: Data after applying Rule set

Stage 3

Data is ready in WUM repository for submitting to reporting tool

Report look and feel depends on the business specific design
6.2.2. Storage and Performance differentiation

6.2.2.1. Current Web Usage Mining Tools Presentation

In a scenario:

- Storage space

  One transaction will take 2mb of disk space, then 1,00,000 transactions will take 194.31Gb of disk space. Therefore, as long as transaction increases, log file size will increase proportionally. This is an additional task to the system administrators to maintain the log file for WUM purpose.

- Search Performance

  Time taken to Search for the required data is always directly proportional to data available in the log file. If the log file is big then it takes longer time to search the data in the file. The time taken to open and close the file also included.

  Let us consider searching the word/given data to search from a 2mb data it would take 4 min and searching the data in 194.31 GB will take 8333.33 hrs (347.22 days)

6.2.2.2. Online Miner Tool Presentation for the above Example

- Storage space

  Here, there is no offline concept, all the data taken is from the live source. Therefore, this tool does not require any additional space for storing log data. Storage space for log file is zero
As data is directly taken from the live source there is no log file, thus no concept of searching for data. Therefore, the performance is high.

Below chart (Fig 4.11) shows the storage and performance difference between current WUM tools and Online Miner tool.

Fig 6.12: Current Web Usage Mining Tools and Online Miner
6.3. Benefits of Online Miner:

- Data is captured live and it starts working when the user sends the request to the server.

- It generates rapid and accurate reports since the data is captured through filters.

- It is an add on tool and can be configured very easily.

- Collects data excluding the error messages and noises and Data mining repository contains only the required information since filters out unwanted data before allowing data in to repository.

- Data transformation is done with dynamic transformation settings and contains straightforward queries.

6.4. Who can get benefited from Online Miner?

- Businessmen extract and collect the market information, product related data, or real estate data.

- Journalists can extensively work with articles.
• Matrimony sites extract the hits viewers’ information and manage it appropriately.

• Bidders and collectors automate extraction of betting and auction

• Information from auction sites.