ISSN 2278-3091
Volume 2, No.5, September - October 2013
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://warse.org/pdfs/2013/ijatcse02252013.pdf

Adaptive Advertisement Selection and Recommendation System based on User's Browsing Style

Ajay Jain
Adobe Systems, India
ajjain@adobe.com

ABSTRACT
A system for use in an online advertisement selection includes tracking, measuring browsing style of online users in order to further interpret, calculate and categorize the information to generate user-specific profiles for targeted advertisement. Auser's style means the way a user interacts and responds to presented situations through personal skills, spontaneous or/and influential decisive actions, interaction needs and is in essence a potentially determinable way of working online. The said actions result in a user-specific speed of transition(across web browser tabs), cursor movement, editing (across blogs), commenting, reading (of published content), etc.

Data pertaining toa user's browsing style is further compared with the cumulative plurality of the web browsing style of other users through the use of a centralized recording and monitoring program running at a designated computer server. Technological improvement represented through this work actually determines profiling values based on a user's web browsing style and uses the categorized profiles to intelligently select the most suitable advertisement for streaming and for a display to the that user. Because the data about the online users' web browsing style along with the data about the web pages content and the online advertisements users access is collected heuristically, the proposed algorithm optimizes and makes precise recommendation about the selection of advertisements

Keywords
Online Advertisements, Browsing Style, Web Analytics,

1. INTRODUCTION

While many techniques and algorithms are traditionally used by advertisers/publishers to select the "so called" right advertisements for online media (for example, online advertisements [1][2] on webpages or video sharing sites etc), the industry average of advertisements click-through rate is still very low (In the range of 2%) [3]. There is scope of optimizing ad selection to match user profiles, thereby enabling the advertisements receive better attention from users.

One of the major challenges faced by online advertisers and publishers is to identify and select the right online advertisement from a huge database of advertisement entries, and stream the right advertisement to the online users to receive timely and more attention from them.

The task becomes daunting because there are billions of online users with diverse styles, such as keyboard usage styles (some people much type faster than others,) and pace of consuming content(some people read much faster than others). These factors result in different internet browsing experiences.

Brisk readers may switch to the next webpage within a short interval while others might spend a relatively longer time on same page before moving on to next page. Similarly, for online videos, some people quickly judge a video's content just by viewing a few frames of the video and jump on to a different video or a new webpage, while other users may spend more time to view the video before going to the next.

The term "Online User" signifies a person going online (i.e. using internet) to browse websites, view webpages, read news, write blogs, watch online videos, movies, play online games, etc.

2. PROBLEM STATEMENT

As mentioned above, different users surf online content in different ways. Sometimes, this depends on the content of the web; on other occasions, it depends on the mood of an online user or on the thing the online user is searching for. As discussed before, even when the web content is the same for all online visitors/users, each user has a unique way of experiencing and assimilating internet content. For example, a user with a) fast typing speed, b) good control over mouse cursor, c) capability to read & comprehend text/pictures/videos quickly d)low attention span have a tendency to open a page and move on to another one within a short interval of time, while other users might show a different tendency .g. spend more time on same web page (due to whatever reasons) and hence might take relatively longer to move to a new webpage.

Selecting and streaming advertisements using traditional methodologies in such cases does not prove to be effective especially when the same (or a set of) advertisement(s) is being streamed to the users irrespective of their web browsing styles (i.e. when different users are browsing at different rates). Low-speed browsing makes a user focus on and give weightage to each element, - content of the web page, while high-speed browsing may not allow a user to spend enough time to read the page content. So, quick reader/user spends very less time in browsing through the content (possibly because his reading habit is so or because he has already seen the content earlier or because the content is not of his interest) and hence moves to next page.

In Summary, within a specific interval of time, a user may have surfed/visited/viewed 10 different webpages whereas another user, in same interval, may have been able to surf only 5 webpages. [Count of 10 and 5 are used for experimental purposes only, actual count can vary]. Therefore, there is a strong

need for devising an algorithm that selects advertisements with the goal of getting timely and full attention of an online user. If devised correctly, the algorithm definitely increases the probability getting higher click-through rates.

3. PROPOSED SOLUTION

A user visits a webpage and performs tasks based on his/her unique style. Any online activity done by a user constitutes a task, including (but not limited to) opening a website, closing a website, playing a song, watching an online video, going to a new website, switching between browsers, tabs, typing a website address, chatting on website, typing on webpages (for example, blogging etc), or reading website content, news, playing online games etc.

Technological improvement demonstrated through this work provides an algorithm that actually determines numerical values based on a user's web browsing style, and uses these numerical values to determine the best and most suitable advertisement for streaming and display to that user. In summary, the presented work is an Adaptive Advertisements selection based on a user's browsing style (and speed) governed by one or multiple following parameters:

1. Fast or slow typing skills, (e.g. on blogging sites, during submitting questions/answers on a forum, filling a former a survey etc),
2. Control over mouse or on-screen cursor movements while selecting a specific link or menu item spread across different coordinates in a webpage,
3. Quick or slow reading styles of presented web content,
4. Interpretation and comprehension capabilities in the context of the streamed content (content can be online videos, movies, pictures),
5. Capability of early (or late) prediction of upcoming content by making quick (or slow) decisions and hence, moving to a new web page slowly (or quickly),
6. Frequent changes in objects of interests, online user gets quickly loses interest in current content and moves to a new content located elsewhere on a different web page,
7. Surfing preferences for images, text or videos on webpages
8. Preferences for traditional static parameters such as Page size, ad size (different sizes), ad type (different types like image, text, video, blinking, static etc), number of words in the streamed ad (incase of text or image ad), ad duration (in case video/animation ad), etc

Each webpage carries a new embedded JavaScript code that is used for tracking and measuring user styles.

4. WORKFLOW

Technical improvements proposed in this work are achieved through following entities:

- **Online user/visitor**: Represents an online user who connects to the Internet and surfs webpages/websites, reads online content, news, watches online videos, movies, online games etc
- **Website:** online sites for examples www.adobe.com,
- **Webpage:** online pages for example www.adobe.com/photoshop/tutorial.html,
- **Web Browsers** (like Microsoft Internet Explorer, Google Chrome etc).

Data Monitoring and Recording Module (DMR Module)

Figure 1 shows an online user opening a specific website using a web browser and then opening additional webpage(s). Data Monitoring and Recording (DMR) Module is a computer running the software program, in the form of an application or a plugin integrated in the web browser the user is using.

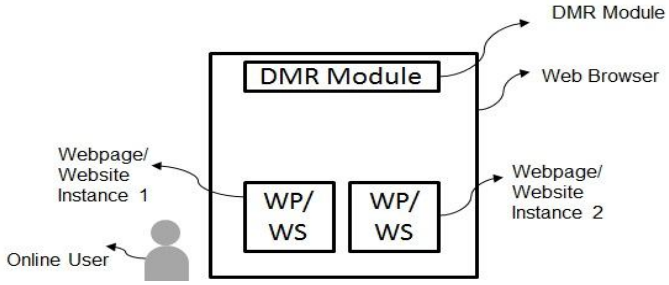


Figure 1: DMR Module

DMR Module collects the following information when a user performs any of these activities with a web browser: surfing websites, opening webpages, chatting, emailing, playing online games, watching online videos, movie clips, reading news articles, writing and reading blogs etc.

Each entry starts with the allocation of a specific ID to each user. User ID remains unique to every user.

{UserID | Date | Website | Webpage | Page size | Staying time on webpage | Time gap in movement to new Webpage |}

Sample data collected by DMR module for online users is shown in table 1.

Table 1: Database capturing User's data under various heads

Date/ Time	Website Visited	Webpage visited	Page size	Ts | Tg (in seconds)		
				UserID 1	UserID 2	UserID N…
mmddyy	www.news.com	wwww.news.com/1.html	2kb	10 |1 , 8 |1	20 |4 , 25 |5	…
mmddyy	www.news.com	wwww.news.com/2.html	15kb	5 |1 , 7 | 2	15 |4, 15 |5	…
mmddyy	www.video.com	wwww.video.com/first.html	20kb	17|2 , 20 |2	35 | 5, 40| 5	…
…	…	…	…	…	…	…

Where, Ts and Tgdenote the duration for which the user stayed on specific page and time gap between opening subsequent pages.

First four columns contain regular data about time;the website/webpage surfed, and page size. Important data covered in the above sample is the paired entry of Time spent on specific webpage by the user in combination with the time the user requires to open a subsequent webpage. {10 | 1} represents 10 units of time are spent by UserID 1, in his interaction with the webpage and 1 unit of time is spent in moving to a different webpage. While for same website and webpage i.e. www.news.com/1.html, UserID 2's data is {20|4} i.e. 20 seconds spent in interaction with webpage while 4 seconds are taken to move to a new website.

The following are the data samples for each UserID

a) {10|1, 8|1} for UserID 1 and
b) {20|4, 25|5} are for UserID 2.

These two data points clearly show that while the content offered by the webpage remains the same, different users spend different amounts of time in reading/viewing/watching the page content.

This differentiation happens because each user has a different reading style, a different writing style, a different mindset and a different style of interpreting the content etc. This user-specific differentiation is the core of this paper.

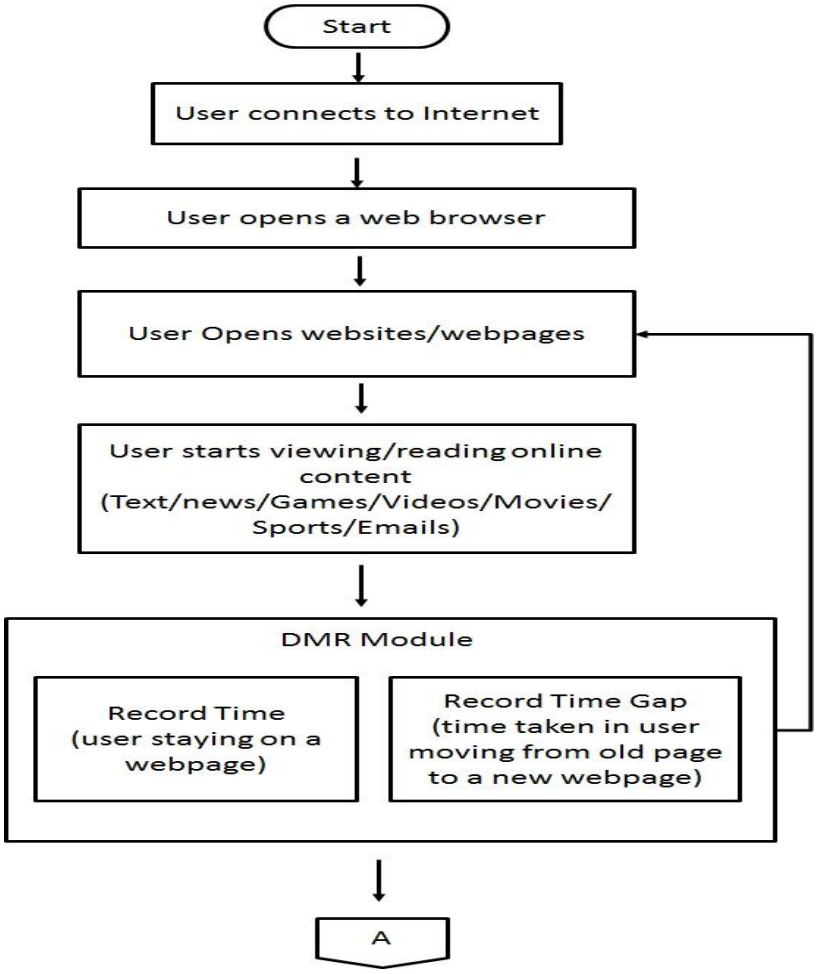


Figure 2: DMR module in Operation

Let's assume that the webpage/website content shown is the same for all users, and still the advertisements streamed to these websites/webpages are intelligently selected in accordance with a user's interaction and time spent on these websites/webpages.

DMR Module records data from each individualuser assigns this data to a unique UserID and sends it to Data Storage Module through internet connectivity as shown in (Figure 2).

Data Storage (DS) Module

Data Storage Module receives data for each user from the DMR module and stores it in a database.

For each UserID data received from DMR module, Data Storage Module synchronizes with Advertisement Server (AS) and gets advertisement data sent to the webpages/websites with respect to specific UserID.

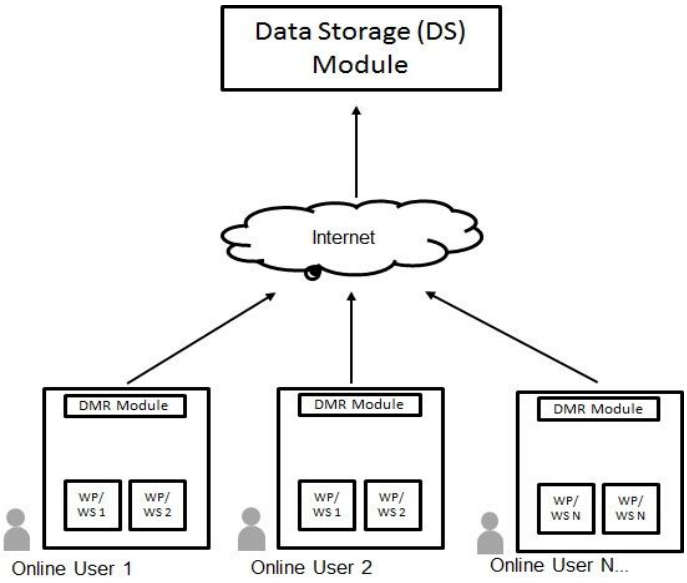


Figure 3: Data Storage Module

Table 2 shows how the advertisement server provides user-specific data and the content (UserID related time and the advertisement served) are saved in a database table. This table helps the algorithm evaluate the efficiency of streamed advertisements and further recommend advertisements that should be selected fora user.

Different advertisement specifics that are captured through the logic are:

- Ad size,
- Ad Type (Text or Image, or Animated, Rich Media video),
- Word count in advertisements (e.g. text ad carries strings etc),
- Words size (i.e. which font size these words are quoted in text strings/images etc),
- Ad duration (in cases when ad is rich media video, animation)

Data Evaluation Module (DE Module)

This module evaluates each advertisement entry specification streamed against a UserID, the associated surfed/visited webpage and websites in congruence with the time spent on the website as well as the time gap kept when moving between 2 websites by specific UserID.

Ad_Efficiency = Function of Users_BrowsingStyle [Ts+Tg] and [Ad_Spec]

whereUserIDii=1 to n

ifAd_Spec == Ad_in_Text

{if (Ad_WordCount>= 0.5*Ts)
Change_Advertisement}

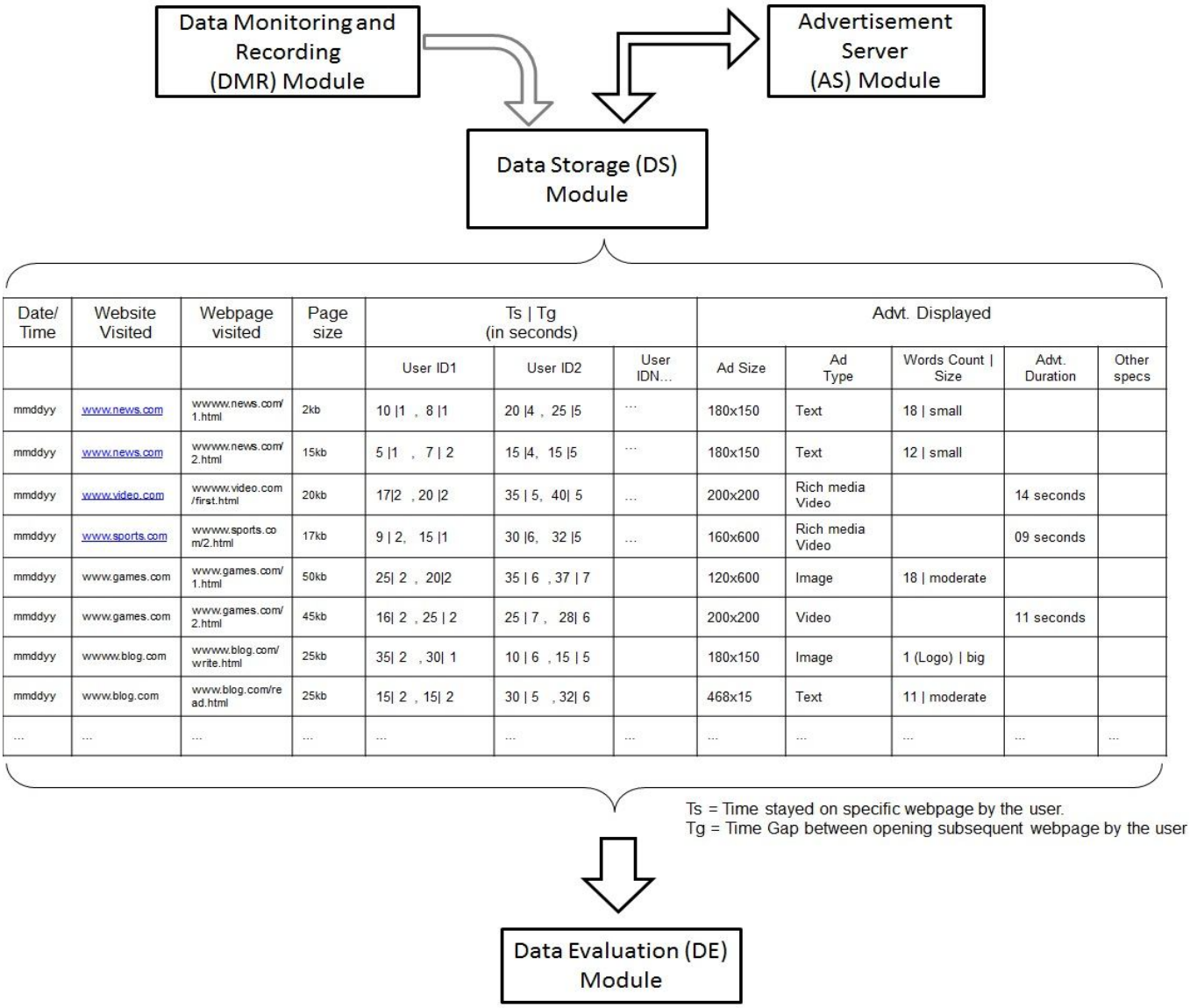
If Ad_Spec == Ad_in_RichMediaVideo

{if{(Ad_RunningDuration>= 0.5*Ts)}
Change_Advertisement}

Table 2: Sample Data

mmd dyy	www.news.com	wwww.news.com/2.html	15kb	5 |1 , 7 | 2	15 |4, 15 |5	…	180x150	Text	12 | small		
mmd dyy	www.sports.com	wwww.sports.com/2.html	17 kb	9 | 2, 15 |1	30 |6, 32 |5	…	160x600	Rich media Video		09 seconds	
mmd dyy	www.sports.com	wwww.sports.com/3.html	27 kb	11 | 2, 13 |1	39 |4, 31 |5	…	160x160	Rich media Video		19 seconds	
mmd dyy	www.youtube.com	wwww.youtube.com/movie.html	04 mb	17 | 3, 11 |1	20 |6, 22 |5	…	160x600	Rich media Video		05 seconds	

Communication channels across components are shown here:



Date/ Time	Website Visited	Webpage visited	Page size	Ts | Tg (in seconds)			Advt. Displayed				
				User ID1	User ID2	User IDN…	Ad Size	Ad Type	Words Count | Size	Advt. Duration	Other specs
mmddyy	www.news.com	wwww.news.com/ 1.html	2kb	10 |1 , 8 |1	20 |4 , 25 |5	…	180x150	Text	18 | small		
mmddyy	www.news.com	wwww.news.com/ 2.html	15kb	5 |1 , 7 | 2	15 |4, 15 |5	…	180x150	Text	12 | small		
mmddyy	www.video.com	wwww.video.com /first.html	20kb	17|2 , 20 |2	35 | 5, 40| 5	…	200x200	Rich media Video		14 seconds	
mmddyy	www.sports.com	wwww.sports.co m/2.html	17kb	9 | 2, 15 |1	30 |6, 32 |5	…	160x600	Rich media Video		09 seconds	
mmddyy	www.games.com	www.games.com/ 1.html	50kb	25| 2 , 20|2	35 | 6 , 37 | 7		120x600	Image	18 | moderate		
mmddyy	www.games.com	www.games.com/ 2.html	45kb	16| 2 , 25 | 2	25 | 7 , 28| 6		200x200	Video		11 seconds	
mmddyy	wwww.blog.com	wwww.blog.com/ write.html	25kb	35| 2 , 30| 1	10 | 6 , 15 | 5		180x150	Image	1 (Logo) | big		
mmddyy	www.blog.com	www.blog.com/re ad.html	25kb	15| 2 , 15| 2	30 | 5 , 32| 6		468x15	Text	11 | moderate		
…	…	…	…	…	…	…	…	…	…	…	…

Ts = Time stayed on specific webpage by the user.
Tg = Time Gap between opening subsequent webpage by the user

Figure 4: Communication Channels across components.

The first datashows, UserID 1 surfing the news.com website and on page 2.html spends 5 units of time in first instance while spending 7 units of time in second instance.

On other hand, the UserID 3 surfs the same page on the website but spends 15 units of time in each first and second instance of visit.

The first phase of evaluation covers: If we compare the advertisement streamed in both these cases, “text” advertisement was streamed where 12 was the word count and the font size was small. Proposed invention algorithm evaluates and determines if a user is spending only 5 units of time on a particular website it is highly unlikely that he will focus on reading descriptive ad of text of 12 words and in a small size. While the same ad might get relatively more attention from a user who is spending 15 units of time and spending approx. 5 units of time when moving to a new page.

Similarly, in the second case where the user is surfing a sports site and watching video, on 2.html page, UserID 1 spends 9 units of time in first instance of visit and spent 15 units of time in second instance with just ½ units of time taken in moving to a new site. While, UserID 2 on same page, stays for a relatively longer time (he might be interested in sports or watching the game first time hence more interest), spends 30 units of time and 32 units of time in his two visits. In these cases, advertisement streamed is a Rich media ad that has a running time of 9 seconds,

while traditional advertisement algorithm selects this rich media correctly for streaming on a sports page but since userID1 spent just 9 seconds on the webpage he wasn't able to view the ad in its entirety because the ad itself was of 9 seconds duration.

The second phase of evaluation is done by graphically plotting a scattered diagram on the browsing style data of each user with respect to the website/webpage the user surfed and is represented here (single time value for each sample is generated by merging Ts and Tg)

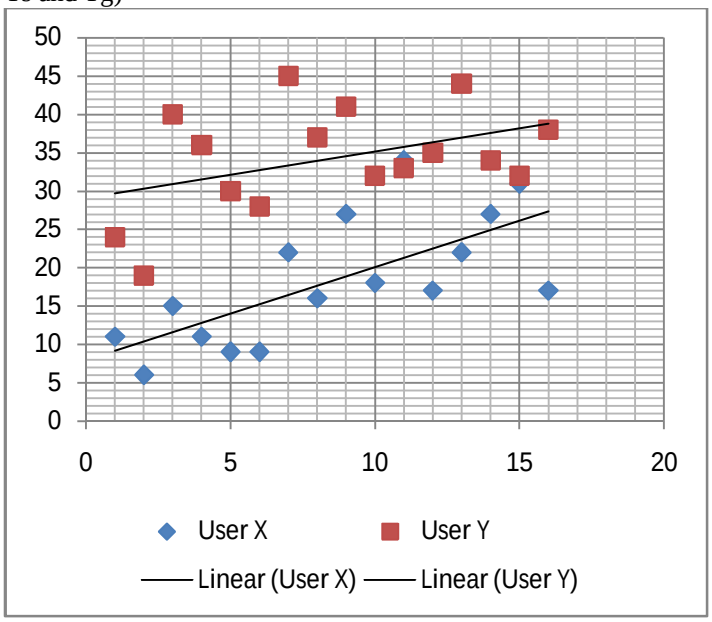


Figure 5: Data spread across all Users.

The third phase of evaluation includes sorting out style values coming from all users and categorizing them in different advertisement recommendation brackets as per the available advertisement specifics as shown below.

Ad Recommendation Module (ARM): This module takes data from the evaluation module, interprets it and finally makes recommendation to the Advertisement server for streaming ads to users as per recommended advertisement configurations.

Different categories are then churned in order to group the evaluated data amongst best set of advertisements that can be streamed for each specific category. Ad categories listed in below table are suggestive and shall be closely monitored about their success when streaming these ad categories in real time. Based on individual publisher's success, further re-categorization of ads shall be performed in order to get best ad clicks from the field.

Because the goal here is to locate precisely the advertisement target specific to the user, different categories are created. Depending upon advertisements variances availability, more categories can also be churned. Keeping a few advertisements common to two nearby categories is also a helpful approach in evaluating the kind of advertisements the user is interested in over a period of time. Depending on the user's click rate on these multi-category located advertisements can then be used to further allocate best matching category for the said advertisements.

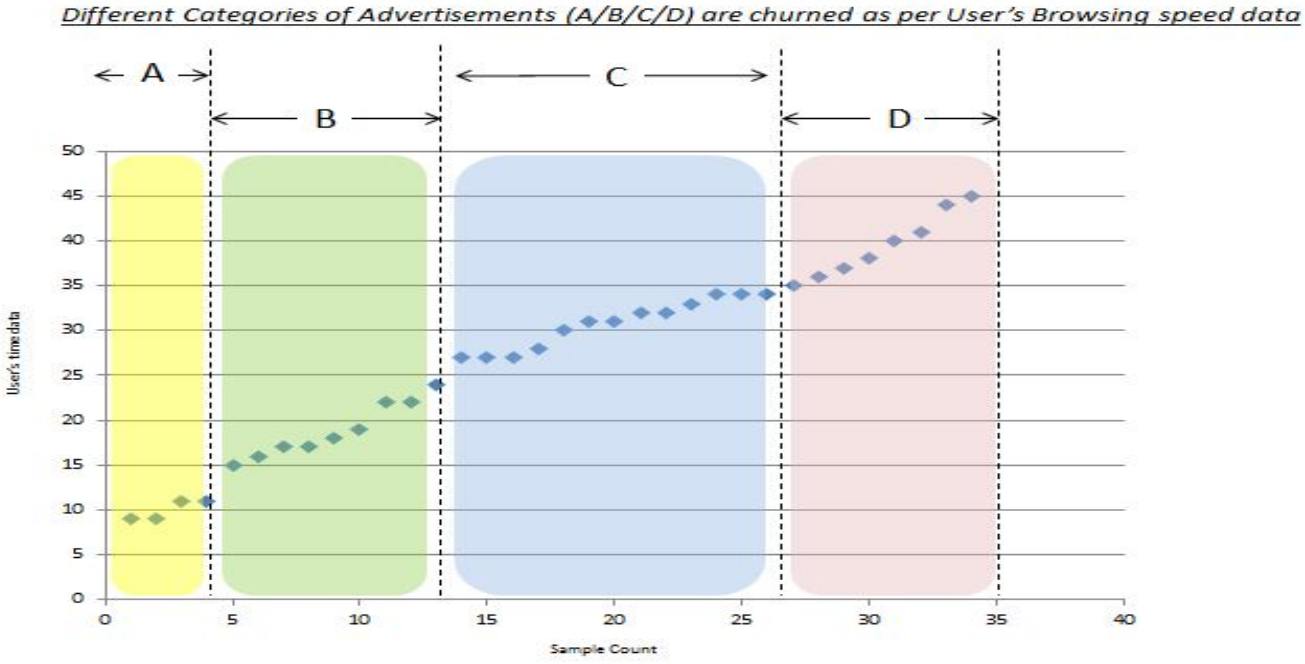


Figure 6: Grouping and Categorization of collected samples.

	Category Created	Advertisement Recommendations	Recommendation Benefits
A	Category A Advertisements (Browsing style: Fastest)	• Ad Size >=400*400 pixel • Ad Duration <2 seconds • Ad Text Count < 3 words • Ad image = Preferred Big image (Say logo)	User ends up spending a very few seconds on the webpage. Recommended ads are that of Bigger image, with say a company logo, almost no textual ads and no ads whose duration is more than couple of seconds.
B	Category B Advertisements (Browsing style: Moderately Fast)	• Ad Size >250*250 pixel • Ad Duration <6-7 seconds • Ad Text Count < 6-7 words • Ad image < Preferred Big /Moderate image (Say logo)	User browsing style is of high speed and the user is spending less time on website but relatively more time when compared to "A" case. Ads recommended are of high size; few text words can be accommodated in textual ads, moderate or bigger logo image will be more beneficial ad.
C	Category C Advertisements (Browsing style: Medium)	• Ad Size >=100*100 pixel • Ad Duration <10 seconds • Ad Text Count < 10 words • Ad image < Preferred Medium-Big image (Say logo or Image with text embed) • Ad Type: Rich media Ad with Ad duration <7 seconds	User browsing style is of moderate speed and spending substantial amount of time on website, Ads recommended are of all generally available size, with and without textual content can be accommodated; Big/ Moderate size image will attract user's attention. Rich media ads with playing duration of 7-10 seconds can be easily viewed by the user.
D	Category D Advertisements (Browsing Style: Lowest)	• Ad Size >=100*100 pixel • Ad Duration <10-15 seconds • Ad Text Count < 10-15 words • Ad image < Preferred Small/Medium/Big image (Say logo or Image with text embed) • Ad Type: Rich media Ad with Ad duration <10 seconds	User's browsing style is very slow. The user is spending significant time on each website/webpage, giving advertisements all the opportunity to get viewed, textual content also having equal chance to get read, images (small, medium, big) all have equal chances of not getting missed. Rich media ads running around 10 seconds can also be fully viewed by the user when we compare the time user is spending on the website.

Table: Advertisements mapping as per User's Browsing categorization.

DMR MODULE DIAGRAM:

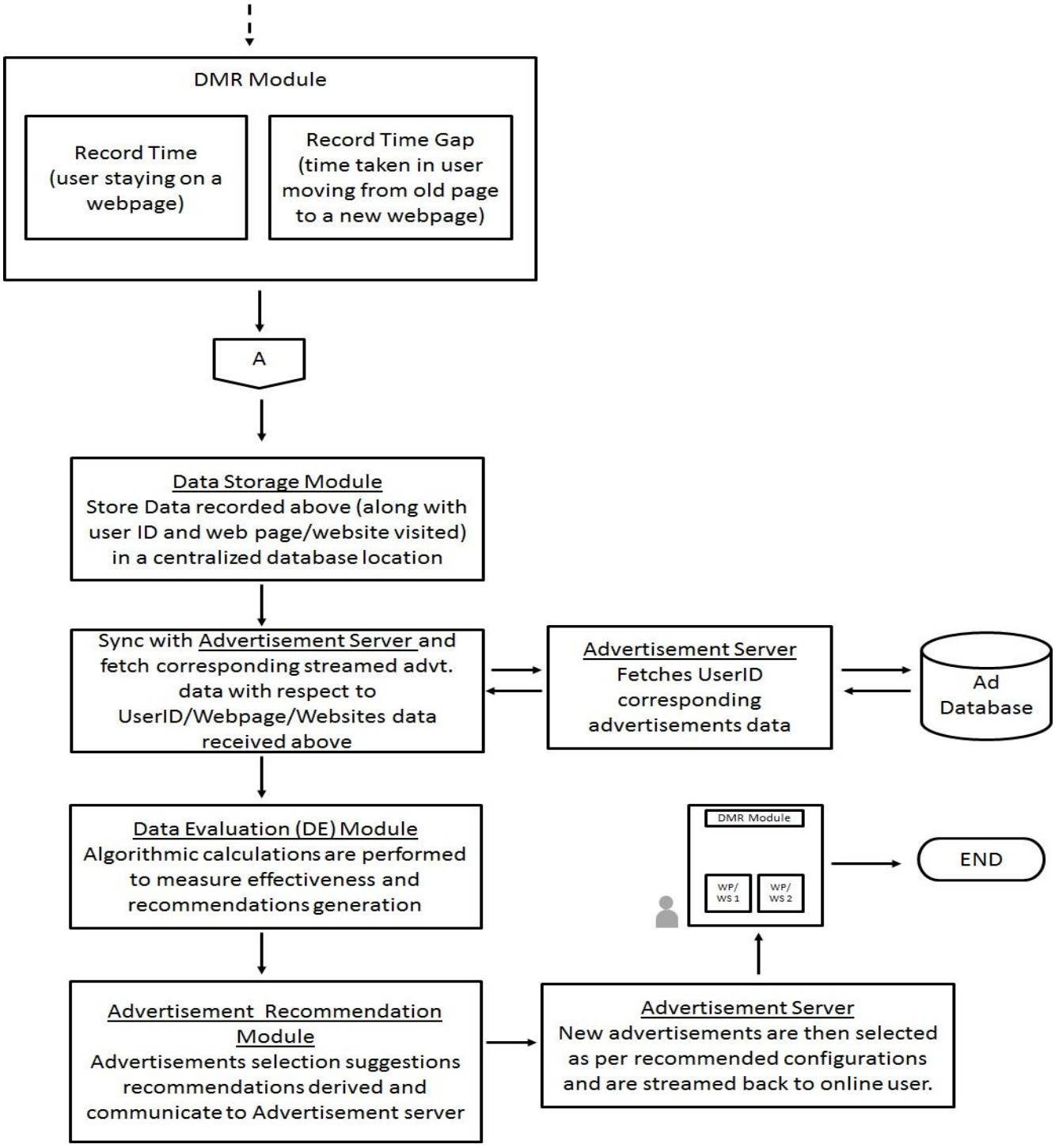


Figure 7: DMR Module

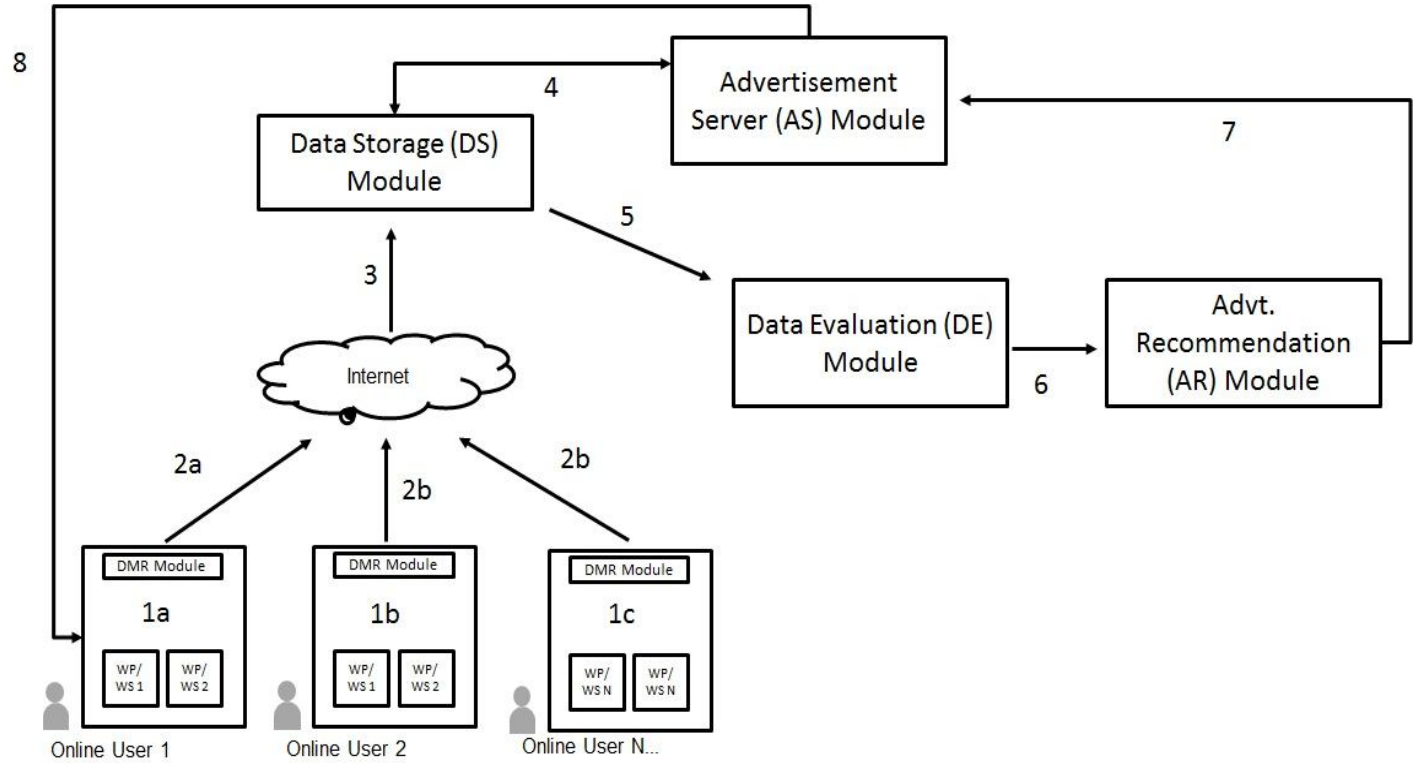


Figure 8: End to End System Workflow Representation

LEGENDS

1a/1b/1c: Three online users, invoking web browsers for accessing/reading/viewing websites and webpages carrying videos, news, text, sports, emails, movies, online games etc.

2a/2b/2c: Communication mechanism through which Data Monitoring and Recording (DMR) Module's collected data on each user's browsing style; user's staying time with respect and mapped to each webpage, websites, performing online activities, time gap value of user's movement from one webpage to a new/different web page, data transferred to internet. Data recorded is in format of

{UserID | Date | Website | Webpage | Page size | Staying time on webpage | Time gap in movement to new Webpage |}

3: Data transferred as part of step 2a/2b/2c is sent to Data Storage (DS) Module

4: Data Storage (DS) Module records user specific data in designated format and contacts Advertisement Server (AS) in fetching data related to advertisements that were streamed with respect to specific User ID.

{UserID | Date | Website | Webpage | Page size | Staying time on webpage | Time gap in movement to new Webpage}

i.e. for each User ID entry, new table entries with respect to streamed advertisement parameters are added and each table row will look like this:

{UserID | Advertisement Streamed ID | Ad size | Ad Type | Total number of Words in streamed Ad | Ad Duration | Other Ad Specifics}

5: Entire User ID and corresponding Advertisement ID data is mapped with real time values are stored in the database table and these are then forwarded to Data Evaluation (DE) Module for data mining and analysis.

6: Advertisement Recommendation (AR) Module generates advertisement recommendations to Advertisement Server for streaming recommended advertisements to website users/ visitors.

5. CONCLUSION

Our work presents an interesting user's browsing style profiling based on the actions a user performs on the web, internet. A complete uniqueness and spontaneity of a user is maintained and monitored. Data collected for each user is stored and compared against plurality of other thousands of user profiles available. Because the data is collected heuristically, there are high chances of locating an advertisement with a high precision rate. This way,the data is collected, mined, and recommendations are then charted with respect to selecting and streaming best online advertisements with a higher chance of getting said user's attention towards the said advertisement.

REFERENCES

[1] Google Ads http://ads.google.com
[2] Yahoo Ads http://advertising.yahoo.com
[3] Display Advertisement Click Rate http://www.smartinsights.com/internet-advertising/internet-advertising-analytics/display-advertising-clickthrough-rates/