



WEB USAGE FORECAST WITH ADAPTED LEAST RECENTLY USED MODEL

Prapai Sridama¹, Somchai Prakancharoen² and Nalinpat Porrawatpreyakorn¹

¹Department of Information Technology, Faculty of Information Technology, King Mongkut's University of Technology, North Bangkok, Thailand

²Department of Computer Science, Faculty of Applied Science, King Mongkut's University of Technology, North Bangkok, Thailand

E-Mail: prapaikmutnb@gmail.ac.th

ABSTRACT

The research focuses on decreased downloading of web pages from other web servers and decreased expenditure of cost. The quantity of web usage from people is increasing. This situation is effected to the networking traffic. The researchers try to solve this problem with new techniques. Unlucky, those techniques can give the hit rate not more 40 percent. Therefore, this research investigates Adapted Least Recently Used (ALRU) algorithm for solving the networking traffic. Many Mathematical Statistics theories are used in this algorithm for increasing the hit rate of the web cache memory as following; Estimated value equation, Interpolation with Cubic Spline, Interpretation of first derivative, Interpretation of second derivative, First Order Condition (FOC) and to find area under curve function. In addition, this model is tested with data sets of web usage from Walailak University. The result of this algorithm is shown that the quantity of downloading decrease more than other algorithms. However, the performance of ALRU algorithm has compared with the LRU algorithm which is a popular algorithm in the present. Furthermore, the maximum of the hit rate of the ALRU algorithm is 72.56 percent but the LRU algorithm can give the maximum hit rate 53.8 percent. However, the average value of the hit rate of the ALRU algorithm is 53.8 percent while the LRU algorithm can give 12.99 percent.

Keywords: web usage pattern, cubic spline, interpolation, first order condition, estimated value.

INTRODUCTION

The internet is used for working in many organizations. It can help to send information between people and people. In addition, the rate of using online applications is quickly growing up because it decreased difficulty of connection and the cost of transferring information. Importantly, the internet is a mechanism tool for business competition.

The linking between people and people are easily. It is very easy to connect when they are at different locations. The communication system was developed from the original model to more. We can send picture files or multimedia files with online applications, which are replaced the outdated communication. Furthermore, international Telecommunication Union (ITU) said that the people more than 27 million uses internet in the present and the quantity of using internet is increasing quickly.

The modern technology and the speed of internet signal affect to the quantity of using internet. The people could use internet from personal computers or notebooks in the part but the people can use internet from mobile phones that they can use the web cache mobile phones everywhere. Furthermore, the cost of using internet and hardware supporting are decreased when we compare the cost.

To quickly increase the internet usage in the present is effect to traffic network problem. We found that the quantity downloading in each day is huge data. If

people would like to look some web pages then they send requesting to servers, which are at other locations. However, if the huge people send requesting in the same time then the waiting time is spent increasingly. In addition, any downloading has the cost. The organization pays the cost every time when users want to open any web pages.

The web cache memory is used for solving traffic network problem. It is a small memory, which is embedded at the server. However, the web cache memory has two types as following; 1) the web cache memory on server and 2) the web cache memory on client. This research focuses on the web cache on memory. The web cache memory can load some web pages from other servers, whenever the people want to look those web pages then their requesting are sent to the web cache memory. If the web cache memory has those web pages and it cannot service any requesting.

Many techniques are created for the web cache memory management. The replacement technique is an algorithm to help web page selection. The popular replacement technique is Least Recently Used (LRU) algorithm. It processes and selects web pages for picking in the web cache memory quickly. But, this algorithm gives less hit rate. While, other techniques give the hit rate not more 40 percent.

The objective of this research is increasing the hit rate of the web cache memory on server. Adapted Least Recently Used (ALRU) algorithm is investigated in this



research. The ALRU algorithm has adapted the LRU algorithm. Many Mathematical Statistics equations are used in the ALRU algorithm for improving the LRU algorithm. Those equations are follows; estimated value, Interpolation, Cubic spline, Interpretation of first derivative, Interpretation of second derivative, first order condition, to find area under the curve function, Normalization-LRU cost, and integration. The section 2 shows theories. The section 3 presents the ALRU algorithm. The section 4 shows the result of experiment and lastly, one presents the conclusion.

THEORIES

This section presents theories for using in the ALRU algorithm. Many Mathematical Statistics equations are explained as following;

Interpolation

The interpolation is a technique for forecasting some values, which miss from the system. In addition, the interpolation is used for fitting curves and this research uses it for fitting curve.

The interpolation can draw the line graph with the function to pass all point of data. The interpolation has three functions for finding values as following; 1) Newton's interpolating polynomials, 2) Lagrange interpolating polynomials and 3) Splines (Akila, 2013). However, this research uses interpolation with spline function.

The Splines function has three methods as following; 1) Linear splines, 2) Quadratic splines and 3) Cubic splines (Soros, 2011) and (Press, 1988). This research uses the Cubic splines for fitting curve.

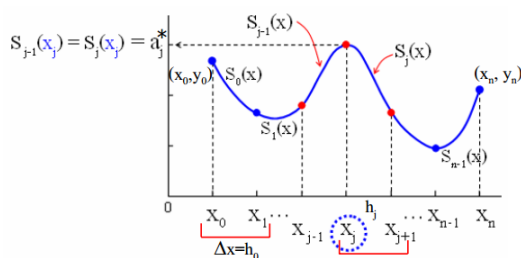


Figure-1. A graph is made with the Cubic Splines.

Figure-1 shows the fitting curve that pass four points with the Cubic Splines. The Cubic Splines makes function $f(x)$ that S_j has made smooth line (Soros, 2011). This function is shown at Equation (1).

$$S_j(x) = a_j + b_j(x - x_j) + c_j(x - x_j)^2 + d_j(x - x_j)^3 \quad (1)$$

where $0 \leq j \leq n-1$ and $S_j(x)$ is spline of (x_j, x_{j+1})

The Cubic Splines was used by many researchers. Because of other spline method could not calculate right estimation. Furthermore, Nipaporn, S. (Nipaporn, 2009) uses the Cubic Splines for 3D modeling, Wandee, J. (Wandee, 2001) investigates the Cubic Splines for approximating function in image processing. In 2008, Jorpsjo, Y. uses it for image interpolation and in 2009, Petrinovic, D. uses it for continuous time domain properties of causal.

Estimated value

The Estimated value is a method for calculating average values (Gerhard, 2010). The Estimated value has two methods that are:

The estimated value of discrete random variable: it calculates the expected values in discrete time. However, the estimated value is shown at Equation (2).

$$E \sum_{i=1}^n X_i = \sum x_i P(x_i) \quad (2)$$

Let $E x$ is estimated value of random variable x .

x_i is random variable i when

$i = 1, 2, 3, \dots, 4$.

$P(x_i)$ is probability of random variable x

The estimated value of continuous random variable: it uses to calculate expected values in continuous time. It presents by Equation (3).

$$E = \int_{-\infty}^{\infty} x_i f(x_i) dx \quad (3)$$

where $E x$ is the estimated value of random variable x .

x_i is random variable i when

$i = 1, 2, 3, \dots, 4$.

$f(x_i)$ is probability function of random variable x .

Interpretation of the first derivative

The interpretation of the second derivative is a function for calculating the derivative value (Rutchawit, J., 2010). The symbol of this function is $f'(x)$. This research replaces the function with polynomial equation, which is the third derivative. However, the positive function and negative function is given in this function.

Interpretation of the second derivative



The interpretation of the second derivative is used to find the changed rate (Ruchawit, J., 2010). The symbol of this function is $f''(x)$. We can use this function after we calculate function with the interpretation of the first derivative. In addition, we know the slope of curve with this function. However, this research uses this function for finding the trend of web usage before selection the top ten.

First Order Condition (FOC)

The first condition is used for testing the second derivative. Usually, the optimized problems use FOC for finding the maximum value and minimum value. This research uses this function to solve the maximum value of web usage in the past. Example of the optimization from a function ($f(x) = 2x^3 - 30x^2 + 126x + 59$).

Firstly, the FOC finds the critical value.

$$\begin{aligned} f(x) &= 2x^3 - 30x^2 + 126x + 59 \\ f'(x) &= \frac{d}{dx} [2x^3 - 30x^2 + 126x + 59] \\ &= 6x^2 - 60x + 126 \end{aligned}$$

$$\text{Let } f'(x) = 0$$

$$\text{So } 6x^2 - 60x + 126 = 0$$

$$x^2 - 10x + 21 = 0$$

$$(x-3)(x-7) = 0$$

The critical values of this function are 3 and 7.

Secondly, the FOC tests the second derivative at both critical values.

$$\begin{aligned} f''(x) &= \frac{d}{dx} [6x^2 - 60x + 126 = 0] \\ &= 12x - 60 \end{aligned}$$

$$f''(3) = 12(3) - 60$$

$$= -24 < 0$$

$$f''(7) = 12(7) - 60$$

$$= 24 > 0$$

The maximum value of this function is $x=3$ (the concave) and the minimum value is $x=7$ (the convex).

To find the area under the curve

To calculate the area under the curve or to calculate area between the graph of function (Ruchawit, 2010) and the phase on the horizontal axis can be presented the example at the Figure-2.

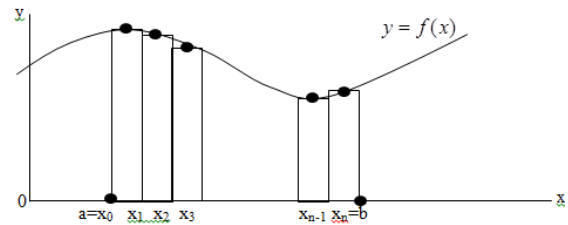


Figure-2. The area between the graph of function and the phase on the horizontal axis.

Figure-2 shows the line graph of the continuous function that $y = f(x)$ and $f(x) \geq 0$ in close phase [a, b]. We get the area between the graph of $f(x)$ and the phase of [a, b] that we represent with A.

If we divide the phase of [a, b] to sub phases then we get $[x_0, x_1], [x_1, x_2], \dots, [x_{n-1}, x_n]$ with the breaking point are $a = x_0 < x_1 < x_2 < \dots, x_{n-1} < x_n = b$ and the widths of each phase (Δx_i). In each of Δx_i have size to approach zero when n is increasing value. However, the procedures of finding the area under the curve are shown as following;

Step-1: we make the rectangles high equal the tallest of each piece. After that, to select x_i point from each sub phases and calculates the height of graph in each point of x_i that we calculate from the $f(x)$. So, the height of each piece is equal $f(x_i)$.

Step-2: we sum the area of rectangles. The equation is used and shown at Equation (4).

$$f(x_1)\Delta x_1 + f(x_2)\Delta x_2 + \dots + f(x_n)\Delta x_n \quad (4)$$

$$\text{or } \sum_{i=1}^n f(x_i)\Delta x_i$$

Step-3: we take the limit or the integration. This part is finding the real area. The total value of all rectangles is the area approximately of area that those values are not the real values. If we want to find the area, which is near the real area of A then we should decrease the wide range in each sub phase. After that, to calculate the area of A make with the limit or the integration that this function is presented at Equation (5).



$$\int_a^b f$$

$$\text{or } \int_a^b f(x) dx \tag{5}$$

where \int is integral sign
 $f(x)$ is integrand or the function is made integration
 dx is the differentiation operator

The performance comparison

The comparing efficiency of web caching compares with the hit ratio and byte hit ratio (Jaynes, 2003). Those equations are shown at Equation (6) and Equation (7).

$$HR = \frac{\sum Hi}{ERi} \tag{6}$$

$$BHR = \frac{\sum (Si * Hi)}{\sum (Si * Ri)} \tag{7}$$

where Hi is the number of requesting of, object i , which are hit.

Ri is the number of all object i

Si is size of object i

METHODOLOGY

The Adapted Least Recently Used (ALRU) model is investigated in this research. We collected using web pages from the log files of the Walailak University between January and July 2013. However, we explain the ALRU model in this section.

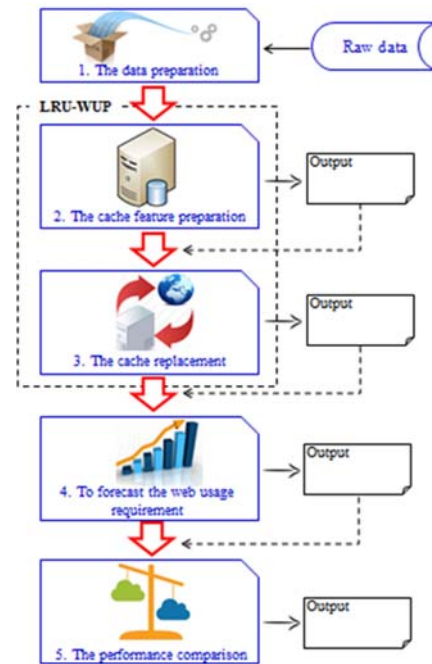


Figure-3. Procedures of the ALRU model.

The Figure-3 shows procedures of the ALRU model. The first procedure cleans the data from log files. The second procedure has used Mathematical Statistics algorithms to calculate the top ten of web pages. Selection of top five from top ten of web pages and simulating the replacement is in the third procedure. The fourth procedure has tested the web usage requirement. Lastly, this procedure has compared the hit rate and the cost of downloading between ALRU model and LRU model.

Data preparation process

The data preparation process is the first of working in ALRU model. This procedure is rather than spending a lot of time because data sets more than 10, 000, 000 records. However, this procedure is shown at Figure-4.

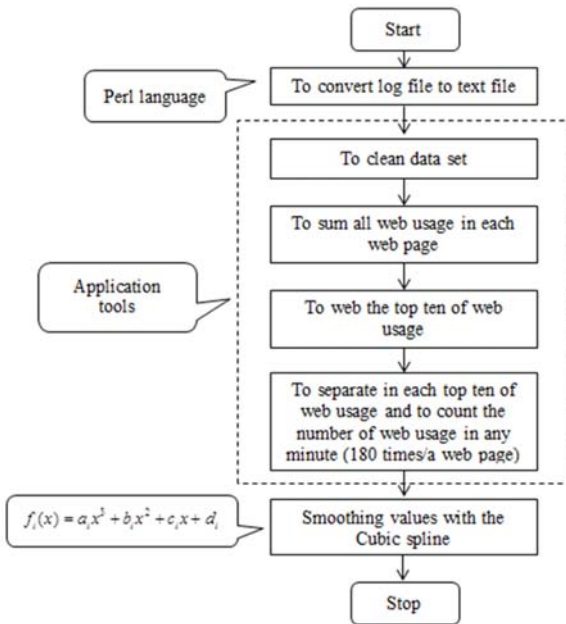


Figure-4. The procedure of the data preparation.

Figure-4 presents the data preparation procedure. This procedure investigates to clean data sets, to find the top ten of web usage, to sum the number of web usage in any minutes between 9.00 a.m. and 12.00 a.m. (180 times) per a web page of the top ten, to separate web usage from all web usage in each web page to each week and to sum those data again (30 times/web page) and fitting curve values from number of web usage in any minutes with the Cubic splines.

The cache feature preparation

The ALRU model is developed for comparing with LRU model in this research. The ALRU model can forecast requests of web usage in the future that it is the main objective of this research.

This part presents many algorithms and data mining technique for solving problems. However, after finishing in this part we get new algorithm to make decision about the web cache replacement and the web usage prediction.

1) To find the value of each pint representing: the estimated value is used in this part for finding the expected values in each web page (top ten). So, in each of the top ten is received the expected value from thirty values.

2) Determination web usage in each period: this process has two steps as following:

Step-1: we calculate the average any ten minutes of all top ten with Equation (8), which is shown at equation bellow.

$$AVG(J) = \frac{\sum_{i=1}^{i=9} J_i}{i+9} \tag{8}$$

where i is the number of time in each minute
 J is the web usage quantity in each minute

Step-2: we find the area under the curve. The area under the curve values can help to know about all web usage volumes. So, those volumes are used for looking forward. The integral is the equation for calculating the area under the curve in this research and it is shown at Equation (9).

$$\int_a^b f(x) dx \tag{9}$$

where \int is integral sign.

$f(x)$ is the function for finding the area under the curve with the integration

dx is the differentiation operator

$\frac{d}{dx}$ is to find the derivative when comparing with x variable.

Equation (9), we will adapt it to Equation (10) for find all under the curve. The Equation (10) is shown at equation below.

$$A = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(t_i) \Delta t_i \tag{10}$$

where t_i is the average value of time at i

Δt is the time between t_1 and t_{18}

n is the time at t_{18}

However, this research divides finding area under the curves any ten minutes so we are gotten those values equal eighteen times. The function in each sub areas is

$$f(t_1) \Delta t_1 + f(t_2) \Delta t_2 + f(t_3) \Delta t_3 + \dots + f(t_n) \Delta t_n$$

So, equations to find the sub areas are shown at equations below.

$$A_1 \Big|_1^2 = \int_{t_1}^{t_2} f(t) dt$$



$$A_2 \Big|_2^{t_3} = \int_{t_2}^{t_3} f(t) dt$$

$$A_3 \Big|_3^{t_4} = \int_{t_3}^{t_4} f(t) dt$$

$$A_{17} \Big|_{17}^{t_{18}} = \int_{t_{17}}^{t_{18}} f(t) dt$$

$$\therefore A = A_1 + A_2 + A_3 + \dots + A_n$$

We conclude the cache feature preparation procedure in Figure-5. However, the $f(x)$ in each period depend on nature of fitting curve values.

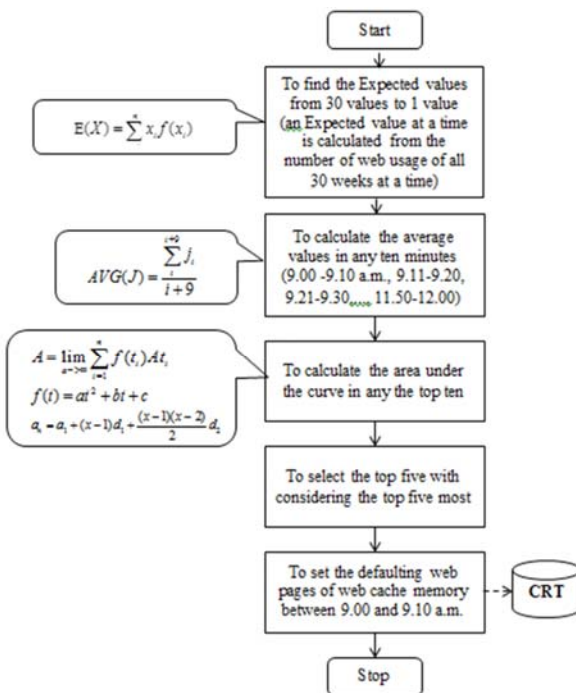


Figure-5. The procedure of the cache feature preparation.

Figure-5 presents processes of the cache feature preparation procedure. There are five processes as followings: 1) finding the expected value of any minute from 30 weeks, 2) calculating average values any ten minutes ($t_1=9.00-9.10$ a.m., $t_2=9.11-9.20$ a.m., ..., $t_{18}=11.50-12.00$ a.m.), 3) finding the area under the curve, 4) selecting the top five from the top ten that one considers from the area under the curve, and lastly, defining the defaulting web pages and transfer to CRT.

The cache replacement

This procedure has used the optimized technique to decide for the cache replacement.

Step-1: planning for the next ten minutes. After ten minutes to pass then one calculates Expected values and areas of under the curve again. But one removes values of the oldest day and increases values of the present day.

Step-2: finding the Normalized LRU (N-LRU). We use Equation (11) for solving this problem.

$$\text{Let N-LRU} = \frac{\text{MaxCacherserved} - \text{LRUofEachCache}\#i}{\text{MaxCacherserved}} \tag{11}$$

where i is a channel of the web cache slot. This research defines the total slots equal five slots.

Step-3: we find the cost of downloading. To find the cost of downloading uses the area of under curve value to solving the cost of downloading. This function is shows at Equation-11.

$$RC = [(t_2 - t_1) * c - \int_{t_1}^{t_2} f(t) dt] * CB \tag{12}$$

where RC is reverse of cost.

t_2 is the average value of a web page from top ten at t_2 .

t_1 is the average value of the same web page of t_2 .

C is the difference equation of $f(t)$.

$f(t)$ is the quadratic/cubic polynomial equation.

$\int_{t_1}^{t_2} f(t) dt$ is the integration of the quadratic/cubic polynomial equation.

CB is cost of downloading (Bath/kb.).

Step-4: Normalization RC is solved with Equation (13).

$$N - RC = \frac{A * CB - RC}{A * CB} \tag{13}$$

Step-5: we decide web pages into the web cache memory.

The last procedure is the decision web page into the web cache memory process. One defines the equation at Equation (14).

$$D = (W_1 * R - LRU) + (W_2 * N - RC) \tag{14}$$



Where W_1 and W_2 are defined by the system administrator that $W_1 + W_2 = 1.0$.

We conclude the processes of this procedure and explain in the Figure-6.

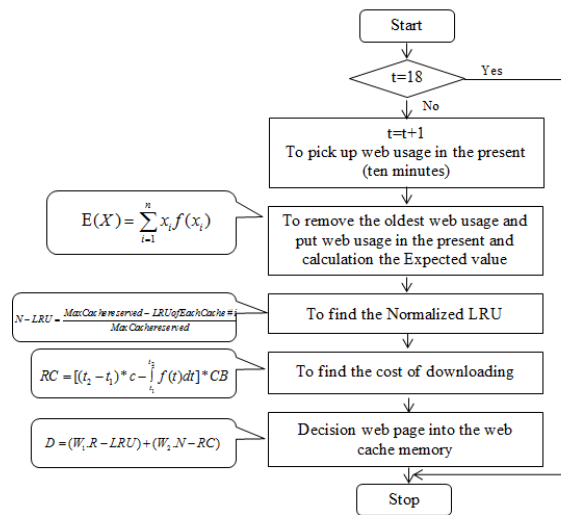


Figure-6. Processes of the cache replacement.

Figure-6 presents the concluding of the cache replacement procedure. The procedure decides any ten minutes.

To forecast the web usage requirement

This part is the testing step that we use historical data for hit ratio finding. However, we will compare hit ratio and byte hit ratio with LRU technique. In addition, we will compare web page qualifying period. For historical data, we would select the same data sets. For formulas in hit values and byte hit values, we can use the Equation (6) and Equation (7).

The performance comparison

This procedure compares the performance between ALRU model and LRU model. We take a random from testing set in each period ($t_1 - t_{18}$). After that, to take those data sets test the hit ratio between ALRU model and model. However, the hit ratio of the ALRU model and the LRU model are shown in the next section.

THE RESULT OF EXPERIMENT

The ALRU model uses data sets (web usage) from January 2013 to July 2013 of Walailak University. However, the procedures of the ALRU model are as follows: 1) the data preparation, 2) the cache feature preparation, 3) the cache replacement. So, one would like to show the results in each procedure.

The results of the data preparation procedure

The data sets are made from log files which are cleaned many steps. Nevertheless, the outputs from this procedure can be shown at Table-1. In addition, one can know the summary of all web pages that users visit between 9.00 a.m. - 12.00 a.m. There are pick up data in any minute.

**Table-1.** The top ten of web usage between January 2013 and July 2013.

Week	The quantitative web usage of the top ten									
	google	youtube	ptvcdn	yting	gstatic	teenee	kapook	adobe	dropbox	4shared
1	20,997	10,122	6,707	4,301	2,171	2,229	1,494	1,629	1,497	,197
2	20,509	10,198	6,230	4,320	2,02,	2,740	1,289	1,682	1,175	1,474
3	19,892	12,163	6,067	4,520	2,790	2,436	1,340	1,500	1,102	1,736
4	19,694	9,024	6,082	4,847	2,547	2,116	1,626	1,094	1,910	1,694
5	21,828	12,598	6,609	4,284	2,848	2,298	1,564	1,877	1,697	1,811
6	20,069	9,855	5,071	4,483	2,911	2,826	1,868	1,858	1,504	1,845
7	20,007	9,726	6,543	4,520	2,513	2,079	1,333	1,541	1,662	1,426
8	19,860	12,928	5,473	4,343	2,801	2,522	1,955	1663	1,834	1,471
9	21,810	12,986	5,247	4,028	2,569	2,930	1,503	1,720	1,894	1,089
10	21,269	9,020	6,586	4,519	2,432	2,229	1,807	1,830	1,926	1,169
11	20,969	11,959	5,431	4,862	2,762	2,799	1,000	1,746	1,560	1,408
12	21,478	12,218	5,238	4,389	2,837	2,145	1,765	1,980	1,491	1,451
13	21,769	9,169	6,694	4,126	2,864	2,641	1,706	1,217	1,544	1,921
14	21,131	11,229	6,784	4,029	2,390	2,931	1,907	1,192	1,047	1,408
15	21,990	12,482	6,236	4,820	2,922	2,147	1,640	1,029	1,466	1,978
16	20,275	9,980	6,308	4,443	2,062	2,401	1,397	1,374	1,733	1,133
17	21,666	11,822	6,636	4,804	2,428	2,410	1,195	1,219	1,561	1,869
18	21,891	12,643	6,182	4,124	2,620	2,198	1,994	1,171	1,968	1,990
19	21,476	9,337	6,511	4,698	2,089	2,639	1,547	1,710	1,066	1,533
20	21,416	11,083	5,258	4,732	2,364	2,967	1,123	1,337	1,888	1,552
21	21,355	11,371	6,480	4,482	2,139	2,899	1,421	1,732	1,142	1,888
22	20,646	12,246	5,948	4,895	2,087	2,143	1,723	1,041	1,583	1,469
23	20,624	9,227	5,133	4,721	2,932	2,610	1,575	1,486	1,843	1,976
24	19,724	12,661	6,904	4,910	2,230	2,674	1,476	1,999	1,409	1,381
25	21,455	10,504	6,701	4,478	2,846	2,754	1,826	1,789	1,328	1,083
26	21,842	12,228	6,872	4,826	2,797	2,390	1,901	1,787	1,968	1,108
27	20,733	9,771	5,005	4,699	2,874	2,783	1,276	1,085	1,019	1,626
28	19,697	12,217	5,345	4,284	2,179	2,012	1,028	1,905	1,705	1,095
29	21,923	9,866	6,585	4,068	2,654	2,666	1,971	1,750	1,386	1,311
30	20,568	12,579	6,584	4,003	2,342	2,205	1,435	1,322	1,063	1,486

The quantitative of web usage of top ten are shown at Table-1. We can know that the Maximum quantitative web usage value is <http://google.com>. Next, <http://youtube.com> and <http://ptvcdn.com>.

The next process is to summary in each top ten web usage. One can sum quantitative web usage with using SQL language.

After we are received the top ten then we calculate then number of web usage in any minutes. Next step, we use the interpolation with the Cubic spline for fitting curve. The Figure-7 shows the result of a web page of the top ten.

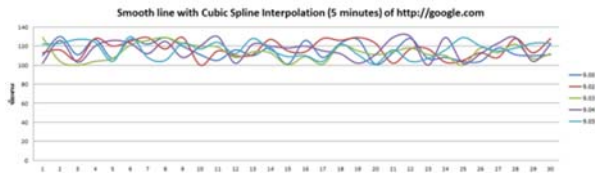


Figure-7. The line graph of <http://google.com> between 9.01 a.m. and 9.05 a.m.

Figure-7 presents the line graph of web usage of <http://google.com> between 9.00 a.m. and 9.05 a.m. However, those lines do not all data. We cannot show all data because the total of lines is 180 lines which are a lot

of lines. However, other lines have quantitative values between 100 and 140.

The results of the cache feature preparation

The Wolfram Alpha is online application (<http://www.wolframalpha.com/input/?i=interpolating+polynomial+calculator>). We use this application solving average values conversion to Cubic equations.

The next process is calculating the area under curve in each top ten. After that, one makes and calculates the normalization RC equations (N_RC). So, the results of N_RC of the top ten are shown at Table-2.

Table-2. N_RC values.

No.	The top ten	$(t_2-t_1)*c$	$\int_{t_1}^{t_2} f(t)dt$	RC (Reverse Cost)	N_RC
1	http://google.com	917096.47	161.15	1,350,785.05	1.00
2	http://youtube.com	1150811.61	86.56	1,293,718.24	0.95
3	http://ptvcdn.net	1344605.03	10.10	861,403.81	0.53
4	http://yting.com	818602.75	217.26	817,119.76	0.49
5	http://gstatic.com	940885.58	108.33	940,028.57	0.61
6	http://teennee.com	2160995.68	204.17	302,199.55	0
7	http://kapook.com	1499292.67	130.82	414,423.33	0.11
8	http://adobe.com	703483.25	207.64	701,751.33	0.38
9	http://dropbox.com	15466363.26	1114.97	746,326.89	0.42
10	http://4shared.com	625612.59	337.85	999,699.20	0.67

$$\int_{t_1}^{t_2} f(t)dt$$

Table-2 presents the outputs of $(t_2-t_1)*c$ values, values, Reverse Cost values and N_RC values.

However, we decide the top five from top ten that one evaluation from $D = w_1.N_LRU + w_2.N_RC$

equation. However, one calculates the N_LRU values in next procedure.

The results of the cache replacement and the hit ratio of ALRU model

We use the training sets from August 2013. The top five in each period is solve with C Language, which is developed for finding top five from the top ten any ten minute. The results are shown at Table-3.

**Table-3.** The hit ratios of ALRU model.

Time	Web ₁	Web ₂	Web ₃	Web ₄	Web ₅	Hit ratio
9.00-9.10	google	youtube	ptvcdn	yting	gstatic	29.15
9.11-9.20	google	youtube	ptvcdn	yting	gstatic	48.90
9.21-9.30	google	youtube	ptvcdn	adobe	dropbox	33.67
9.31-9.40	google	youtube	ptvcdn	gstatic	adobe	57.55
9.41-9.50	youtube	ptvcdn	Yting	gstatic	dropbox	63.83
9.51-10.00	google	youtube	ptvcdn	yting	4shared	61.11
10.01-10.10	google	youtube	ptvcdn	adobe	4shared	42.34
10.11-10.20	google	youtube	ptvcdn	yting	4shared	72.56
10.21-10.30	google	youtube	ptvcdn	yting	dropbox	69.80
10.31-10.40	google	youtube	ptvcdn	yting	4shared	70.35
10.41-10.50	google	youtube	ptvcdn	yting	dropbox	57.30
10.51-11.00	google	youtube	ptvcdn	4shared	dropbox	42.95
11.01-11.10	google	youtube	ptvcdn	yting	gstatic	65.50
11.11-11.20	google	youtube	ptvcdn	yting	dropbox	38.77
11.21-11.30	google	youtube	ptvcdn	yting	dropbox	64.62
11.31-11.40	google	youtube	ptvcdn	yting	kapook	30.04
11.41-11.50	google	youtube	ptvcdn	yting	teenee	51.95
11.51-12.00	google	youtube	gstatic	yting	teenee	69.12

Table-3 shows selecting the top five values and the hit ratio values from 9.00 a.m. to 12.00 a.m. In addition, the average value of the hit ratio is 53.8 percent. In addition, the maximum of the hit ratio is 72.56 percent (between 10:11 a.m. and 10:20 a.m.) and the minimum of the hit ratio is 29.15 percent that the situation was happen between 9:00 a.m. and 9:10 a.m.

However, twelve times for the hit ratio average value are higher than 50 percent. In addition, seven times for the average values of the hit ratio are less than 50 percent.

The results of LRU model

The ALRU model is compared with LRU model. We use the same data sets. This part shows the hit ratios of LRU model which is a popular model. We divide the testing sets any ten minutes and those data are executes with C Language. This application is written for testing the hit ratio of LRU model that this application is design for supporting five channels only. If we would like to run with less than or more than five channel then this application is adapted before to execute. There are the hit ratios of LRU model at Table-4.

Table-4. The hit ratios of LRU model.

Time	Hit ratio	Time	Hit ratio	Time	Hit ratio
9.00-9.10	8.21	10.01-10.10	8.12	11.11-11.20	6.70
9.11-9.20	18.25	10.11-10.20	8.12	11.21-11.30	4.35
9.21-9.30	7.65	10.21-10.30	6.70	11.31-11.40	9.31
9.31-9.40	12.76	10.31-10.40	14.72	11.41-11.50	53.34
9.41-9.50	8.00	10.41-10.50	5.30	11.51-12.00	9.10
9.51-10.00	9.30	10.51-11.00	3.24	11.11-11.20	40.76



Table-4 shows the hit ratio values of the LRU technique. The average value of hit ratio is 12.99 percent. However, In addition, the time gives the highest hit ratio value is from 11.41 a.m. to 11.50 a.m. (53.34 percent) and the time gives the lowest hit ratio value is from 10:51 a.m. to 11.00 a.m. (3.24 percent).

CONCLUSIONS

The ALRU model is created in this research for decreasing the downloading web pages from other web server. Many Mathematics Statistics are used in this model such as the Estimated Value, the interpolation with Cubic Spline, the First Order Condition (FOC), an interpretation of the first derivative, an interpretation of the second derivative and to find the area under the curve. In addition, the quantity web usage in the past is used to test in this model.

The objective of this research reduces downloading web pages from web servers in other locations and reduces the cost. However, the performance of the ALRU model will compare with the LRU model. In addition, this research not compares the performance of the time of processing.

The performance can be evaluated from the hit ratio. We divided the hit ratio any ten minute so we were received the hit ratio of the ALRU model equal 18 times. However, this model gives the hit ratio more than 25 percent in any periods and this model offers the hit ratio less than 40 percent in any periods too.

The advantage of ALRU model increases the hit ratio higher than LRU model and the disadvantages of this model are using a lot of the training sets, not supporting unpredictable events and spending a lot of time for calculation in each procedure.

REFERENCES

Akila W. 2013. Mathematics for Economics. Wellesley College, MA.

Gerhard B. and Günter Z. 2010. Introduction to Statistics and data analysis for physicists. Siegen University.

Hiroshi Y. 2008. Image interpolation by estimation and deconvolution of wavelet approximate subband. Master of Engineering Thesis. King Mongkut's University of Technology Thonburi, Bangkok.

Jaynes E.T. 2003. Probability theory: the logic of Science. Cambridge University press. Cambridge. UK.

Nipaporn S. 2009. Modeling of detected pleural thickening through thin plate spline interpolation. Master of Computer Science Thesis, King Mongkut's University of Technology North Bangkok.

Petrinovic D. 2009. Continuous time domain properties of causal cubic splines. Proceeding of ACM Signal Processing. 89: 1941-1958.

Press W.H. and et al. 1988. Cubic spline interpolation, numerical recipes in C: the art of scientific computing. Cambridge University Press.

Rutchawit J. 2010. Fundamental of Mathematical Economics: Optimization and Dynamic Analysis. Sukhothaihammarhirat Press. pp. 10-20.

Soros J. 2011. Computational techniques in food engineering. Ph.D of Food Engineering Thesis, Kasetsart University.

Wandee J. 2001. Approximating functions by cubic splines. Master of Mathematics Thesis. King Mongkut's University of Technology, Thonburi, Bangkok.