ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

ADAPTIVE WEB CACHING WITH INTERPOLATION AND WEB USAGE PATTERNS

Prapai Sridama[1], Somchai Prakancharoen[2] and Nalinpat Porrawatpreyakorn[1]
[1]Department of Information Technology, Faculty of Information Technology, King Mongkut's University of Technology North Bangkok, Thailand
[2]Department of Computer Science, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Thailand
E-Mail: prapaikmutnb@gmail.ac.th

ABSTRACT

The objective of this research increases efficiency of web cache memory. The Adaptive Web Caching with Interpolation and Web Usage Patterns (AWC-IWUP) simulation is investigated to solve for decision making about web objects replacement based on web usage and dynamic decision. More than 1,000,000 data sets of web usage between January 2013 and October 2013 from Walailak University are used for testing in this simulation. The AWC-IWUP simulation combines many mathematical statistics theories as follows: Expected value algorithm, Cubic spline technique, MarSpline technique, First Order Condition (FOC) and Bayesian algorithm. The experimental results of this research can conclude that the AWC-IWUP simulation is proposed replacement technique that the AWC-IWUP solves replacement problem of the web cache memory better than the LRU technique. The AWC-IWUP technique increases the hit ratios higher than the LRU technique approximately 25%.

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

INTRODUCTION

Web cache memory is used to reduce web traffic on web servers and network bandwidth providers (Cobb and ElAarag, 2008; Chetan, 2009; Koskela et al., 2009; Kumar and Norris, 2008; Chiang et al., 2007). The working of web cache memory is implemented at three levels: client level, proxy level and original server level (Chen, 2008, Chen 2007). However, the cache replacement is the most important of the web cache mechanisms (Koskela et al., 2008). Many researchers investigated web cache replacement techniques but the cache hit ration is not grew up much with caching schemes. The hit ratio is still not over 50% of the cache scheme (Lee et al., 2009; Jianhui etal., 2008; Abhari et al., 2006). The significant limit of web cache memory is the less web cache memory. So, the most researchers must consider both replacement approach and prefetching approach too. However, the allocation of web cache memory is not optimized (Acharjee, 2006; Huang and Hsu, 2008; Pallis et al., 2008; Feng et al.2009). Therefore, both replacement technique and prefetching technique should be created carefully to eliminate those limitations.

The simplest and most easy web cache management approach is Least-Recently-Used (LRU) algorithm (Cobb and ElAarag, 2008; ElAarag and Roman, 2009). However, this approach is still not efficient enough (Liu, 2009). The disadvantage of LRU algorithm ignores download latency and the size of web objects (Koskela et al., 2003). Thus, this research investigates new technique for efficient increasing of web cache replacement. AWC-IWUP is a new approach which combines many Statistics Mathematics such as Cubic Spline, MarSplines, Expected Value algorithm, FOC, Dynamic Programming with Bellman equation, Bayesian technique and Maximum Likelihood estimation. It is created for web usage forecasting that predicted from web usage in the past. In addition, this approach considers hit ratio, download latency and the size of web objects.

The remaining parts of this paper are arranged as follows. Related researches and theories are presented in Section 2. AWC-IWUP algorithm design is proposed in Section 3. The result of AWC-IWUP technique testing is shown in Section 4. Finally, Section 5 and 6 are presented the discussion and conclusion of the paper.

LITERATURE REVIEW

Related research

This section presents a comprehensive overview of different proposals for web cache replacement strategies on LRU foundation only. The LRU strategy removes the least recently referenced object. MineCache (Jarutut, 2006) is a replacement technique on LRU foundation. It increases hit ratio approximately 60-70 percentages. A disadvantage of this technique is slow process than LRU technique. LRU-Threshold by Abrams (Abrams et al., 1996) gives higher hit ratio than LRU technique if web cache is large size. Some algorithm, applied LRU (Fang et al, 2013) will remove the biggest object firstly if both objects are used least. However, LRU-Min algorithm (Carlos, 2013), Applied the Apriori technique (Fudaila, 2010) MD5, SHA-1, SHA-226 (Jakkraphan, 2005) are high performance if the web cache memory are large memory. Some techniques such as Applied Lru algorithms with Naïve Bay algorithm (Waleed, 2012), ICWCS technique (Waleed et al., 2012), Markov model for predicting web page accesses (Mukund and George, 2004), Mix model which applied GAM (William et al.,

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

2014) are presented on LRU too. Anywise, all techniques give hit ratios higher than LRU but not over 20 percentages. Some researches investigate techniques of Data Mining to compound with LRU such as Neuro-fuzzy system (Ali and Shamsuddin, 2009), Applied the Apriori technique and Bitmap indexing algorithm, ANN (Artificial Neural Network) (Ibrahim and Xu, 2004, Farhan, 2007) and replacement algorithm based on fuzzy logic (Calzarossa and Valli, 2003). Those researches predicted with complex algorithms but it could predict to correct less higher than LRU. However, they still use items for training and testing their simulations.

Theories

Interpolation is a technique for prediction about to estimation missing values (Akila, 2013). Some researches use the interpolation technique for to estimate some data have values equal zero that this research use interpolation for solving this problem too. However, the interpolation technique has many algorithms that each algorithm can fit curves to well base on of each data. This research investigates interpolation with cubic spline algorithm.

Cubic spline is a technique from three techniques (Soros, 2011). The profit of cubic spline is giving concave and convex values (Press et el., 1988) but other techniques cannot do that.

MarSplines is a model that is combined with the basis functions (Robert et al, 2009) and those functions must be the model parameters (estimated via least squares estimation). The MARSplines algorithm can search over the space of all inputs and can give predictor values.

Expected value is an average value which depends on the random variable (Gerhard and Günter, 2010) that one is interesting, it has two types that are: Expected value of discrete random variable and Expected value of continuous random variable. However, this research presents with Expected value of discrete random variable because it uses to calculate expected values in discrete time (Gharles and Laurie, 1997).

FOC is used to calculate for average slope values (Akila, 2013). The result of FOC algorithm can specify trend values. However, this research knows that the request of client is growing up or falling down at the time. In addition, FOC is used from some researches (Boris and Moritz, 2013, Chao et al., 2013) to approximate convex values and concave values.

Bayesian analysis applies Bayes's theorem (Liu, 2003, Zhu et al., 2005). It uses parameters with random variables and distributes probability of parameters.

Least-Recently-Used (LRU) algorithm (Waleed A., Siti M.S., and Abdul S.I., 2011) is general to implement the web cach replacement. This algorithm does not consider the size of web page and the download latency of web page. The LRU algorithm searches web pages which are least used at the present time.

METHODOLOGY

This research focuses on new technique for improvement of web cache memory only. It decreases bandwidth and the time of downloading.

The data sets of web usage were recorded with the squid program.

Those data sets were log files of Walailak University between January 2013 and October 2013 (approximately 1,000,000 records).

AWC-IWUP is investiaged in this studying. It can request the prediction which may happen in the future. The detail of elements is shown by Figure-1.

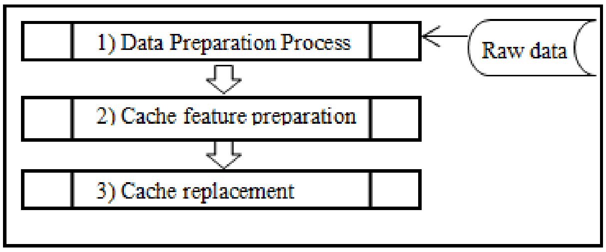

Figure-1. Elements of the AWC-IWUP simulation.

From Figure-1 elements of the WCO-CW technique are as follows:

Data preparation process

Data preparation process, it is processed as follows:

The cleaning, this section the observations from log files were changed to text files with Perl programming, after that noise elimination is under taken in the next step. It can be shown with Figure-2.

```
Wed Aug 21 06:52:55 2013   5439 10.52.0.35 TCP_MISS/000 0 GET http://imgcdn.ptvcdn.net
Wed Aug 21 06:52:56 2013    319 10.52.0.29 TCP_MISS/200 853 GET http://clients2.google.c
Wed Aug 21 06:52:56 2013   1119 192.168.74.246 TCP_MISS/404 491 POST http://somicroso
Wed Aug 21 06:52:56 2013   2498 10.52.0.35 TCP_MISS/000 0 GET http://imgcdn.ptvcdn.net
Wed Aug 21 06:52:57 2013      0 10.52.0.30 TCP_NEGATIVE_HIT/404 269 GET http://toolbar.li
Wed Aug 21 06:52:58 2013      1 10.52.0.30 TCP_HIT/200 1451 GET http://img.weather.weatl
Wed Aug 21 06:52:58 2013    241 10.52.0.30 TCP_MISS/200 769 GET http://datafeed.weathe
```

Figure-2. Example of a text file after the cleaning process

Figure-2 shows the example of web usage of users from Walailuk university between January and October 2013 that those data set are not changed to use in this research.

The summary process, the summation of web usage each webs are calculated in this process. It summarizes each web between 9.00 a.m. – 12.00 a.m. that was separated from each day. Finally, only top ten of webs are selected each day.

The dividing process, this part summarizes web usage in each top ten of web pages every one minute between 9.00 a.m. – 12.00 a.m.. However, the summary in this part were separated each day (Monday- Friday) wherefore that it can be shown with Table-1.

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

Table-1. Partial example of the web usage in a web of top ten.

Monday	The number of a web of top ten (in each second) between 9.00 a.m. – 12.00 a.m.								
	9.00	9.01	9.02	9.03	.	.	.	.	12.00
No.1	1,200	1,204	1,242	1,265	.	.	.	.	0
No.2	1,234	1,303	1,303	1,324	.	.	.	.	45
No.3	1,505	2,500	2,500	2,304	.	.	.	.	1,003
.	.	.	.	.	.	.	.	.	0
.	.	.	.	.	.	.	.	.	432
.	.	.	.	.	.	.	.	.	1,723
No.30	1,709	1,754	1,776	4,743	.	.	.	.	4,743

Figure-3 shows the web page usage of a website from top ten of webpages on every Monday (equal to 30 days) in the past. It can be seen as one cannot select a line from other lines now. This step, the 10 lines are obtained and the AWC-IWUP technique is created to solve this problem in the next step. Lastly, forecasting the web page usage is used.

Cache feature preparation. This section presents feature preparation for decision making.

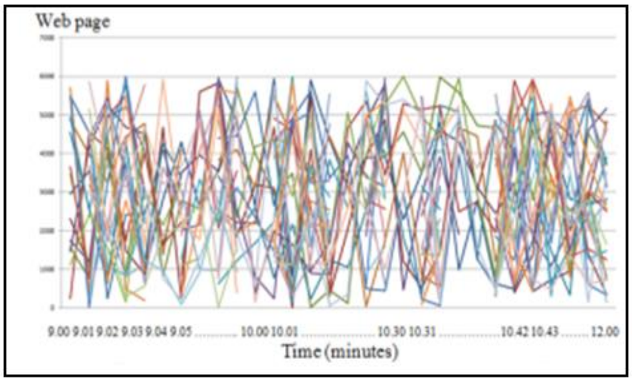

Figure-3. Example of web usage every Monday (30 days) between January 2013 and August 2013.

The estimation of agented values are calculated in this step. The Expected Value of random variables is used to estimate values. The Expected Value is seen in the Equation 1

$$E(x) = \int_{\infty}^{\infty} x f(x)\,dx \qquad (1)$$

Let x is the number of web usage in each minute of 30 days The f(x) is a function that can calculate with cubic spline. It is shown in the Equation (2).

$$f_i(x) = a_i x^3 + b_i x^2 + c_i x + d_i \qquad (2)$$

where $i = 1, 2, 3, \ldots, n$

Each function has 4 constant values that we do not know. Hence, if ones have n phases then ones do not know $4n$ values. However, the conditions are made to give the number of equations.

Adjacent two functions are equal at interior points. If interior points have n-1 points (x_1, x_2, …, x_{n-1}) then i is equal 3, 4, …, n.

$$a_{i-1}x_{i-1}^3 + b_{i-1}x_{i-1}^2 + c_{i-1}x_{i-1} + d_{i-1} = f(x_{x-1}) \qquad (3)$$

$$a_i x_{i-1}^3 + b_i x_{i-1}^2 + c_i x_{x-1} + d_i = f(x_{i-1}) \qquad (4)$$

Ones can get 2(n-1) or 2n-2 equations from this condition. For example, let i is equal 2 or x_i are connecting points between session 1 [x_0, x_1] and session 2 [x_1, x_2]. Both equations show at Equation (5) and Equation (6).

First session

$$f(x_1) = a_1 x_1^3 + b_1 x_1^2 + c_1 x_1 + d_1 \qquad (5)$$

Second session

$$f(x_1) = a_1 x_1^3 + b_1 x_1^2 + c_1 x_1 + d_1 \qquad (6)$$

The first function passes a point (x_0, $f(x_0)$) and the function at n passes the point of (x_n, $f(x_n)$). Hence, ones can get 2 equations. These equations are presented by Equation (7) and Equation (8).

$$f(x_0) = a_1 x_0^3 + b_1 x_0^2 + c_1 x_0 + d_1 \qquad (7)$$

$$f(x_n) = a_n x_n^3 + b_n x_n^2 + c_n x_n + d_n \qquad (8)$$

The first level derivative of adjacent functions that the values are equal at interior points.

So, ones can get n-1 equations from quadratic function ($f(x) = ax^3 + bx^2 + cx + d$). The first level derivative is shown by Equation (9).

$$f'(x) = 3ax^2 + 2bx + c \qquad (9)$$

For example, if the interior points have n-1 points (x_1, x_2, …., x_n-1) then the equation for i = 2, 3, …, n is shown at equation bellow.

$$3a_{i-1}x_{i-1}^2 + 2b_{i-1}x_{i-1} + c_{i-1}x_{i-1} = 3a_i x_{i-1}^2 + 2b_i x_{i-1} + c_i x_{i-1}$$

The second level derivative of adjacent functions that the values are equal at interior points.

Hence, ones can get n-1 equations from the second level derivative. This equation is shown by Eq.10.

$$f'(x) = 6ax + 2b \qquad (10)$$

For example, if i = 2, 3, …, n then $6a_{i-1}x_{i-1} + 2b_{i-1} = 6a_i x_{i-1} + 2b_i$.

The second level derivative of the first and the last functions are zero.

This equation is shown by Equation (11).

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

$$f'(x_0) = 6a_1x_0 + 2b_1 = 0 \quad and$$

$$f'(x_n) = 6a_1x + 2b_n = 0 \qquad (11)$$

If the cubic spline uses the matrix to solve equation then the matrix is large. In addition, to solve the equation is difficult. However, the values estimation of interpolation can use the polynomial by lagrange equation. It is shown at Equation (12).

$$f'(x) = \frac{x - x_i}{x_{i-1}x_i} f'(x_i) \frac{x - x_i}{x_{i-1} - x_i} + f'(x_i) \qquad (12)$$

where i = 1, 2, 3, …, n

To integrate two times can give $f(x_i)$ and two constants. These functions are presented by functions bellow.

$f(x_i) = f(x_{i-1})$ at the point of $x = x_{i-1}$

$f(x_i) = f(x_i)$ at the point of $x = x_i$

From the functions above can give Equation (13).

$$f_i(x) = \frac{f' x - x_i}{6x_i - x_{i-1}(x_i - x)^3} + \frac{f' x - x_i}{6(x_i - x_{i-1})}(x - x_{i-1})^3$$
$$+ \left[\frac{f(x_{i-1})}{x_i - x_{i-1}} - \frac{f'(x_{i-1})(x_{i-1})}{6}\right](x_i - x)$$
$$+ \left[\frac{f(x)}{x_i - x_{i-1}} - \frac{f(x)(x_i - x_{i-1})}{6}\right](x - x_{i-1}) \qquad (13)$$

The equation of $f'(x_{i-1})$ and $f'(x_i)$ from the condition of derivative level one of adjacent functions, that values are equal, shows at Equation (14).

$$f'_i(x_i) = f'_{i+1}(x) \qquad (14)$$

where $i = 1, 2, \ldots, n-1$

To solve Equation (12) by the derivative and to use relation of Equation (14) get Equation (15).

$$(x_I - x_{i-1}) f''(x_{i-1}) = 2(x_{i+1} - x_{i-1}) f''(x_i)$$
$$+ (x_{i+1} - x_i) f''(x_{i+1}) \qquad (15)$$
$$= \frac{6}{(x_{i+1} - x_i)} + f(x_{i+1}) - f(x_i) + \frac{6}{(x_i - x_{i+1})} + [f(x_{i+1}) - f(x_i)]$$

For example, let the data for estimating of interpolating values are 9.00=1,200, 9.01=1,204, 9.02=1,242 and 9.03=1,265 then let us to estimate by cubic spline interpolation at x = 9.015. The condition to estimate is seen in the Equation (11).

Hence, to solve the first point ($f'(9.01)$) by Equation (15) , ones uses the data from x_{t0}=9.00, x_{t1}=9.01 and x_{t2}=9.02.

$$= \frac{6}{(9.02 - 9.01)}[1{,}242 - 1204] + \frac{6}{(9.01 - 9.00)}$$

$$+[1{,}200 - 1{,}204]$$

or $0.04 f'(9.01) + 0.01 f'(9.02) = 20{,}400$

To solve the second point $f'(9.02)$) by Equation (14), ones uses the data from x_{t0}=9.01, x_{t1}=9.02 and x_{t2}=9.03.

$$(9.02 - 9.01) f''(9.01) + 2(9.03 - 9.01) f''(9.02) + (9.03 - 9.02) f''(9.03)$$

$$= \frac{6}{(9.03 - 9.02)}[1{,}265 - 1{,}242] + \frac{6}{(9.02 - 9.01)} + [1{,}204 - 1{,}242]$$

or

$$0.04 f''(9.02) + 0.01 f''(9.03) = -9{,}000$$

Thus to solve at the first point and the second point, ones get outputs bellow.

$$\begin{pmatrix} 0.04 & 0.01 \\ 0.04 & 0.01 \end{pmatrix} \begin{pmatrix} f''(9.01) \\ f''(9.02) \end{pmatrix} \begin{matrix} = \\ = \end{matrix} \begin{pmatrix} 20{,}4000 \\ 9{,}000 \end{pmatrix}$$

To solve the equation of $f'(9.01)$ and $f'(9.02)$, ones get outputs bellow.

$f''(9.01) = 170{,}000$ and

$f''(9.02) = 75{,}000$ To bring those outputs replace by Equation (12). It shows equation bellow.

$$f_{t2}(9.015) = \frac{170{,}000}{6(9.02 - 9.01)}(9.02 - 9.015)^3 + \frac{-75{,}000}{6(9.02 - 9.01)}(9.015 - 9.01)^3$$
$$+ \left[\frac{1{,}204}{9.02 - 9.01} - \frac{170{,}000(9.02 - 9.01}{6}\right](9.02 - 9.015)$$
$$+ \left[\frac{1{,}242}{9.02 - 9.01} - \frac{-75{,}000(9.02 - 9.01)}{6}\right](9.015 - 9.01)$$

= 1,224.526

Cache feature preparation

The MARSplines is used in this section for the automatic adaptation of values in each minute.

$$Y = f(x) + B_0 + \sum_{m=1}^{M} B_m h_m(x) \qquad (16)$$

The Eq.16 can separate to calculate with Eq.17 in any ten minute.

$$Y_i = \alpha_0 + \beta_{j=1} f(x_1) + \beta_{j=2} f(x_2) + \ldots \qquad (17)$$

Where ∝ is the value from the expected value.

β is the value that calculates with a spline algoritm and $\bar{x}$ is the average value from 30 days in each minute.

For example, Let $Y_1^{9.00-9.10am.}$ is $Y_{i(1)}$ value. So, MARSplines will calculate with Eq.17 ten times in each i. However the number of Y_i values are 18 numbers in this research.

Cache replacement

The slot of memory cache has five channels. Hence, web_2, web_3, web_4, web_5 and web_9 are chosen at 9.00 – 9.10 a.m.. Normally, this system transfers the average slope values of top ten web pages to the Cache Replacement Table (CRT). It can be shown in Table-2. However, this process compares the average slope values

ARPN Journal of Engineering and Applied Sciences

every fifteen minute. The Table-4 shows the CRT at 9.10 a.m.

Table-2. The CRT at 9.00 a.m.

Web	Web usage	FOC	Decision
http://google.co.th	1,200	0	
http://facebook.com	1,234	0	✓
http://youtube.com	1,505	0	✓
http://dropbox.com	1,632	0	✓
http://yahoo.com	1,320	0	✓
http://pantip.com	764	0	
http://kapook.com	903	0	
http://teenee.com	1,231	0	
http://manager.co.th	1,763	0	✓
http://matichon.co.th	597	0	

Table-3. The CRT at 9.11 a.m.

Web	Web usage	FOC	Decision
http://google.co.th	1,532	33,200	✓
http://facebook.com	1,214	-2,000	
http://youtube.com	1,312	-19,300	
http://dropbox.com	1,082	-55,000	
http://yahoo.com	1,721	40,100	✓
http://pantip.com	1,083	31,900	
http://kapook.com	1,321	41,800	
http://teenee.com	1,424	19,300	✓
http://manager.co.th	1,412	-35,100	✓
http://matichon.co.th	1,325	72,800	✓

From Table-3, the LRU values are considered to select webpages when average slop values are equal. Hence, web_1, web_5, web_8, web_9 and web_{10} are chosen at 9.00 – 9.10 a.m.

THE RESULT OF EXPERIMENT

This section shows the result of this research. The numbers of data set in this test are divided into 4 groups. Those data sets are presented in Table-4. The simple random sampling technique is used for sampling because all data sets have an equal chance.

Table-4. The Result of experiment.

Number of data set	**Hit ratio** (%)	**Download saving** (Mb)
100	68.18	268.98
1,000	71.83	2,834.72
10,000	73.10	30,559.56
100,000	68.41	283,749.84

The table above shows the result of the experiment from this research. This result can be seen the percentage of hit ratio of all group from any data set (100, 1,000, 10,000 and 100,000) more over 50%. In addition, AWC-IWUP technique can decrease the quantity of downloading from other web servers.

CONCLUSIONS

This article has given exhaustive outputs of the AWC-IWUP technique. The percentage of the hit ratio and the quantity of web download saving from any groups of data sets with AWC-IWUP technique are not over 75%. However, the hit ratios of the AWC-IWUP technique are higher than LRU (the hit ratios are not over 50%).

In this paper, we have proposed a new algorithm called the AWC-IWUP technique for web usage forecasting. The algorithm improves LRU technique with the mathematic and statistics algorithms. The AWC-IWUP algorithm reduces bandwidth and reduces the quantity of downloading from external web servers more than the LRU algorithm. We are investigating about a new algorithm which indicates the hit ratio is higher than the LRU technique. The AWC-IWUP technique is a technique to manage the web cache memory. The process of web page selection for loading into web cache memory is the important thing of this technique because all downloading has charges and the webcache memory is limited with the size of memory. However, if the AWC-IWUP technique selects to match with the wanting of users then the hit ratio increases. Lastly, we will research a new technique to enhance efficiency of web cache replacement.

REFERENCES

Abhari A., Dandamudi S.P. and Majumdar S. (2006). Web Object- Based Storage Management in Proxy Caches. Journal of Generation Computer Systems, 22(1-2), pp. 16--33.

Abrams M., Standridge and Fox E. (1996). Chaching proxies: limitations and potentials. Proceedings of the International World Wide Web Conference, p. 4.

Acharjee U. (2006). Personalized and Artificial Intelligence Web Caching and Prefetching. Master Thesis, University of Ottawa, Canada.

Ali W. and Shamsuddin S.M. (2009). Intelligent client-side web caching scheme based on least recently used algorithm and neuro-fuzzy system. International Symposium on Neural Networks (ISNN), 6, pp. 70--79.

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

Boris Houska and Moritz Diehl. (2013). Nonlinear Robust Optimization Via Sequential Convex Bilevel

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

Programming. Journal of Mathematical Programming, 142(1-2), pp. 539-577.

Calzarossa and Valli G. (2003).A fuzzy algorithm for web caching. Journal of Simulation Series, 35-4, pp. 630-636.

Carlos R. (2013). VFC for wikis and web caching. Phd. Thesis of Information systems and Computer Engineering, Lisboa University.

Chao Ding., Defeng Sun. and Jane J. Ye. (2013). First Order Optimality Conditions for Mathematical programs with Semidefinite Cone Complementarity Constraints. Journal of Mathematical Programming.

Chetan Kumar. (2009). Performance Evaluation for Implementations of a Network of Proxy Caches. Journal of Decision Support Systems, 46, pp. 492-500.

Chen H.T. (2008). Pre-fetching and Re-fetching in Web Caching Systems : Algorithms and Simulation. Master Thesis, Trent University, Peterborough, Ontario, Canada.

Chen T. (2007). Obtaining the Optimal Cache Document Replacement Policy for the Caching System of an EC Website. Journal of Operational Research, 181(2), pp. 828.

Chiang I.R., Goes P.B. and Zhang Z.(2007). Periodic Cache Replacement Policy for Dynamic Content at Application Server. Journal Decision Support Systems, 43(2), pp. 336-348.

Cobb J., and ElAarag H. (2008). Web Proxy Cache Replacement Scheme based on Back-propagation Neural Network. Journal of systems and software, 81, 2008, pp. 1539-1558.

ElAarag H., and Romano S. (2009). Improvement of the neural network proxy cache replacement strategy. Proceeding of the 2009 Spring Simulation Multiconference (SSM'09), pp. 90.

Fang Chao. (2013). Fast convergence caching replacement algorithm based on dynamic classification for content-centric networks. ScienceDirect Trans, on China Universities of Posts and Telecommunications, 20-5, pp. 45-50.

Farhan. (2007). Intelligent web caching architecture. Master of Computer Science and Information System Thesis, UTM University.

Feng W., Man S. and Hu G. (2009). Markov Tree Prediction on Web Cache Prefetching. Journal of Software Engineering, Artificial Intelligence (SCI), 209, pp. 105-120.

Freund J.E. (1992).Mathematical Statistics. fifth, London, Prentice-Hall International.

Fudaila D. (2010). An algorithm for mining frequent itemsets based on the similarities of transactions. Journal of Network and Computer Applications.

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

Gharles M.G., and Laurie S.J. (1997). Introduction to probability. American Mathematical Society.

Huang Y.f., and Hsu J.M. (2008). Mining Web Logs to Improve Hit Ratios of Prefetching and Caching. Journal of Knowledge-Based Systems, 21(1), pp. 62-69.

Ibrahim T.I., and Xu C-Z. (2004). A Keyword-Based Semantic Pre-fetching Approach in Internet News Services. Journal of Knowledge Management, 16-5, pp. 601--611.

Jakkraphan M. (2005). A study of replacement algorithms and object placement algorithms for decentralized web cache. Master of Computer Science Thesis, Kasetsart University.

Jarutut P. (2006). An efficiency improvement model of web caching with web log mining: case study the PSU campus network. Master of Computer Science Thesis, Prince of Songkla University.

Jianhui L., Tianshu X., and Chao Y. (2008). Research on WEB Cache Prediction Recommend Mechanism Based on Usage Pattern. First International Workshop on Knowledge Discovery and Data Mining(WKDD), pp. 473-476.

Koskela T., Heikkonen J. and Kaski K. (2003). Web cache optimization with nonlinear model using object feature. Journal of Computer Networks, 43(6), pp. 805-817.

Kumar C. and Norris J.B. (2008). A New Approach for A Proxy-level Web Caching Mechanism. Journal of Decision Support Systems, 46(1), pp. 52-60.

Koskela T., Heikkonen J. and Kaski K. (2008). Web Cache Optimization with Nonlinear Model Using Object Feature. Journal of Computer Networks, 43(6), pp. 805-817.

Lee H.k., An B.S. and Kim E.J. (2009). Adaptive Prefetching Scheme Using Web Log Mining in Cluster-Based Web Systems. International Conference on Web Services (ICWS), 2013 First International Conference on. IEEE , pp. 903-910.

ARPN Journal of Engineering and Applied Sciences

Liu J.S. (2003). Bayesian Methods in Biological Sequence Analysis. Handbook of Statistical Genetics 2nd Edition, John Wiley & Sons, Ltd.

Liu Q. (2009). Web Latency Reduction with Prefetching. Phd. Thesis, University of Western Ontario, London, UK.

Mukund D., and George K. (2004). Selective Markov models for predicting web page accesses. Journal of Internet Technology, 4, pp. 163-184.

Pallis G., Vakali A. and Pokorny J. (2008). A Clustering-based Prefetching Scheme on a Web Cache Environment. Journal of Computers and Electrical Engineering, 34(4), pp. 309-323.

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

Soros J. (2011). Computational techniques in food engineering. Phd. of Food Engineering Thesis, Kasetsart University.

Waleed A., Siti M.S. and Abdul S.I. (2012). Intelligent Naïve Bayes-based approaches for Web proxy caching. Journal of Knowledge-Based Systems, 31, pp. 53-579.

Waleed A., Siti M.S. and Abdul S.I. (2011). A Survey of Web Caching and Prefetching. ICSRS Publication.

William S.R., Alain Z.F. and Kristynn S.J. (2014) Using generalized additive (mixed) models to analyze single case designs. Journal of Analysis and Meta-Analysis of Single-Case Designs, 52, pp. 149-178.

Zhu S., Dellaert F. and Tu Z. (2005). MCMC for Computer Vision. A Tutorial at 10th International Conference on Computer Vision 2005, Beijing.