



UNIVERSITI PUTRA MALAYSIA

**WEB ACCESS ANALYSER FOR
DISTANCE EDUCATION ENVIRONMENT**

RODZIAH BT HJ. ATAN

FSAS 1998 31

**WEB ACCESS ANALYSER FOR
DISTANCE EDUCATION ENVIRONMENT**

By

RODZIAH BT HJ. ATAN

**Thesis Submitted in Fulfillment of the Requirements for the
Degree of Master of Science in the Faculty of
Science and Environmental Studies
Universiti Putra Malaysia**

August 1998



DEDICATION

Bismillah.....

Alhamdulillah.....syukur ke hadratNya kerana dengan limpah kurniaNya saya dapat menghabiskan tesis Master ini.

Ayahanda Hj. Atan bin Awang dan bonda Hjh. Halijah bte. Awang, jasa kalian berdua mengasuh dan membesarkan anakanda tidak sia-sia dan tidak akan dipersiakan.

Suami tersayang Mohd. Nazri bin Hj. Yusuf, kesabaran dan pengorbananmu selama ini telah menjadi kenyataan. Terima kasih di atas segalanya.

Abang-abang yang dihormati, Abang Kaya, Abang Li dan Abang Nuar, dorongan dan nasihat selalu menjadi pegangan hidup.

Kakak-kakak yang dikagumi, Kak Long, Kak Ngah, Kak Uda dan Kak Cik, kelembutan serta kebijaksanaan selalu menjadi contoh peribadi yang mulia.

Ayah dan Ibu mertua, ipar serta anak-anak saudara, kalian adalah semangat dan penghibur di kala kesunyian.

ACKNOWLEDGEMENTS

I wish to express extreme gratitude to my supervisor, Dr. Md Yazid Mohd Saman for his highly valuable guidance and stimulating suggestions that were enabled me to complete this thesis. A thousand thanks goes to my co-supervisors, Dr. Md. Nasir Sulaiman and Dr. Ali Mamat who has given me their full commitment and constructive views for me to accomplish the project.

Special thanks goes to Universiti Putra Malaysia for granting me PASCA fellowship schemes which funding me with a large amount of money during my study.

Thank you to all my colleagues especially Kak Zura, Along, Karen, Maya and Kak Liza for the endless support and sweet -sour memories we have been through each day in UPM. Heartfelt thanks goes to all Computer Science Department staffs for their hospitality and patient in providing me facilities that make it possible for me to finish the project.

Last but not least, a very special gratitude goes to my beloved parents, Tn. Hj. Atan Awang and Hjh. Halijah Awang, my brothers and sisters especially Kak Cik and family who has provided me shelters during my study. Finally, not forgetting my beloved husband, a person so special with endless support and encouragement to me all these years. Thank you all and may God bless all these individuals for their kindness.



TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vii
LIST OF FIGURES	viii
LIST OF ABBREVIATION	x
ABSTRACT	xii
ABSTRAK	xiv
CHAPTER	
I INTRODUCTION	1
Distance Education Technology	2
Problem Statement	4
Objectives of the Study	6
Problem Solution	7
Thesis Overview	8
II LITERATURE STUDY	10
Distance Education Technology	11
Distance Education and the Internet	15
Module Delivery Mode Technologies	18
Visualisation Concepts	21
Internet and the World Wide Web	24
Client-Server and Network Computing	26
WWW Browser and HTTP Server Data Communication	29
Combined Log File System	31
Distance Education in UPM	32
Implementation History	33
Implementation Process	34
A Review on Network Authentication System	39



PGP/PEM Encryption	39
Kerberos Network Authentication System	40
Terisa System	41
A Review on Access Analyser Systems	42
FTPWebLog	43
Wwwstat 2 0	45
Wusage	47
AccessWatch	49
Musage	50
III METHODOLOGY	52
Research Plan Overview	52
Model Development Approach	54
CGI Programming Concepts in DLAnalyser	56
Perl The Language for DLAnalyser	57
Access Log File for DLAnalyser	59
Access Authentication for Distance Module	61
User Access Control Implementation	63
Per-Directory Access Control	64
Accesses to Authenticated Files	68
IV SYSTEM DESIGN	71
System Requirement Definition	71
System Conceptual Design	73
Function-oriented System Design	75
System Architecture	76
Function Detail Description	77
HTML Interface (input1.html)	77
Daily Access Analysis (log_harian)	78
Weekly Access Analysis (log_minggu)	79
Monthly Access Analysis (log_bulan)	81
User-defined Period (semak_tarikh)	82
Analysing Accesses (analisa)	83
System Platform	84
V RESULTS AND DISCUSSION	85
Submitting User Inputs	86
Summarisation Outputs	88
VI CONCLUSION AND FUTURE RESEARCH	96
DLAnalyser Development Summary	97
Authentication Process	98

Visualisation of Access Summary	99
Future Work	99
Implementation Enhancements	100
Related Topics	100
BIBLIOGRAPHY	102
APPENDICES	
A HTML Form Directive Options	106
B Combined Log System Design	109
C Access Analyser Systems Summaries	110
D HTML Interface Source Code	112
E System Installation and Configuration Description	115
VITA	117



LIST OF TABLES

Table		Page
1	Open Universities World Wide	12
2	Development Phases Summary	56
3	Functional Specification	74
4	System Configuration	117



LIST OF FIGURES

Figure		Page
1	Basic Description of Visualisation Process	21
2	An Integrated Visualisation Model	23
3	Web Client-Server Structure	27
4	Data Flow in CGI Programming Environment	30
5	UPMnet Network Schematic Diagram	37
6	A Telnet Session with Kerberos System	41
7	GraphFTPWebLog Summary Output	44
8	FTPWebLog Summary Output	45
9	Wwwstat Textual Access Summary Report	46
10	Wusage Monthly Access Report for Ibnu sina Web Server ...	48
11	Wusage Daily Access Report for Ibnu sina Web Server	48
12	AccessWatch Graphical Hourly Statistic	49
13	AccessWatch Page Demand Textual Report	50
14	Research Activities Summary	53
15	Incremental Development Approach	55
16	Traditionally Compiled Program and Perl Interpreter Processes	58
17	Sample Data in Access_Log File	61
18	Communication between Browser and Restricted Files	62



19	Path and File Mode for <i>Conf</i> Directory in FSAS NCSA Server	64
20	Access Control Process	65
21	Htaccess File	66
22	.Htpasswd File	67
23	Username and Password Dialog Box	69
24	Data Flow for Authentication Process	70
25	DLAnalyser System Environment	72
26	DLAnalyser Conceptual Diagram	73
27	DLAnalyser System Design	75
28	DLAnalyser System First Level Chart	76
29	HTML Interface	86
30	Listed Subjects	87
31	User Access Frequency Summary Report	90
32	Textual Summary Report	91
33	Graphical Version of User Access Frequency	92
34	Graphical Version of User Access Percentage	93
35	User Access Detail	94



LIST OF ABBREVIATIONS

WWW or Web	World Wide Web
NCSA	National Center for Supercomputing Application
CGI	Common Gateway Interface
HTML	Hypertext Markup Language
FTP	File Transfer Protocol
IDEAL	Institute for Distance Education and Learning
UPM	Universiti Putra Malaysia
STPM	Sijil Tinggi Pelajaran Malaysia
TCP/IP	Transfer Control Protocol/Internet Protocol
UPMnet	UPM Network
ATM	Asynchronous Transmission Mode
WAIS	Wide Area Information Service
CERN	European Laboratory for Particle Physics
URL	Uniform Resource Locator
HTTP	Hypertext Transfer Protocol
TCP	Transfer Control Protocol
MIME	Multipurpose Internet Mail Extension
DNS	Domain Name Server
FSAS	Faculty of Science and Environmental Studies
PEM	Privacy Enhanced Mail



HTTPd	Hypertext Transfer Protocol daemon
S-HTTP	Secure Hypertext Transfer Protocol
CLF	Common logfile format
SSI	Server Side Include
SDLC	System Development Life Cycle
Perl	Practical Extraction and Report Language
PGP	Preety Good Privacy
DLAnalyser	Distance Learning Access Analyser



ABSTRACT

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirements for the degree of Master of Science.

WEB ACCESS ANALYSER FOR DISTANCE EDUCATION ENVIRONMENT

By

RODZIAH BT. HJ. ATAN

August 1998

Chairman: Dr. Md Yazid Mohd Saman

Faculty: Science and Environmental Studies

Distance Education implementation has been greatly influenced by the changing of the current sophisticated technologies. Internet and World Wide Web (WWW) have enabled the rapid growth of distance education technology. The Internet is a huge database that connects most of the information source centers worldwide. This technology is equipped with the existence of WWW that ease accesses to the sources.

With these technologies, distance education can be implemented properly. An approach been developed is the implementation of distance education through WWW. Learning modules are kept in a Web server while the student can access the



modules from scattered terminals linked to the server via the network. It is a method of module delivery to the students that which are easier and faster.

Even though distance education requires individual capability to manage the time and learning effort, the educators still have the responsibilities in evaluating their students' achievements. It is easier to evaluate on-campus students compared to distance learners. If the educators can evaluate on-campus students through their attendance and work performance, an appropriate approach to evaluate out-campus students' effort and commitment is necessary.

There is a file that is specially created by a Web server which keeps and lists all access requests and transaction information made to the files kept within the server. This particular file is known as an "access_log" file. There has been a substantial effort to manipulate the file, for example, this study used "access_log" file to filter out information to produce students' distance module access statistics which is able to assist distance educators to evaluate distance students' commitment as well as to measure modules' design efficiency.

Access restriction is one of the lightest security alert approaches used by a Web server. The study of module access analysis has adopted this approach as its main processing attribute. The study intends to peel the problems regarding learning module efficiency and distance students' commitment measurements through statistical access analysis approach.



ABSTRAK

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Master Sains.

PENGANALISIS CAPAIAN “WEB” BAGI PERSEKITARAN PENDIDIKAN JARAK JAUH

Oleh

RODZIAH BT. HJ. ATAN

Ogos 1998

Pengerusi: Dr. Md Yazid Mohd Saman

Fakulti: Sains dan Pengajian Alam Sekitar

Perlaksanaan Pendidikan Jarak Jauh atau PJJ telah banyak dipengaruhi oleh perubahan arus teknologi canggih masakini. Pengenalan kepada “Internet” dan “World Wide Web” (WWW) telah membantu perkembangan teknologi pendidikan jarak jauh. Internet merupakan pangkalan data yang menghubungkan pusat-pusat sumber maklumat di seluruh dunia. Teknologi ini telah diperlengkapkan lagi dengan kewujudan WWW yang memudahkan pencapaian sumber dilakukan.

Dengan teknologi ini, kaedah Pendidikan Jarak Jauh (PJJ) dapat dilaksanakan dengan lebih sempurna. Satu kaedah yang sedang giat dibangunkan adalah perlaksanaan PJJ melalui WWW. Modul pembelajaran disimpan di dalam pelayan



Web manakala capaian pelajar dilakukan ke modul-modul ini melalui terminal-terminal pelanggan dari serata tempat yang disambungkan ke pelayan tersebut melalui jaringan. Ia merupakan pendekatan mod penghantaran modul kepada pelajar yang lebih mudah dan cepat.

Walaupun pembelajaran jarak jauh memerlukan kepakaran individu dalam mengurus masa dan usaha pembelajaran masing-masing, namun begitu para pengajar juga tidak terlepas daripada tanggungjawab untuk menilai pencapaian pelajar-pelajar mereka. Penilaian dapat dilaksanakan dengan mudah jika pelajar tersebut berada di dalam kampus tetapi agak sukar untuk merealisasikan perlaksanaannya bagi pelajar di luar kampus. Sekiranya pengajar menilai komitmen pelajar dalam kampus melalui kehadiran dan hasil kerja yang dilakukan, satu kaedah yang tertentu adalah perlu untuk menilai daya usaha dan komitmen para pelajar luar kampus.

Terdapat satu fail yang dibina secara khusus oleh pelayan Web bagi menyimpan serta menyenarai kesemua maklumat permintaan dan perpindahan capaian fail yang disimpan di dalamnya. Fail penyimpanan data ini dikenali sebagai fail “access_log”. Terdapat banyak usaha dilakukan bagi memanipulasi fail ini, antaranya, kajian ini telah menggunakan fail “access_log” untuk mencerna maklumat bagi menghasilkan statistik capaian pelajar terhadap modul pembelajaran jarak jauh bagi membantu pengajar jarak jauh menilai komitmen yang ditunjukkan oleh pelajar jarak jauh disamping menjadi pengukur keberkesanan rekabentuk modul.

Kekangan capaian adalah merupakan satu pendekatan keselamatan yang paling longgar digunakan di dalam pelayan Web. Kajian ke atas analisis capaian modul telah mengadaptasi pendekatan ini sebagai satu atribut yang terpenting di dalam pemrosesannya. Kajian ini bertujuan untuk mengupas permasalahan penilaian tahap keberkesanan modul pengajaran dan komitmen pelajar jarak jauh melalui kaedah analisis statistik capaian modul pembelajaran.

CHAPTER I

INTRODUCTION

Summarised and visualised data is more easily understandable than browsing through a collection of raw data. An analysis system for Web accesses is one way of summarising and visualising a collection of data. This approach is adapted by various programs emerged to provide statistical analysis of World Wide Web (WWW) access log. These programs are used to list the occurrence of accesses in a detailed and easy to view form. A fully utilised network resource can lead to the development of a useful access analysis system.

Distance Education has evolved for many years. Technological advancement in computing and networking has changed the implementation of distance education these days. Web pages are been widely used by many universities and open learning institutions as a transmission medium for course materials. It is beneficial to visualise accesses made by the students. A better understanding of student's access behaviour can lead to a more efficient evaluation of the student's achievement and materials published.



Distance Education Technology

Recent years have seen rapid development in information technology. Many applications were build, used and maintained in a globally networked environment. Distance Education is not a new thing in educational theory. It was considered born in year 1892 when the University of Wisconsin used the term in its correspondence courses catalogue.

Distance education technologies are expanding at an extremely rapid rate due to the fast sophistication of computer and information technology systems (Sherry, 1996). The technologies employed by a particular distance learning system have a direct impact on the number of sites supported, the instructional media supported, the nature of the interaction, the level of quality, the type of system whether it is private or public and the security and confidentiality (Umich, 1995).

The main objective of distance education is to allow students remotely sited to access course materials. These materials are extracted and understood in their own way to add the knowledge to be tested at the end of semester. New technology has enabled these materials to be passed world-wide. It is the intention of the designers of the World Wide Web (WWW) to provide accesses to the information resources of the Internet through easy-to-use software which operated in a consistent manners (Berners-Lee, 1992). Whilst these were achieved, the implementation of the Web pages has changed the traditional approach of distance education, i.e. from sending

materials to the students by mail to accesses to the materials using computers linked to the materials kept in the server

The World Wide Web or the WWW is a world wide string of computer databases which use a common information retrieval architecture It also offers a means of transparently transporting all forms of data, (i e video, text and image) concurrently Integrating the use of the Web into distance education has more to offer than simply fast data transmission It offers more convenience in communication for those who have access (Ferguson, 1996) Modern technology supports a wide range of interactive communication between the instructors and the students Teleconferencing, chat and CUSeeMe, are a few examples of the facilities supported by the technology

In investigating the student access towards the materials, there is a log file that logs all the transactions done by the students The National Centre for Supercomputing Application's (NCSA's) HTTP 1.0 server has created default file called *access_log* file which records relevant data into formatted fields This file is useful in many ways to track user access behaviour Various applications manipulate this file to solve various series of problems The sample log entry below shows a line of data stated in an *access_log* common file format

```
202 184 29 128 - - [15/Oct/1997:11:29:50 -0800] "GET /~rodziah/const/mainmenu.html HTTP/1.0" 200 3452
```

```
202 184 29 128 - gee [15/Oct/1997:11:29:50 -0800] "GET /~rodziah/const/mainmenu.html HTTP/1.0" 200 3452
```

↳ the authentication name field (the 3rd field)

The third field of the entry represents the authentication user name. The dash means that no entry was made to that particular field. Most WWW page developers ignore the field and it is most probably because they do not want to authenticate accesses to the sites. It is different when implementing access analysing system for distance education environment. The users have to be authenticated so that they can be traced individually and this is the main objective of the study.

Problem Statement

All successful distance education has to overcome the problems associated with the instructor and the student being separated by geographical distance (Ferguson, 1996). This distance complicates building a trustful and constructive student-teacher relationship. The initiation of distance education must include the transition from teacher-centred to student-centred instruction, support of faculty, and the knowledge of how to restructure the coursework to meet the needs of the new world-cohort of learners (Parker, 1996). This gives the distance students chances to examine thinking and learning processes.

Even though the paradigm has changed from teacher-centred to student-centred learning, the teachers still have to do their usual monitoring task. This is to ensure that the students were on the right track. It was stated by Parker (Parker, 1996) that “the syllabus of distance course must become the road map and the student become the driver when instruction is mediated by technology”.

Although there are continuous academic evaluation for student's achievements, it is still not enough. Distance educators should not jump to the conclusion over their students by using the academically evaluated results only. Harrington-Luker as cited by Aloia (Aloia, et al., 1996) said that, "we have learned that you have to encourage learning by doing, that the learning has to be relevant to classroom experience and that you need to model good practice". The statement shows that distance educator is more than just a module writer.

As part of distance students' evaluation process, measuring students' commitment, and planning and evaluating distance modules, it is essential to study students' module accesses traffic. There are systems built to summarise accesses made to a collection or particular server, but none were built to particularly to solve distance education material access behaviour. It is believed that there exist a relation between student access behaviour with the understanding level and the effectiveness of the learning modules.

In addition to the students' formal academic evaluation process, there are several issues that should be discussed. In dealing with distance learning environment, it is normal when people ask questions on the effectiveness of Internetly-published distance education material. There are also question regarding the performance of a student compared to the learning material pages, and the criterion of a well-designed Web page to meet the highest understanding level of the students. These questions are simple but it takes some time before they can be answered.

The initial step is to plot the students' access behaviour. Students' access behaviour covers questions of the accesses capacity made by the students to the materials published, who has made the most access, what is the rate of bytes transferred per student access, or what is the time most accesses are made. Distance educators also need to know how their module is being utilised, does the module can easily be understood or does the module is being heavily downloaded due to the many graphics been presented. These questions require an access analyser to give accurate answers.

Objectives of the Study

The main objective of the study is to develop an analysing system for distance education materials access. The system developed will summarise and generate statistical reports of accesses made to the learning modules.

Upon achieving the objective, several other important issues are also been studied. The issues covered are the study on distance education implementation over the WWW and the identification of access information to be analysed. The access analyser system for distance education course materials generates and displays statistics based on individual access. User identification is the key information to be processed by the system.

Economic factors and user identification are two main reasons for the implementation of user authentication. User access restriction has been embedded to

most distance education modules to identify users through their username. Access restriction is also important in the system as it produce user identification field in access_log file.

Problem Solution

The main focus of the study is to analyse accesses of students to a particular course material and produce statistics out of the analysis. Traditional approach of evaluating students' achievement based only from marks and grades that the students get. The evaluation approach can be enhanced with access statistics of a particular student. The access statistic of a particular student to a particular subject will give a wider range of evaluation input.

To generate a particular person's access behaviour statistic to a document, a unique field that can differentiate a user from another should be identified. The study upon NCSA's server access_log file shows that user login identification is stored every time an authenticated user accesses a restricted document. The authenticated user login is a unique attribute to be analysed by the system.

User access restriction is embedded to distance module subjects. The embedded function is vital for further processing events as it supply login information data in access_log file. The field will be the key data throughout the access analysis process.

The system is developed using a programming language suitable for text manipulation and report generation. The language used is Perl programming language. Perl offers many features for data processing and report generating.

Thesis Overview

The thesis is organised in such a way so that the reader can get clear view over the objectives of the study. The current chapter gives a brief introduction about distance education and how technological growth has changed the approach of delivering learning materials to remotely sited students. It is also focused on the use of Web pages as the main delivering method. Then, there are issues concerning the instructors' responsibility over guiding their learners. The instructors have to aware of the contents of their materials. They should know how their materials were utilised and they also should know the correlation between students' achievements and the modules they have created. These need to be solved by visualising the trend of students' access traffic. This is the main starting point of the purpose for the study.

The next chapter describes literature surveys and background study on related work. It summarised the problems to be solved and the approaches taken by the systems. These works are then synthesised to show their strengths and how these systems are not suitable to solve the stated objectives of this study. The background study also covers the example of system development methodology adapted by the systems and some explanation on the terms and concepts used by the systems.