



UNIVERSITI PUTRA MALAYSIA

**DEVELOPMENT OF AN ENVIRONMENT FOR
VIRTUAL INFORMATION RETRIEVAL
IN UPMNET**

HANAFIZAN HUSSAIN

FSAS 1997 5

**DEVELOPMENT OF AN ENVIRONMENT FOR
VIRTUAL INFORMATION RETRIEVAL
IN UPMNET**

By

HANAFIZAN HUSSAIN

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

January 1997



Dedicated to:

My father:

Hussain Darus

My mom:

Siti Hagar Hashim

My elder brother:

Huzeri

My brothers:

Harisul, Hasnoor Hanif, Hasnoor Amir

My sister:

Haslinda

And my fiancée:

Samsuri Hashim

I love all of you



ACKNOWLEDGEMENTS

In the name of Allah, the Beneficent, the Merciful.

First and foremost, I would like to express my gratitude to Allah, The All-Mighty. It is with His ascendancy, I have completed this study.

Next, I would like to express my earnest gratitude and heartfelt thanks to the chairman of my supervisory committee, Associate Professor Dr. Abu Talib Othman, for his guidance and support throughout my study. He introduced me to the field of networks and laid the foundation of the virtual campus, on which this thesis is based.

It is also a great honour and pleasure to acknowledge Dr. Abdul Azim Abdul Ghani and Dr. Ismail Abdullah members of the supervising committee, for their technical support, helpful suggestions and insight.

I am grateful for the assistance of many people, specially Adzman Shah Mohd Kasim, who introduced me to the world of Internet; to Encik Mohd Zul Mohd Yusoff, Encik Mohd Faisal, Encik Rosmi and other staff of the Computer Centre, for letting me use the facilities. My heartfelt thanks to Encik Khalil



Awang and Puan Saadiah Yahya for their generous help in spending a considerable amount of time advising me on the organisation of the thesis while doing their PhD in UPM.

I would like to thank to Encik Nor Azman, Encik Suwardi, Encik Nasir, Encik Aziz, Encik Hazli and the other Computer Science Department staff for their help. Thank you for all your valuable time extended to me during this study.

To my friends, Ummu Salma, Hazura, Wan Bayani, Paizah, Fadhilah, Mazlah, Haryati, Noraini, Asmah, and my housemates, I would like to express my thanks for your ideas and help. I really appreciate it. To all the lecturers of Computer Science Department, my thanks for their support and guidance.

Last but not least, I must mention the tremendous support I received from my parents and family and also my fiancée; Samsuri Hashim. I could not have finished this thesis without their support.



TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS	xii
ABSTRACT	xv
ABSTRAK	xvii
 CHAPTER	
 I INTRODUCTION	 1
Computer Network	1
Information Retrieval	10
Objectives of the Study	15
Structure of the Thesis	18
 II LITERATURE REVIEW	 20
Historical Perspective of Internet and Related Networks	20
Internet Tools to Retrieve Information	30
On-Line Discussion	30
Remote On-Line Databases	36
Real-time Discussion	48
UPMNET	52
Learning on Demand	57
Existing Retrieval System in UPM	60
Conclusion	61



III	METHODOLOGY	64
	Hypertext and Hypermedia	64
	<i>HyperText Transfer Protocol (HTTP)</i>	72
	<i>Uniform Resource Locator (URL)</i>	75
	Browser	78
	<i>HyperText Markup Language (HTML)</i>	86
	Conclusion	89
IV	DEVELOPMENT OF AN ENVIRONMENT FOR VIRTUAL INFORMATION RETRIEVAL	91
	System Design	91
	Module Design	96
	On-Line Discussion Submodule	98
	Remote On-Line Databases Submodule	99
	Real-Time Discussion Submodule	100
	PERTANIKA Submodule	101
	Conclusion	104
V	RESULT AND DISCUSSION	106
	On-Line Discussion	106
	Remote On-Line Databases	110
	Real-Time Discussion	112
	PERTANIKA	114
VI	CONCLUSION AND RECOMMENDATION	117
	Conclusion	117
	Recommendations for Further Research	126



BIBLIOGRAPHY	129
INTERNET REFERENCE	136
APPENDIX	138
A The Internet Map	139
B Telnet Commands	140
C File Transfer Protocol Commands	141
D Gopher Commands	144
E Internet Relay Chat Commands	145
BIBLIOGRAPHICAL SKETCH	146



LIST OF TABLES

Table		Page
1	The Main Hierarchies of <i>USENET</i>	33
2	Dialects of the <i>Web</i>	47
3	Machine-Human Combination via Internet Tools	62
4	Examples of <i>URLs</i> and Types	77



LIST OF FIGURES

Figure		Page
1	The Concept of a Communication Network in General	3
2	Basic Connections of a Network	5
3	Part of Perl-Digest	35
4	General Diagram of the Information Industry	37
5	<i>Hytelnet</i> Menu	39
6	<i>FTP</i> Anonymous	40
7	<i>Gopher</i> Menu	42
8	<i>Veronica</i> via <i>Gopher</i> Menu	43
9	<i>Talk</i> on Half Screen of the Computer	49
10	Basic <i>Talk</i> Conversations	50
11	Connection to <i>IRC</i> Server at ' <i>irc.bu.edu</i> '	51
12	Layout of UPMNET	54
13	LAN at the Faculty Science and Environmental Studies	56
14	Layout of Learning on Demand	58
15	Basic Hypertext Document	65
16	Hypertext as a Subset of Hypermedia	68
17	Hypermedia in Relation to Other Computer-Based Technologies and Dicipines	69



39	Output of Remote On-Line Databases	111
40	Output of <i>Telnet</i> Session	111
41	Flow of Output for Real-Time Discussion	112
42	Output of Real-Time Discussion	113
43	Output of <i>IRC</i> Gateway	113
44	Flow of Output for PERTANIKA	114
45	Output of PERTANIKA	115
46	Output of PERTANIKA Journal of Science and Technology	115
47	Output of Text File of PERTANIKA	116
48	Layout of On-Line and Office Automation	119
49	Communication of ACM (CACM)	120
50	UPM News Bulletin	122
51	A Prototype of a Virtual Library for UPMNET	127
52	High-speed Network Demonstration - <i>ITX'95</i>	128



LIST OF ABBREVIATIONS

ARPA	- Advanced Research Project Agency
ARPANET	- Advanced Research Agency NETwork
ATM	- Asynchronous Transfer Mode
BBN	- Bolt Beranek and Newman
BBS	- Bulletin Boards System
BITNET	- Because Its Time NETwork
CERN	- Centre Europeenne puorla Recherché Nucleaire (European Centre for Nuclear Research)
CMU	- Carnegie-Mellion University
CSNET	- Computer Science NETwork
CWRU	- Case Western Reserve University
DDN	- Defence Data Network
DNS	- Domain Name Server
DoD	- Department of Defense
EMAIL	- Electronic Mail
ENIAC	- Electronic Numerical Integrator And Computer
FTP	- File Transfer Protocol
HTML	- HyperText Markup Language
HTTP	- HyperText Transfer Protocol



IMP	- Interface Message Processor
INWG	- Internet Networking Group
IP	- Internet Protocol
IPTO	- Information Processing Techniques Office
IRC	- Internet Relay Chat
IT	- Information Technology
JARING	- Joint Advanced Research Integrated NetworkinG
LAN	- Local Area Network
LOD	- Learning On Demand
MILNET	- MILitary NETwork
MIMOS	- Malaysian Institute of Microelectronics System
MIT	- Masschusetts Institute of Technology
NCP	- Network Control Program
NCSA	- National Computer System Association
NPL	- National Physical Laboratory
NSFNET	- National Science Foundation NETwork
OCLC	- On-line Computer Library Centre
OPAC	- On-line Public Access Catalogue
PC	- Personal Computer
RIPE	- Reseaux IP Europeans
SDC	- System Development Cooperation



SGML	- Standard Generalised Markup Language
SRI	- Stanford Research Institute
TCP	- Transmission Transfer Protocol
TCP/IP	- Transmission Transfer Protocol/ Internet Protocol
THEORYNET	- THEORY NETwork
UCLA	- University of California of Los Angeles
UCSB	- Culler-Fried Interactive Mathematics
UIUC	- University of Illinois Urbana Campaign
UPM	- Universiti Pertanian Malaysia
UPMNET	- UPM NETwork
URL	- Universal Resource Locator
USENET	- USE NETwork
USSR	- United State Soviet Russian
UUCP	- Unix-to-Unix CoPy
VERONICA	- Very Easy Rodent-Oriented Net-wide Index to Computerised Archives
WAIS	- Wide Area Information Service
WAN	- Wide Area Network
WWW/Web/W3	- World Wide Web



Abstract of thesis submitted to the Senate of Universiti Pertanian Malaysia in fulfilment of the requirements for the Degree of Master Science.

**DEVELOPMENT OF AN ENVIRONMENT FOR VIRTUAL
INFORMATION RETRIEVAL IN UPMNET**

By

HANAFIZAN BINTI HUSSAIN

January 1997

Chairman : Assoc. Prof. Dr. Abu Talib Othman
Faculty : Science and Environmental Studies

In this study, the development of an environment for virtual information retrieval in UPMNET that uses Internet tools such as *email*, *USENET*, *BITNET*, *Gopher*, *WAIS*, *FTP*, *Telnet*, *Archie*, *Hytelnet*, *IRC*, *WWW* etc is proposed. The purpose of this project is to gather all these tools to access or retrieve the information at any time. The methodology for this project is based on the concept of hypertext and hypermedia including *HTTP*, browser, *HTML* and *URL*.

In this project, Internet tools have been categorised into three groups; on-line discussion, remote on-line database and real-time discussion. On-line discussion has been used for sharing ideas continuously, for instance using email while remote on-line database can be used for searching information in other



various databases. Real-time discussion allows for academicians to have real life conversations as though they are in a 'face to face' situation.

Another feature of this project is the development of a prototype of an electronic version for PERTANIKA Journal of Science and Technology. This prototype will consist of information on author, title, year of publishing, abstract of article and email of the author. This will enable academicians to share the ideas and disseminate their knowledge. It also represents one of the steps towards the implementation of an environment for virtual information retrieval in UPMNET.

The study concludes with recommendations for further research on enhancing UPMNET applications particularly in the realms of office automation, research and dissemination of scholarly information.



Abstrak thesis yang dikemukakan kepada Senat Universiti Pertanian Malaysia sebagai keperluan untuk mendapatkan ijazah Master Sains.

**PEMBANGUNAN PERSEKITARAN UNTUK PENCAPAIAN
MAKLUMAT MAYA DI UPMNET**

Oleh

HANAFIZAN BINTI HUSSAIN

Januari 1997

Pengerusi : Profesor Madya Dr. Abu Talib Othman
Fakulti : Sains dan Pengajian Alam Sekitar

Dalam kajian ini, pembangunan persekitaran untuk pencapaian maklumat maya di UPMNET yang menggunakan perisian yang boleh didapati dari Internet seperti *email, USENET, BITNET, Gopher, WAIS, FTP, Telnet, Archie, Hytelnet, IRC, WWW* dan sebagainya. Tujuan kajian ini adalah untuk mengabungkan perisian-perisian ini bagi membolehkan pencapaian maklumat boleh diperolehi pada bila-bila masa sahaja. Kaedah yang digunakan dalam kajian ini adalah berdasarkan konsep hypertext dan hypermedia seterusnya merangkumi *HTTP, browser, HTML* dan *URL*.

Dalam kajian ini, perisian-perisian Internet telah dikategorikan kepada tiga bahagian iaitu '*on-line discussion*', '*remote on-line database*' dan '*real-time dicussion*'. '*On-line discussion*' digunakan apabila mereka ingin berkongsi idea



secara berterusan seperti menggunakan *email* manakala '*remote on-line database*' pula boleh digunakan bagi mendapatkan bahan-bahan yang terdapat di pelbagai pangkalan data. '*Real-time discussion*' pula membolehkan mereka mengadakan perbincangan seolah-olah mereka berada di dalam situasi 'berhadapan antara satu sama lain'.

Satu prototaip untuk PERTANIKA Jurnal Sains dan Teknologi versi elektronik telah dibina. Prototaip ini akan mengandungi nama penulis, tajuk, tahun diterbitkan, abstrak artikel tersebut dan alamat email penulis. Ini membolehkan para akademik berkongsi dan menyebarkan idea dan pengetahuan mereka. Ianya merupakan permulaan implimentasi untuk pembangunan persekitaran untuk pencapaian maklumat maya di UPMNET.

Kajian ini juga turut memberikan beberapa idea untuk penyelidikan akan datang bagi meluaskan penggunaan applikasi UPMNET terutamanya di dalam bidang automasi pejabat, penyelidikan dan penyebaran maklumat.



CHAPTER I

INTRODUCTION

Computer Network

Computers are found everywhere in shops, banks, schools and even at home. The computers constantly challenge us with ‘new ways of doing science’ and it becomes an important aspect of the scientific community (Piore, 1980). We are reaching an era when ‘small science’ can survive only with the use of the new small computers. Computers usage in the various field of education, research and development, business and administration, to name as few, has increased tremendously over the decade.

The decreasing cost of computers and the advances in data communications and computer networking technology have extended computer applications in many areas of business, government and the academia. Spurred on by the proliferation of microcomputers, computer networks are gaining popularity. Thus, the merging of the computer and communications has had a profound influence on the way computer systems are organized.



Once a computer is put into the communication loop, the potential for structuring facilities and argumenting the communication and information exchange among a number of user groups are unvirtually limited. The early 1970s until early 1980s saw a merger of computer science and data communication that profoundly changed the technology, products and companies of the now combined computer communications industry. The computer revolution has resulted in several remarkable developments:

1. There is no fundamental difference between data processing (computers) and data communications (transmission and switching equipment).
2. There are no fundamental differences among data, voice and video communications.
3. The differences between a single processor computer, multi-processor computer, local network, metropolitan network and long-haul network have blurred.

An effect of these trends has been the growing overlap of the computer and communication industries, from component fabrication to system integration. Another result is the development of integrated systems that transmit and process all types of data and information. Both the technology and the organisations technical standards are moving towards a single public system that integrates all

communication and makes virtually all data and information sources around the world easily and uniformly accessible (Stallings, 1994).

In its simplest form, data communication takes place between two devices that are directly connected by some forms of a point-to-point transmission medium. Figure 1 illustrates the concept of a communication network in general (Sphonn, 1993).

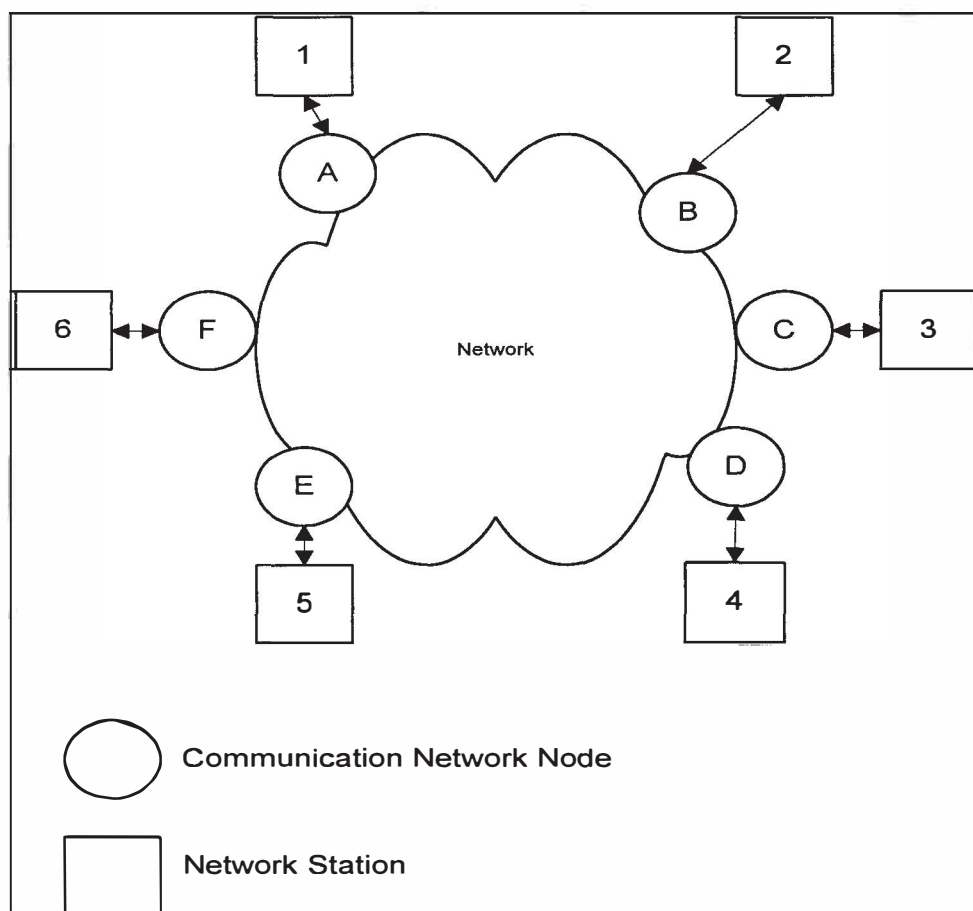


Figure 1: The Concept of a Communication Network in General

A communication network is a collection of interconnected functional units that provides data communication service among stations attached to the network (Stallings, 1994). A computer network is described as an interconnected collection of autonomous computers; two computers are said to be interconnected if they are able to exchange information (Tanenbaum, 1989).

Computer networks are derived from a combination of computers and communications; two technologies with different histories and traditions. Before the advent of computers, machines of very great organisational complexity had been produced for automatic telephone switching. Then in the mid-twentieth century, two rapid changes occurred. Organisational complexity leaped ahead with the concepts of the stored program and of programming languages while solid-state electronics set in motion a rapid improvement in the speed and cost of digital processing that is still on-going. The concepts and techniques that computers stimulated were found to apply just as effectively to communications.

Computer networks in the late-twentieth century have served to move information and reduce the effect of distance. Networks began as extensions of computers and to a large extent, computer networks will be used to serve human communications and be part of the essential infrastructure of the post-industrial world.

A network is a distribution system composed of interlinked spatially dispersed channels, subsystems and/or elements (Samuelson et al., 1977). A computer network is only the first step. Figure 2 illustrates the basic connections of a network that operates a number of links interconnected by nodes (Black, 1993).

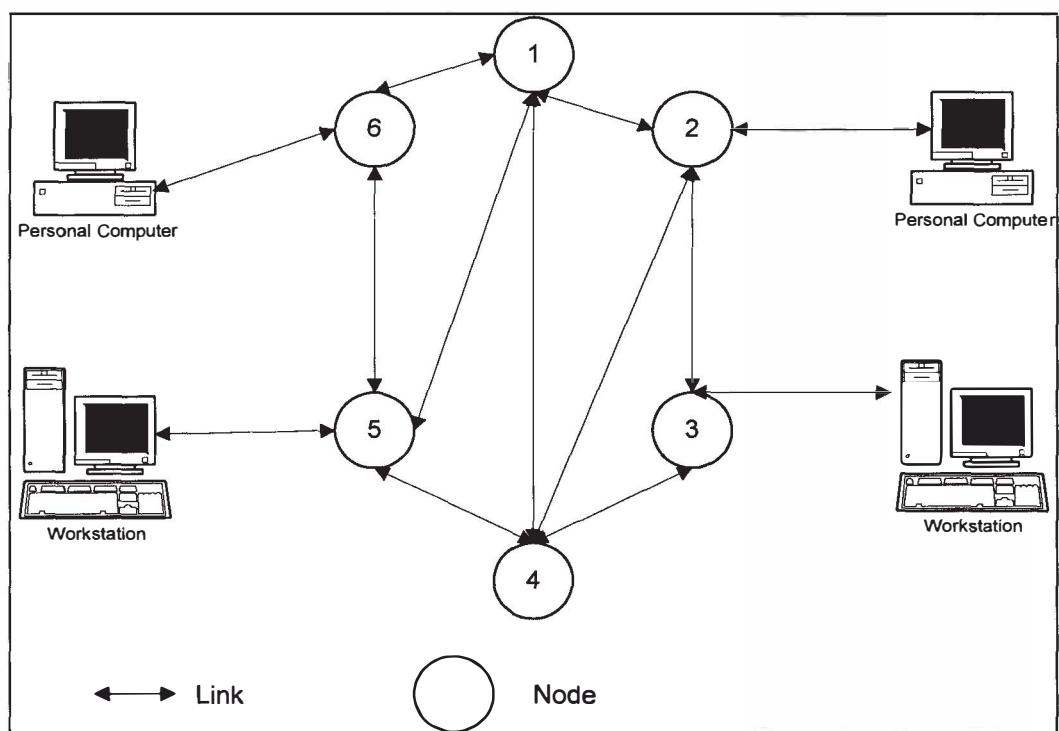


Figure 2 : Basic Connections of a Network

Computer and communication hardware and software technologies provide an effective means for building an information system (Rzevski, 1985). An information system is a combination of human and computer-based capital

resources which results in the collection, storage, retrieval, communication and use of data for the purpose of efficient management of operations (planning, decision-making, reporting and control) in organisations (Samuelson et al., 1977). Connecting machines in a network gives us even more computing and information resources than we can get from simply by having a computer at our desk.

The Internet is 'network of networks' and is used to describe the massive world wide network of computers (Hughes, 1994). The Internet is the largest computer network in the world. Nobody is really sure just how big the Internet is but it is huge. The network connects in excess of two million personal computers and hundreds of servers throughout the world (Meeks, 1995).

The Internet represents many things to many people but in actuality, it is an immense global computer network. It started out in the 1960s as a research and development project in the US. Today, however, it has grown to such an extent, that it has acquired the status of a vast library as well as a world-wide communications medium.

The Internet currently spans about 170 countries and has approximately 50 million users which is rapidly growing. In Malaysia, the Internet subscriber has been growing at around 20 per cent per month (IT Malaysia, 1995). These

subscribers are a part of the new 'communications age' society. The Internet is one of the most important elements in the world of information technology today. The Internet map can be seen in Appendix A.

Initially the Internet was a network designed to facilitate the work of researchers and academics. Currently, it is being used to enhance teaching and learning in a variety of subjects at various levels of education (Zoraini, 1995). Today researchers and academicians especially in the field of computer science have access to Internet with some 2 million Internet nodes and an estimated 30 million people can connect to Internet, many of them directly from their desk.

In the last couple of years, the Internet has seen a phenomenal user growth rate of about 10 per cent per month . It can be seen that the emergence of the *World Wide Web (WWW or W3 or Web)*, (sometimes referred to as multimedia on Internet, with interactive *WWW* pages, voice and graphics) has contributed largely to this growth. The *WWW* refers to a body of information while the Internet refers to the physical side of the global network, a giant mass of cables and computers. The *WWW* merges the techniques of information retrieval and hypertext to make an easy and powerful global information system (Polly, 1993).

The era of an 'information society' is definitely possible with the creation of information highway. The information highway has received much attention

