



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

**Charles University in Prague
Institute of Information Studies and Librarianship**

Modul No.1
Information and the Internet
Milena Medková

Překlad: Anna Neumanová

For the project:

Title: Studies of Information and Knowledge Management in the European Context

Reg. No: CZ.1.07/2.2.00/07.0284

OP: Education for Competitiveness

Support area. 2.2 University Education

Realization: VŠB-Technical University of Ostrava, Faculty of Economics, separate unit:
Business Academy and HPS Valašské Meziříčí

CONTENT

Content	2
Introduction	3
1. Brief History of the Internet	3
1.1 From the ARPANET to the Internet	5
1.2 Internet Operation	6
1.3 The Internet in the Czech Republic	7
2. Internet Protocols	9
2.1 Layered model TCP/IP	9
2.2 Documents about the Internet	11
2.3 Addresses of Computers on the Internet	12
2.3.1 IP addresses	12
2.3.2 Domain addresses	13
2.3.3 Rules for creating domain names	13
3. Internet Services	15
3.1 Remote Login	15
3.2 Transferring files using FTP	16
3.3 Electronic mail	17
3.4 Gopher	19
3.5 World Wide Web - WWW	20
3.6 Electronic conferences and discussion groups	22
4. Introduction to Internet Search Strategy	25
4.1 How to search – tips for better searching	26
4.1.1 Main steps of search strategy	26
4.2 Search queries	27
4.2.1 Simple queries	27
4.2.2 Advanced queries	29
5. Classification of Searching Services	32
5.1 Resource List	32
5.2 Subject catalogues (Web portals)	32
5.3 Search Engines	34
5.4 Meta-search Engines	36
5.5 Other types of search tools	36
5.6 Virtual Libraries	37
5.7 Invisible web	37
5.8 Social Networks	38
6. Google.com	40
6.1 Chronology of crucial events	40
6.2 Search Strategy Google.com	41

„Learning about the Internet is not something one can complete in an afternoon – learning never stops because the Internet keeps changing“

Douglas E. Comer

INTRODUCTION

This material Internet and Information is an introduction to information processing. Its aim is to support information literacy of university students in their first year of study. The Internet becomes an integral part of their preparation for studies, it is necessary to use the tool of the Internet effectively and above all critically. In view of the fact that everything around the Internet is constantly changing and developing, the text does not go into details but follows the broader not so changeable frame. The teaching material was created within the frame of the project Study of Information Science and Knowledge Management in European Context, which was financed from the Operational Programme Education for Competitive Advantage.

1. BRIEF HISTORY OF THE INTERNET

The Internet has changed the world of the information and communication technology like nothing before. The inventions of telegraph, telephone, radio and computer broke new ground for integration possibilities of the Internet. The history of the Internet started before 1962, in the period when computers, despite being primitive with low memory capacity, cost hundreds, thousands of dollars. At that time the company RAND Corporation (<http://www.rand.org>) was solving the considerable task – to think up a plan how individual administrative authorities could successfully communicate also after a nuclear war. It was sure that the solution should be based on a suitable communication network. The RAND Corporation came up with the solution which was published in 1964. It is based on two following principles:

- the net will have no central component
- right from the start the net will be designed so as it could work even if some parts of it are in ruins

The realization of the principles was based in the fact that the network was designed so as all its nodes had equal status and beforehand it was taken into account that transfers among particular nodes were not reliable, (that the switching nodes and network links were unreliable). The very second requirement stood at the birth of the idea which was based on the fact that transmitted data would be divided into appropriately big parts called packets, and they would be transmitted as separate units – similarly like letters sent by post. Each data packet will have a full receiver's address, the path of its transmitting will be set separately, independently on other packets. Different packets will have different paths, so when a path to the destination will be damaged other packets will be transmitted via other paths.

So the technology called Packet Switching came into being on top of that in the new version, now called Datagram Service. Usually it is rather less effective than the data transfer which simulates telephone network functioning called Circuit Switching but on the other hand it provides maximum robustness. [3].

It is necessary to point out that not only the RAND Corporation worked on the project of packet switching but contemporarily also MIT and the National Physical Laboratory in Great Britain. Paul Baran, Donald Davies and Leonard Kleinrock etc. worked on the research. In fact the packet switching is no modern invention, it is only a suitable application of the ancient idea of dynamic usage of means of transport having been used by post and telegraph systems for ages. Consignments or messages are usually transported or handed over at one go when the carriage is loaded fully or the transmission route is made full use so as the transport pays off.

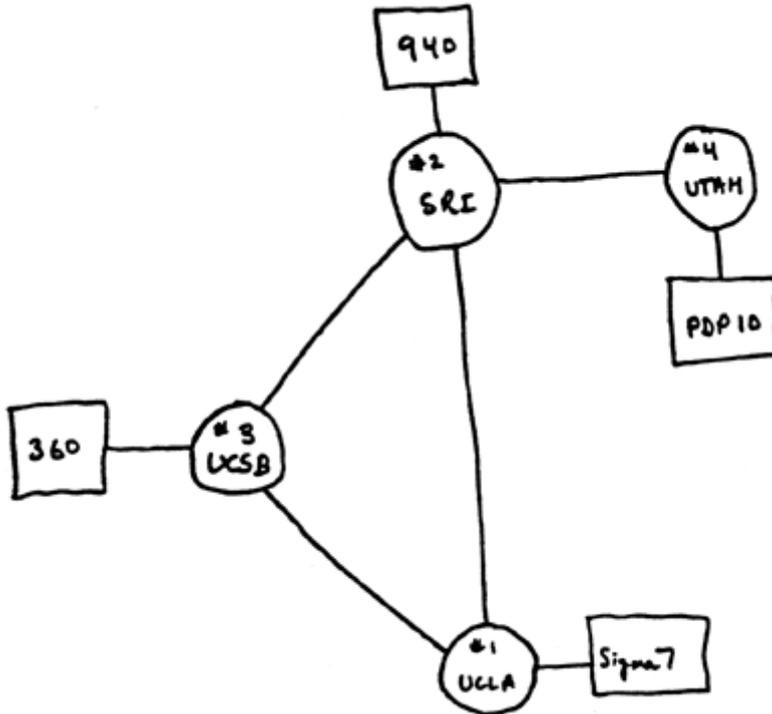
The idea of decentralized maximally robust and “explosion proof” network running on the packet switching principle was implemented in practice for the first time in Great Britain. The experimental network came into being in the National Physical Laboratory in 1968

The USA plucked up the courage for it in 1969. The initiative and financial support came from the Department of Defence, specifically from the science foundation of the Department of Defence called Advanced Research Projects Agency (ARPA).¹ After this agency the experimental network was named as ARPANET. The first nodes of ARPANET were placed at four American universities, - UCLA (University of California Los Angeles, Vint Cerf, Steve Crocker, Jon Postel), UCSB (University of California Santa Barbara, Glen Gulle, Burton Fried), at Stanford (Stanford Research Institute, SRI, Doug Engelbart) and at university in Utah. The actual node was realized by the universal computer (it was the Honeywell DDP516 computer), which was programmed to be able to work as so called Interface Message Processor (IMP). For the intercommunication the nodes used IMP circuits transmission speed 50kbps, and transmission protocol NCP (Network Control Protocol). Before the end of 1969 the network was put into operation and its remote users could use facilities and computing capacity of supercomputers, which the net interconnected

The ARPANET grew very quickly, in 1971 it had 15 nodes, although the number rose continually: in 1972 ARPANET had 37 nodes and in 1973 it was joined by nodes from abroad, from the UK and Norway.

In October 1972 Kahn successfully presented the ARPANET at the International Conference on Computer Communication, the ARPANET had functions such as e-mail, long-distance connection and file transfer.

¹ ARPA was later renamed DARPA (Defense Advanced Research Project Agency) <http://www.darpa.mil>



Picture No. 1: Original diagram of the Arpanet, 4-node ARPANET diagram

Source: http://www.computerhistory.org/exhibits/internet_history/ [2]

A group of teachers and postgraduate students from the University of California Los Angeles, established the Network Measurement Center where the behaviour of the ARPANET was tested and measured. Members of the group were people such as Vinton Cerf, Steven Crocker or Jon Postel. The results of the research were published in the form of working documents which were called “Request For Comment”. The very first material RFC (Request For Comment) was published 7.April 1969. It was written by Steven Crocker and dealt with the problems of software for network nodal computers. <http://tools.ietf.org/html/rfc1>

The key person in the further development of the ARPANET was Vinton G. Cerf who after finishing his postgraduate studies assumed office of associate professor at university in Stanford. Shortly after that he began organizing networking seminars, in which the foundations of new generation of communication protocols for the ARPANET – TCP/IP protocols family.

1.1 From the ARPANET to the Internet

The fate of the ARPANET network took place completely under the control of the U.S. Department of Defense which funded all activities of the network. In 1983 the parts of the ARPANET which had something in common with military affairs were separated from the original ARPANET. A separate network MILNET was established. The ARPANET gained more civil job content but it was still funded from the budget of the U.S. Department of Defense.

ARPANET, however, was far from a single computer network in the world or at least in the USA. Computer networks were built by other departments. Due to the qualities and public availability of TCP / IP these networks were increasingly built on the basis of these protocols. For network operators it was very useful when they could be linked to ARPANET. It was thus increasingly becoming more rudimentary network, which "snowballed" more and more networks so much that a conglomerate of interconnected networks was created and called the Internet.

Moreover it was associated with the advent of local networks (LANs), which occurs sometime after 1983 when the best known and most widely used technology for the transmission LAN - Ethernet fully matures. Also, for a countless number of local networks it was very convenient to use TCP / IP to connect to the ever-expanding system of interconnected networks based on these protocols, i.e.the Internet.

Probably the most important network that "snowballed" on the ARPANET and thus became part of the Internet was the NSFNET network. Thanks to the generous funds the agency NSF could finance literally mass connecting of academic institutions to NSFNET and therefore to the Internet. NSFNET with its transmission capabilities, and also on the basis of a conceptual plan, gradually borrowed the original ARPANET role - a role of the network backbone, through which the most of the traffic passes.

1.2 Internet Operation

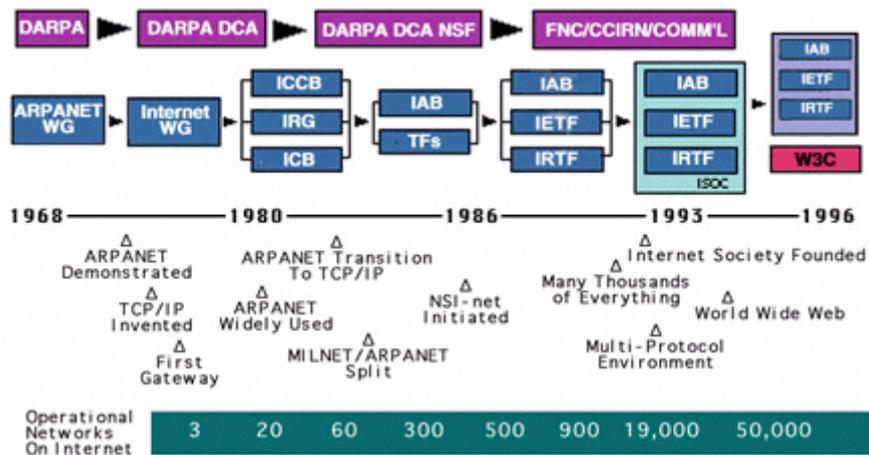
One of natural trends in this area was the fact that institutions that took care of building significant networks soon entrust the care of the routine operation with specialized institutions. The agency ARPA did it so in July 1975, when he handed over the ARPANET as a fully functional network management organization in the DCA (Defence Communications Agency, later renamed the DISA Defence Information Systems Agency). Management of the Internet itself (more precisely the operation of NCC the Network Control Centre), the Ministry of Defence instructed the firm of BBN (Bolt, Beranek and Newman), who was already at the very birth of ARPANET and "built" his first nodes. The company operating the Control Centre NCC continued until 1993 (later, on the basis of grants coming from the agency NSF).

The NSF acted similarly, when, after less than a year after launching its NSFNET backbone network, gave the management company Merit Network, Inc.. (Founded by a consortium of Michigan universities), and together with IBM and MCI to it by the required implementation of an upgrade NSFNET backbone - especially its acceleration.

Besides its own network NSFNET and its operation the agency NSF funded the whole highly-developed system of collection and distribution of various information about the Internet. This system was based on the information centre NIC (Network Information Center), which for many years worked in Stanford for the local SRI (Stanford Research Institute). This resort secured registration of all new networks connected to the Internet and it allocated globally

unique network addresses (called IP addresses). Moreover, the centre NIC in Stanford was also one of the main depositories of Internet documents (eg RFCs).

From April 1993, however, most information and other services around the Internet were entrusted to a consortium of three companies that provide them on a commercial basis: Network Solutions, Inc., AT & T and General Atomics. The consortium, known as the InterNIC, is partly funded by the NSF Agency. [1]



Picture No. 2: Time axis

Source: <http://www.isoc.org/internet/history/brief.shtml> [1]

1.3 The Internet in the Czech Republic

The history of the Internet in the Czech Republic started in November 1989.

At that time the political barriers, which until then had made our connection to the outstanding world networks impossible, were broken.

With the abolition of the political obstacles, suddenly obstacles of technical nature were of great importance. In our country there was not a suitable communication infrastructure for larger development of computer networks. The first networks, which we received after the Velvet Revolution, placed minimum demands on necessary communication infrastructure. In March 1990, FIDO Network gets to our country and the Bulletin Board System was connected to it.

The Eunet network came to Czechoslovakia in May 1990 and particularly Unix computers were connected. In October 1990 we received the European branch of Bitnet network, the European Academic and Research Network which requires a permanent connection over fixed circuits. The first node of the network in this country, called the National Network EARN node, became the IBM 4381 mainframe computer at the Regional Computer Center of the Technical University in Prague. The EARN network provided only services of the batch nature (especially electronic mail and file transfer), so it made do even with relatively slow circuits. The node CSEARN began with a line on the transmission speed of 9600 bps, and it was connected to the national Austrian EARN node at Linz.

The first attempts to connection to the Internet appeared in November 1991. This date is also

in the various official and unofficial statistics stated as the date of Czechoslovak connection to the Internet. Initially, there was only a commute connection from Prague (specifically VC CTU) to the node in Linz, Austria, and later a telephone line to Linz was artificially “halved” so that the first half transmitted operations in the EARN network, and the other the Internet operation. [4]

In November 1991, Czech universities connected to the European Internet via the CESNET network and the history of the Internet in the Czech Republic started.

Already in 1991a proposal to build a nationwide backbone network was submitted. It was to connect all domestic academic centres and the Internet should then be distributed over the metropolitan network. On the ground of the proposal two projects on a national backbone network, Slovak and Czech were created. The Czech project was called FESNET – the Federal Educational and Scientific Network.

On 13th February 1992 Czechoslovakia is connected to the Internet. The festive act took place at the Technical University in Prague in the presence of representatives of the NSF agency.

In June 1992 the Czech Ministry of Education approved the project and released for it 20 million crowns. During 1992, the letter F in FESNET changed to C and in this way CESNET was created (United Educational and Scientific Network). In the Slovak Republic the network SANET (Slovak Academic Network) was simultaneously built.

Literatura:

1. *A Brief History of the Internet* [online]. c2010 Internet Society. [cit. dne 2010-03-17]. Dostupné z: <http://www.isoc.org/internet/history/brief.shtml>.
2. *Internet history: A history of the Internet 1962 - 1992* [online]. c2006 Computer History Museum. [cit. dne 2010-03-17]. Dostupné z: http://www.computerhistory.org/internet_history/.
3. PETERKA, J. Internet. *Computerworld*, 1995, roč. 6, č. 4, s. 17 – 19. ISSN 1210-9924.
4. PETERKA, J. *Internet u nás* [online]. [cit. dne 2010-03-17]. Dostupné z: <http://www.earchiv.cz/a95/a504c504.php3>
5. *TCP/IP Overview and History: Modern TCP/IP Development and the Creation of TCP/IP Architecture* [online] c2001 - 2005 Charles M. Kozierok [cit. dne 2010-01-30]. Dostupné z: http://www.tcpipguide.com/free/t_TCPIPOverviewandHistory-2.htm.

2. INTERNET PROTOCOLS

In connection with the Internet and computing network we often meet the term protocol. A protocol is a specific set of rules (in other words standards and regulations) which a particular device has to meet to be able to communicate with another corresponding device or programme. The process of creating network protocols is long-term and protocols are continually being developed. The mixture of the protocols TCP/IP and ITU and ISO is used on the Internet. The protocols ITU and ISO occur mainly in the two low layers, physical and link.

TCP/IP(Transmission Control Protocol/Internet Protocol) marks the whole network architecture, not only two protocols included in its name, the two protocols form its core. According to one of the designers, Doug Comer, its correct and entire signification should be the TCP/IP Internet Protocol Suite. [PUŽMANOVÁ, 2009, s.83]

2.1 Layered model TCP/IP

Unlike the seven-layered reference model OSI², the protocol architecture of TCP/IP is formed by four layers:

- Network interface
- Internet layer
- Transport layer
- Application layer

FTP, TFTP, HTTP							
HTTPS, TELNET,							
POP3, SMTP, IMAP,							
RPC, XDR,		RIP	III. layer				
TCP		UDP	II layer				
IP			ARP	RARP	ICMP , IGMP		I. layer
Ethernet		Token Ring		Other media		hardware	

Source: *Počítačové sítě...*[online]. Available at: <http://site.the.cz/index.php?id=3>

²Reference model OSI (Open Systems Interconnection) was created by Mezinárodní normalizační organizace a přijala ho jako mezinárodní normu IS 7498 v roce 1984. OSI se začala zabývat i Mezinárodní telekomunikační unie (ITU), která přijala a vydala kompletní text zmíněné mezinárodní normy jako své doporučení X.200. [PUŽMANOVÁ, 2009, s.42]

Hardware (network interface layer)

The lowest layer of the architecture enables an access to the physical transmission medium. It is responsible for the access to the network and so it is specific for each network according to its implementation.

ARP - Address Resolution Protocol

The Address Resolution Protocol (ARP) transmits 32 bit IP address into 48-bit MAC address.

RARP - Reverse Address Resolution Protocol

On the contrary RARP transmits MAC address into IP address. This protocol is used by diskless workstations which do not know their address.

ICMP - Internet Control Message Protocol

It is used for indicating of error messages and other substandard situations (only for indicating not for correcting)

IGMP - Internet Group Management Protocol

supports so-called multicasting.

The lowest layer of the architecture enables an access to the physical transmission medium

II. Layer (transport layer)

It provides the mechanism for end-to-end communication between two workstations.

TCP - Transmission Control Protocol

It is an acknowledged protocol. TCP creates the so-called virtual connection. This connection lasts until the application is disconnected.

UDP - User Datagram Protocol

It is an unacknowledged protocol. Unlike IP it has more ports. An enquiry can be sent to a programme. It is not used very often, more likely for corporate communications. For example some routers every 30 seconds report who is connected.

III. layer(application layer)

The application layer includes protocols which are used directly for communication on the network.

FTP/TFTP - File Transfer Protocol/Trivial FTP

FTP (File Transfer Protocol) is used for transferring files among computers connected in a network.

HTTP/ HTTPS - Hypertext Transfer Protocol is used for having an access to the World Wide Web. HTTPS is secured (encrypted) transfer of www pages.

TELNET - Telecommunication Network

It creates terminal operating. We can work on a remote computer as if we were sitting at a terminal connected to it. Through Telnet, a user can access someone else's computer remotely.

Using this protocol poses a risk because the communication runs en clair. In consideration of TELNET, Secure Shell (SSH), sometimes known as Secure Socket Shell, is for securely getting access to a remote computer.

POP3 - Post Office Protocol

It is used for receiving e-mails by an e-mail client.

SMTP - Simple Mail Transfer Protocol

It is used for sending e-mails by an email client.

2.2 Documents about the Internet

All information about the Internet and TCP/IP (conventions and network protocols) has been available from the very beginning on the web since 1969 and is called Request for Comments (RFC).

RFC types [PUŽMANOVÁ, 2009, s.31 - 34]

:

- Standards track – Proposed (first level), Draft (second level) Internet standard (STD, full de facto standard)
- Best current practice (BCP)
- Informational (some marked FYI)
- Experimental
- Historic

Note

More information about RFC is in [RFC 2026](#)

Available free of charge at the address:

<http://www.ietf.org/rfc.html>

Editor RFC:

<http://www.rfc-editor.org/>

2.3 Addresses of Computers on the Internet

There are two types of addresses

1. IP addresses
2. Domain addresses

2.3.1 IP addresses

Computers and other devices interconnected in one common network have its own address. For a particular computer it is unique, it is not repeated anywhere else on the Internet. Each address determines unequivocally a computer which was allotted to. Packets passing on the Internet carry.

IPv4

Internet Protocol version 4 uses the IP address of a length of four bytes. IP address addresses a unique network interface - unicast. If the system has multiple network interface cards (more than one network interface) and all is running the IP protocol, then each interface has its IP address.

It is perhaps the opposite option, on the one network card (physically one network interface) supports multiple IP addresses. The first address is usually called primary, secondary, then other addresses or aliases. Using secondary IP addresses as it is common for web servers, when one computer running the Web sites of several companies and each has a look as a standalone Web server.

IPv6

IPv6 was created in the middle of 1990s.

The most obvious improvement in IPv6 over IPv4 is that IP addresses are lengthened from 32 bits to 128 bits (RFC 4291). While Ipv4 contains roughly 4 addresses, IPv6 has enough space for 3.4×10^{38} unique addresses. Ipv6 addresses consist of two logical parts: 64-bit subnetwork prefix and 64-bit part for a host, either created automatically on the basis of MAC address interface or allocated consequently. Since the globally unique MAC address enables to trace up a user facility – so a user – IPv6 addresses change over time. IPv6 addresses are usually registered as eight groups of four hex numbers. For example 2001:0db8:85a3:08d3:1319:8a2e:0370:7334 is a valid IPv6 address.

If one or more from the quadripartite groups is 0000, zeros can be omitted and replaced by two colons (::). For example 2001:0db8:0000:0000:0000:0000:1428:57ab can be replaced by 2001:0db8::1428:57ab.²

Note

- More information: RFC 4291: <http://tools.ietf.org/html/rfc4291>

2.3.2 Domain addresses

The numerical IP addresses are difficult to remember so instead of the IP address of the network interface the name of the network interface is used. For each IP address there is the name of interface (computer), more precisely, the domain name. One IP address can be assigned none, one or more domain names. The system Domain Name Service (DNS) converts numeric addresses to the domain addresses. A domain name consists of chains each separated by periods.

2.3.3 Rules for creating domain names

Domain names must conform to standards RFC1 1034, 1035, 1122, 1123 and any standards replacing them. In particular, they must follow these conditions:

- A domain name can contain only characters [a-z ,0-9, -];
- The length of the domain name is 63 characters;
- A domain name can not begin or end with the character '-';
- A domain name can not contain two characters "-" in a row.

More: http://www.nic.cz/files/nic/doc/Pravidla_registrace_CZ_ccReg_20080930_fin.pdf³

The most important and interesting RFC documents related to the DNS:

- [RFC 1034](#) – Domain Names - Concepts and Facilities
- [RFC 1035](#) – Domain Names - Implementation and Specification
- [RFC 1537](#) – Common DNS Data File Configuration Errors
- [RFC 1886](#) – DNS Extensions to support IP version 6
- [RFC 1912](#) – Common DNS Operational and Configuration Errors
- [RFC 1996](#) – A Mechanism for Prompt Notification of Zone Changes (DNS NOTIFY)
- [RFC 2181](#) – Clarifications to the DNS Specification
- [RFC 2308](#) – Negative Caching of DNS Queries
- [RFC 2870](#) – Root Name Server Operational Requirements
- [RFC 3330](#) – Special-Use IPv4 Addresses
- [RFC 3467](#) – Role of the Domain Name System

Source: ŠŤASTNÝ, P.

³ The association of legal entities CZ.NIC (www.nic.cz) is responsible for the registry of domain names

Literatura:

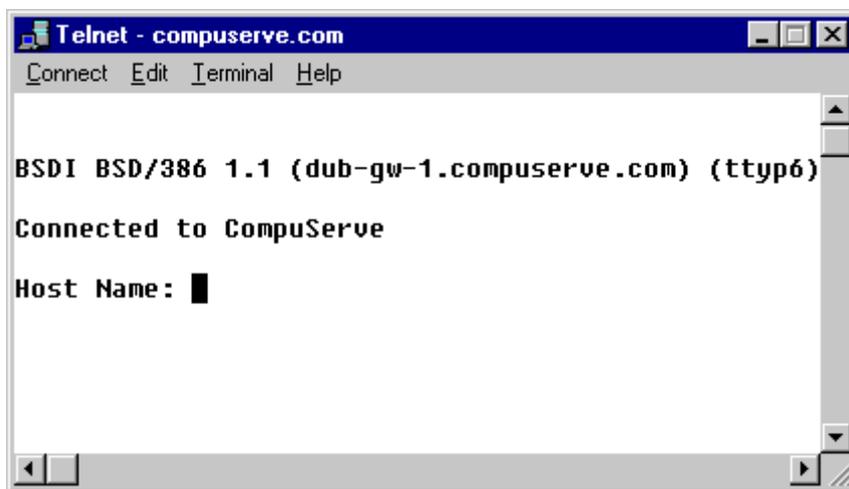
1. PUŽMANOVÁ, Rota. *TCP/IP v kostce*. 2. uprav. a rozšiř. vyd. České Budějovice: Koop, 2009. 619 s. ISBN 8978-80-7232-388-3.
2. DOSTÁLEK, Libor, KABELOVÁ, Alena. *Velký průvodce protokoly TCP/IP a systémem DNS*. 5. aktualit. vyd. Brno: Computer Press, 2008. 488 s. ISBN 978-80-251-2236-5.
3. MUSIL, Marek. *Historie internetu* [online]. Poslední modifikace 23. července 2003. [cit. Dne 1. března 2010]. Dostupné z: <http://ihistory.webzdarma.cz/chap/sites/RFC.php>
4. ŠŤASTNÝ, Petr. *Dokumenty RFC a jiné – DNS* [online]. Poslední změna 29. dubna 2007. [cit. Dne 1. března 2010]. Dostupné z: <http://www.pweb.cz/dns/rfc.html>
5. *Vlastnosti protokolů* [online] Poslední editace 6. srpna 2008. [cit. Dne 1. března 2010]. Dostupné z: https://www.ipv6.cz/Vlastnosti_protokolu

3. INTERNET SERVICES

Nowadays, the Internet offers a wide range of services. The oldest service was electronic mail, transfer of files and remote login. Today www is used at the maximum.

3.1 Remote Login

Remote login is done via Telnet, by which utilities (programs) are named that implement it. The user gets to your computer Telnet and log on to the remote computer (enter name of your user account on your computer and password). Then you can work on that computer as if sitting on an actual terminal physical proximity of the remote computer. The problem of remote login is small but its user friendliness - it works only in text mode.



Picture No. 3 : Login to telnet

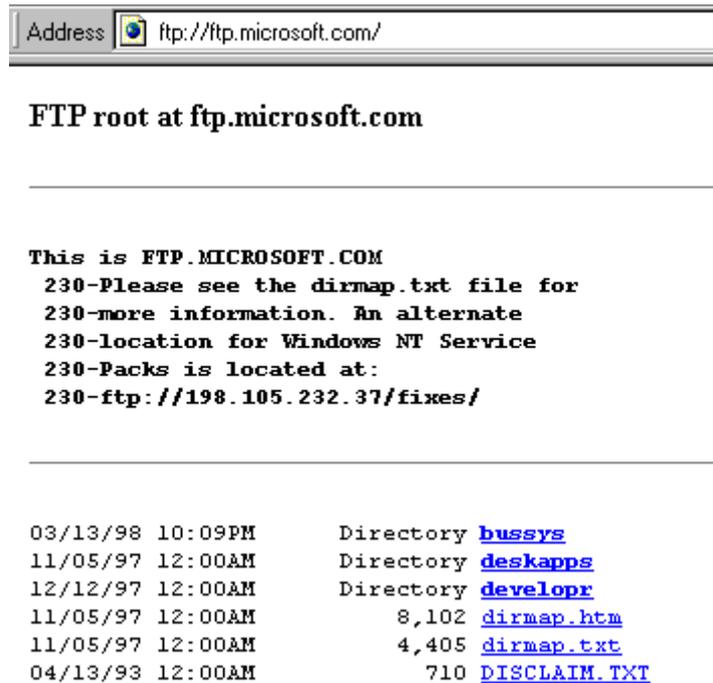
Source: <http://www.techtrainteam.com/services/olt/internet/ftp-0.html>

3.2 Transferring files using FTP

FTP or File Transfer Protocol is a service that allows you to work on remote files, mainly the upload (copy to a remote server) or download (free download), then delete, change to the rights, working with the FTP directory, etc. can be used in two cases:

- Download from anonymous FTP servers
- Administration of the website.

The first draft specification of this protocol is already in the year 1971 (RFC 114), the current version of the standard described in RFC 959



Picture No. 4: Root of FTP <ftp://ftp.microsoft.com>
Source: <http://www.techtrainteam.com/services/olt/internet/ftp-0.html>

Example: Run command line in Windows; for example. Start > Run: command .

In comand line:

```
c:\>ftp ftp.linux.cz
User: anonymous
Password: vas@email.cz
ftp>help
ftp>ls
ftp>quit
```

3.3 Electronic mail

The newspaper MF Dnes, 18.2.1994, Washington: "American president Bill Clinton and Swedish prime minister Carl Bildt became the first top officials (politicians), who exchanged diplomatic mail through the Internet".

Electronic mail belongs to the oldest and most popular services of all computer networks including the Internet. Its popularity and importance is so big that it even changes habits and behaviour of users.

The advantages of electronic mail

- speed
- its electronic form itself – text messages are electronic documents, they can be archived, processed, printed, searched and so on., a huge amount of operations can be automatized and so they can increase comfort and efficiency of users – directories, templates, e-signature and so on.
- the possibility to send messages to many receivers at the same time – each message can have three kinds of addresses all at once:
 - **main recipient** (To: - to him/her the message is sent)
 - **secondary recipient of a message:** (Cc = carbon copy :that who is to receive a copy of a message directed to another. The secondary recipient can see that the main recipient is somebody else and he gets only the copy for his attention. And similarly the main recipient finds out who received the copy of the message.
 - **recipient of a blind copy** – he/she can see who are the main and secondary recipients, he can see the content of the message but it does not apply to vice versa. Neither the main recipient nor the secondary recipient know who gets the blind copy.

For creating and delivering an electronic letter two programs are needed:

1. Mail Transfer Agent (MTA) - this program runs on the server and its task is to ensure that an email message is delivered to the mailbox of the appropriate recipient. A Mail Transfer Agent (MTA) transfers email messages between hosts using SMTP (Simple mail transfer protocol)

2. Mail User Agent (MUA) - is a program, which a user works with. This program creates a user interface for managing user's email. For access to the user's mailbox the Mail User Agent has to use many special protocols.

a) Post Office Protocol 3 (POP3) – it enables to collect the email that is to retrieve email from a remote server. The protocol has some disadvantages – messages cannot be filtered, all messages are transferred. It is not possible to select messages according to a subject. It is not possible to transfer only parts of messages.

b) Internet Message Access Protocol 4 (IMAP4) – tries to eliminate the disadvantages of the protocol POP3 – for example it enables to transfer only the letterhead etc.

Email and Netiquett

Netiquett (from Network Etiquett)

7 rules, how to write email according to the Netiquett:

1. Addressing or how many people we are writing to

If we send email to more recipients, it is more discrete to send copies as blind carbon copies. If you want recipients to know about each other – for example you are organizing a reunion of your class ... the particular addresses of your classmates should be separated by a comma.

2. Message subject

The task of a message subject is to capture recipient's attention (similarly to headlines in newspapers). If you write a suitable subject, the message will not disappear in the huge amount of reviewed messages, it is easier to handle the messages. **Subject is a brief and apt text, which characterize the content of the document.**

Example: To the editor-in-chief of the magazine "Gardener" you are sending an article about snail (slug) outbreak in your region.

Wrong: Message. Slugs have outbroken. Vermin outbreak.

These subjects are not clear, we do not know if the article is from a furious gardener or an environmentalist.

Better subject: Slug outbreak in the central Bohemia – an article for publishing.

If the email address is general, it is possible to state a particular recipient. For Ms Novak, subject: curriculum pro HPS. If a company states only one undetailed (general) email address on their www pages, it shows the amateur approach to customers.

Examples of wrong subjects: Message, Problem, Request, What shall I do?

3. Priority or importance of the message

To each message can be assigned so called priority, we can choose from three levels of importance: low-normal-high. Whichever priority we choose, it has no connection with the speed of delivering the message. It is only a notification for the recipient.

4. Open your email with a salutation

Even an email message should be opened with a salutation.

5. Text of the message

There is a rule "one topic, one mail", we should be brief.

It is not suitable to write in capitals in email. Typing in capitals is the equivalent of shouting. If we want to emphasize a word and we cannot use for any reason text formatting (for example the bold style) we can use the asterisk or the underscore character. * at the beginning

and at the end of a word means the bold type, the underscore character stimulates the italic type.

Handle emoticons with caution. 2-3 emoticons at the most in private correspondence, do not use them in business correspondence.

The Ten Commandments of Computer Ethics

Thou shalt not use a computer to harm other people.

Thou shalt not interfere with other people's computer work.

Thou shalt not snoop around in other people's computer files.

Thou shalt not use a computer to steal.

Thou shalt not use a computer to bear false witness.

Thou shalt not copy or use proprietary software for which you have not paid.

Thou shalt not use other people's computer resources without authorization or proper compensation.

Thou shalt not appropriate other people's intellectual output.

Thou shalt think about the social consequences of the program you are writing or the system you are designing.

Thou shalt always use a computer in ways that ensure consideration and respect for your fellow humans.

(from The Computer Ethics Institute, <http://www.computerethicsinstitute.com/>)

3.4 Gopher

The service Gopher was developed in the U.S., the University of Minnesota, slightly earlier than the World Wide Web. Then the two services competed together. Initially the Gopher had more servers. Around 1995 there were over 6,000 servers in operation worldwide. Over time, the Web got ahead the Gopher. The main reason for its defeat was a particularly unattractive "wrapper". Currently there are still some Gopher services working on the Internet that are maintained by a few enthusiasts.

More information about the service:

<http://galerie.earchiv.cz/videos/gophers.php3> - Videos about the Gopher service

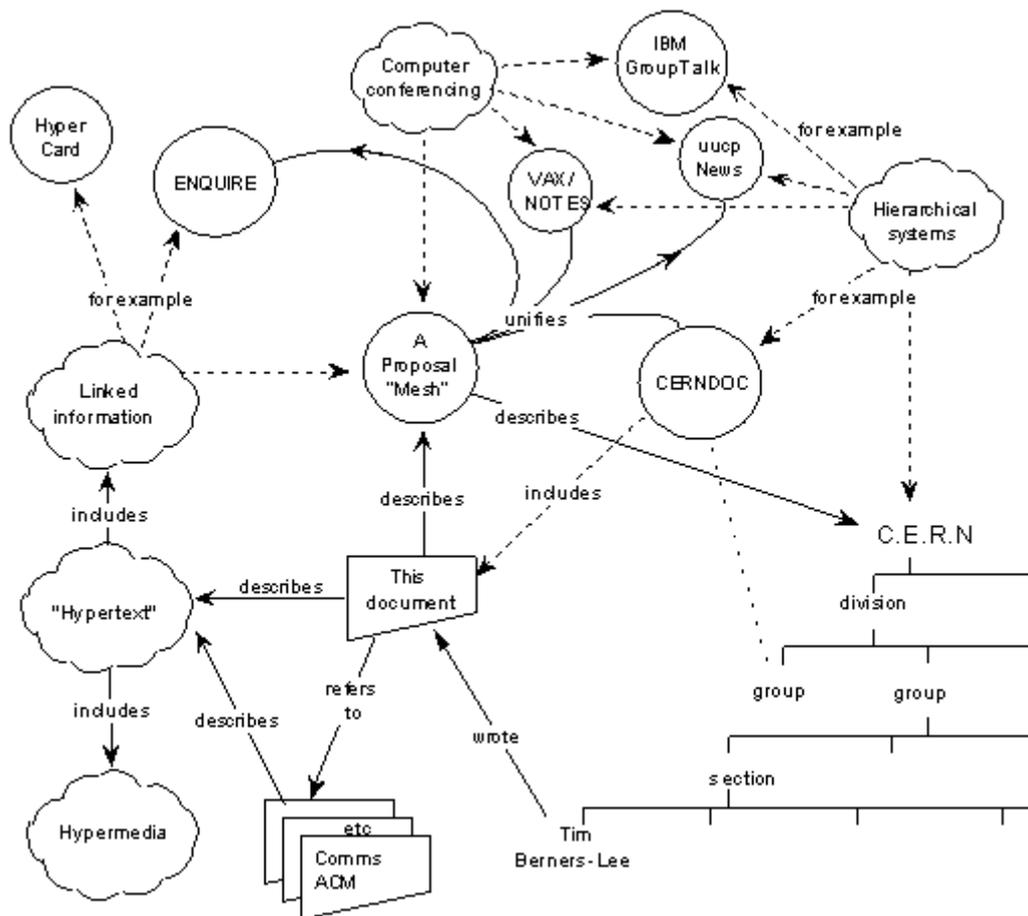
<http://galerie.earchiv.cz/scshots/gopher/index.php3> - Gopher Service - Screenshoty

3.5 World Wide Web - WWW

The most popular service on the Internet. The service started as a simple service providing interlinked documents access via the Internet. Nowadays it is a platform for providing all range of other specific services.

Brief history:

- 1945
 - Vannevar Bush – (1890 – 1974) published the article „As We May Think“ in the magazine The Atlantic Monthly [1], where he considers the problems of growing numbers of data. The article deals with the use of devices for communication, refers to the device Memex (MEMory and EXtended, extended memory), which operates on the principle of hypertext., prediction of the possible linking of several senses in communication with the device - now multimedia
 - 60th. of the 20. century
 - Doug Engelbart - Stanford Research Institute, ONLine System (NLS) – the first hypertext facilities
 - Ted Nelson – 1963 defined the concept of hypertext. In 1965 he published the concept of hypertext
- 1989
 - Tim Berners Lee, CERN – design of a web document



Picture No.5 : Information Management. A Proposal
 Source: <http://www.w3.org/History/1989/proposal.html>

The Web is an information system based on hypertext technology which provides end-users with a very progressive and attractive communication of electronic (digital) information sources (textual, image...). Electronic text sources are made on the basis of HTML and are physically stored on computers called web servers. Saving is not spontaneous. It has fixed rules. The user uses these documents in the communication web server / web client via the protocols HTTP. Optimal use of all www utilities ensure only clients (PC, terminal) working in graphics mode (eg, Netscape, Explorer, etc.) An access to documents is realized simply through the document address on the web registered in the form of a URL. The address can clearly identify any object on the Internet

3.6 Electronic conferences and discussion groups

A. Electronic conference

- works using e-mail
- conferees discuss pre-appointed topics
- after login the electronic address is added to **the list of participants**
- any letter sent to the conference address **is automatically forwarded to addresses of all participants** of a specific conference

Conferences can be divided according to the ways of entry distribution into:

unmoderated – each entry sent to a conference is automatically distributed

moderated – each entry goes through **the conference moderator**

Conferences can be divided according to the membership into:

open

close

How the electronic conference works:

The whole conference operation is managed by a **special software application**, which is continually running on a computer – a server.

The most often used mailing list managers are **majordomo, listserv** or **listproc**.

The program processes messages which were delivered to the so called **administrative** or **distribution** conference address

Administrative address enables participants to

- log in
- log out
- temporary log out
- get the list of participants of a particular conference
- get a list of conferences which the server conducts

Administrative address can be found in one of the following forms:

listserv@počítač.firma.cz

→ for example listserv@cesnet.cz

Only commands to the server are sent to the administrative address.

Distribution address

- is the address to which conference participants send their entries (text file)

How to subscribe to a conference:

1. To an administrative address for example

listserv@cesnet.cz

send a simple letter containing only the simple phrase:

subscribe title of the conference , first name and surname

How to unsubscribe

To an administrative address for example listserv@cesnet.cz send a letter with only the simple phrase:

unsubscribe the title of the conference

Commands:

subscribe,

unsubscribe/ signoff

the description of commands - help

the list of participants - who the title of the conference

to remain subscribed, but not receive posts: - set nomail the title of the conference

to receive mail again – set mail the title of the conference

For experimentation you can use for example these conferences

administrative addresses	their titles	Distribution addresses
listserv@cesnet.cz	demo-1	demo-1@cesnet.cz
listserv@cesnet.cz	cs-test	cs-test@cesnet.cz
majordomo@list.zcu.cz	Test	test@list.zcu.cz

B. Discussion groups

- the difference: at the electronic conferences “*everybody gets everything*” at discussion groups “*nobody gets anything*”

Discussion groups function on the same basis as web.chat, but all messages are stored on the server and they can be displayed any time.

C. Other communication services

MSN Messenger: <http://messenger.msn.cz/>

SKYPE: <http://www.skype.com/>

Yahoo! Messenger: <http://messenger.yahoo.com/>

ICQ: <http://www.icq.com>

Google Talk: <http://www.google.com/talk/>

Literatura:

1. BUSH, V. *As We May Think* [online]. July 1945, Atlantic Magazine [cit. 2010-04-02]. Dostupné z: <http://www.theatlantic.com/magazine/archive/1969/12/as-we-may-think/3881/>
2. *Čeština* [online]. [cit. 2010-04-17]. Dostupné z: <http://www.cestina.cz>
3. *Doug Engelbart Institute* [online]. c2008. [cit. 2010-04-02]. Dostupné z: <http://dougengelbart.org/>
4. NEJEDLÝ, J. – VRÁNOVÁ, D. *Výklad zákona o některých službách informační společnosti*. 1. vyd. Vyškov: TSM, 2005. 82 s. ISBN 80-239-5020-7.
5. PETERKA, Jiří. *Galerie archivu aneb Muzeum českého Internetu od roku 1994 do roku 1999*. [online] 2008. [cit. 2010-04-17]. Dostupné z: <http://galerie.earchiv.cz/gopher.php3>
6. *TechTrain* [online]. cTechnology Training Team. [cit. 2010-04-17]. Dostupné z: <http://www.techtrainteam.com/services/olt/internet/ftp-0.html#gopherheading>

4. INTRODUCTION TO INTERNET SEARCH STRATEGY

The Internet can be likened to a vast library in which to save someone a lot of jewelry world knowledge, practical manuals, guides and entertaining reading as well as many explicitly pulp literature .. Especially in a library, but destroyed all the files, registers or other types of breakdowns, all the magazines and books thrown in a pile. To be in this big pile of easily identifiable, we need search engines, search engine services that will help us solve the problem. With some exaggeration we can say that the Internet is indeed everything, but it is hard to find. By contrast, in libraries is just something, but that something is systematically arranged, and therefore it can be found easily. It is also important to realize that the Internet itself does not produce any information, only the transmission infrastructure, which allows you to work with information.

J. Peterka [1, 1998] states several reasons why works with information on the Internet is so complicated

- **lack of authorization and approval process**

Perhaps the biggest difference today from the Internet, traditional media is its accessibility to a large publishing entities. If someone wants to broadcast their own TV programs, shall establish its own television company, be licensed, equipped television studios, television get a transmitter, etc. If someone wants to publish a newsletter, has found the editor, find the printer, to get advertisers to raise publishing needed permits, etc. In contrast, published on the Internet can join. Cost and performance of other publishing your own information on the Internet is much lower than any other media. In addition, the Internet, the one who wants to access their own information to other people, do not need anyone to ask whether it may do or do not have to approve the way he wants to do. On the one hand, it's very democratic, on the other hand, there are things on the Internet, which borders the law.

- **lack of cataloging duties**

The absence of any approval is in addition to their ethical and moral implications of the major implications of a technical nature - no one does or does not affect publishers as in how a particular way to publish their information and what they observe while konvence.Chybí mechanism for the dissemination of information has created a character.

- **the problem of information assurance**

In the absence of any approval or authorization, it depends on the problem of assessing the credibility of information. So how do you build information on the Internet and how to evaluate them? A possible guide is the identity of sources of information were those who desert the information on the Internet - whether it is a reputable company or site on the page below which has never signed.

- **information quality problem**

Given the vast quantity of information on the Internet, there is a problem with the quality of information

- **lack of reporting obligations**

Internet has no owner, so there is nobody who would maintain a central repository of information received .. There are only partial records.

• **lack of absolute retrieval services**

The absence of any central records and the lack of notification requirements have major implications for all users who want to find something on the Internet. There is no mandatory central register, then, there is no search service that could accurately answer the question of whether anything concrete on the Internet exists or not. Conclusion: If the user is looking for specific information and does not find it using a search engine does not mean that no information on the Internet. You need to use another search engine and compare results.

4.1 How to search – tips for better searching

- 1 Be precise, ie. use multiple keywords. Search engines work better when you combine multiple keywords, which will distract the irrelevant pages. Request "Armstrong NASA Apollo Moon Landing" will take you to a result more easily than if you try the words one by one.
2. Read the instructions. Each search engine has its own syntax. If you insert the query in the right form, you will find what you're looking for, otherwise you will go through a pile of links that lead nowhere.
3. Use Boolean. Some search engines allow you to upload queries with boolean logic.
4. Be advanced. Many sites offer the advanced search with multiple passwords and categories. This can be useful if your query is too general.
5. If you fail the first time. Try another search.
6. Save your page. Once you find the page you are looking for, do not lose it again. At least save the bookmark.

4.1.1 Main steps of search strategy

For a successful search on the Internet it is important to learn the basic search strategy, which consists of the following steps:

1. Verbal formulation topic (coherent sentence or several sentences)
2. Choosing the right keywords that reflect the search topic, including synonyms and related terms
3. Preliminary formulation of a query, including decisions about whether to use a simple search or advanced search methods,
4. Selecting search engine
5. Help Reading the selected search service
6. Query formulation in accordance with the functions of the search service
7. Exploration and evaluation of search results

8. If you are not satisfied with the results, it should follow treatment (repair) and re-evaluate the query search results

9. If necessary, select a different search service

4.2 Search queries

4.2.1 Simple queries

A simple query is offered by all search engines and is known as Simple Query. The advantage of simple queries is that the user does not have to remember anything, nothing to learn. The disadvantage is the breadth of the results, the search engine returns thousands of links, most of them related to the information requirements of users.

A. Key words

Key words are very important to successful searching. One of the biggest mistakes that users make is the lack of key words. On average, users use two key words in a query. Key words can be categorized:

- Stop words (prepositions, conjunctions, etc., so words that do not have any information to search for meaning in most cases they are ignored by search engines.
- Objects (nouns)
- Actions (verbs)
- Modifiers (adjectives, adverbs, predicate)

What to do when choosing the right keywords?

- You should be as specific as possible.
 - the meaning of the word as narrow as possible
 - use phrases or complete sentences
 - use multiple keywords
- The first word is decisive
 - Discography Beatles or Beatles discography - the first query brings the different pages on the Beatles, where you can also find their discography. The second query will bring different pages of different bands discographies, where Beatles discography may also occur. When entering a query we must know what is more important for us.
- It is necessary to use unusual, significant words and their combinations
 - When you enter a word or phrase that is very common, occurring too often, we will return millions of results. For example, if you want information about the band The Frames and enter this word into the computer - the word is often used in English, it is better to enter a search name of the band members, song names, etc.

Other tips

Do not use synonyms in a search (for the simple search)

It is especially dangerous if we do not know what operator the search engine has predefined.

Browse away, where the results are promising.

Try it again and again, one of the rules is the Internet - the Internet is of much more than you can find

Find the right place is not always reach good search engine. If you re looking for a product and the price it is better to look into the store or auction center.

In the document, use Ctrl F (example: search for a word thrombocytopenia, appears a very long document, which is "somewhere" that word. When using the shortcuts do not read the whole document, but we cursor appears in the word

How to find synonyms - using synonyms is very important. We are looking for eg the word "language", but the site may be the word "speech" "speech". If we can not think of any synonyms, use utility Word. Move the cursor to the word that you want to know the synonyms, then press **Shift F7** appears Thesaurus.

C. Phrases

One way of the exact search is using phrases. Phrases are a combination of words, which must be present in the document exactly in the order in which they were entered by the user. The query is written using the quotation marks phrases such as "financial accounting".

The phrase should be used whenever a connection is established or phrase, such as Prague Spring, the Berlin Wall, the solar system, etc.

Possibilities of typing the phrase in Alta Vista:

"John Lennon"

John-Lennon

John;Lennon

John:Lennon

John,Lennon

4.2.2 Advanced queries

Boolean operators

In 1844 George Boole was given a gold medal of the Royal Society for his article on the operators. Boolean algebra is the basis for structured search techniques, it is also one of the cornerstones of modern computers. The basic starting point of Boolean logic, set theory. AND operator is equivalent to the set intersection operation, the operator OR corresponds to the unification of sets. For visual observations devote charts are used.

AND - the dates on both sides of the operand must also exist somewhere in the document. The AND operator can be advantageously used for concatenating the required terms or phrases in one query with the document for inclusion in search results to satisfy the requirement of simultaneous occurrence.

OR - in a document at least one of these terms must appear. The OR is equivalent to a simple sum. The documents that contain both the query, results are included only once. Simple reasoning can work out the argument that the guaranteed size of the result can never be less than the number of documents relevant to the most readable of these terms or phrases. The OR can be used for the concatenation of required terms or phrases in one query with the document for inclusion in the outcome of the Condition of the presence of any of them. With the OR operator should be treated very carefully. It is suitable in situations of search terms using the same or similar meaning (synonyms). Excessive use the OR operator leads to unmanageable number of documents found.

AND NOT - a document that contains the date indicated on the right hand operand is excluded from results. AND NOT operator excludes documents that contain the specified term or phrase. It covers only the term or phrases that are immediately listed for operator .. If we imagine the search terms and phrases set of documents in which they occur, then the operator AND NOT the difference between the two character sets. AND NOT operator must use great care to avoid exclusion of relevant documents.

NEAR - has a similar meaning as the default. But both terms must occur at some distance (usually measured in words) from each other. The condition of "close proximity" is evaluated as a maximum distance between the two dates. Most search engines is the maximum distance is set to ten words. The main advantage lies in the NEAR operator limitations of search terms, which usually means in terms of document structure to a sentence or paragraph. In this way, are excluded large documents with links to everything possible and impossible, but in terms of direct search of specific information have no value. It should be noted, however, also some disadvantages of the NEAR operator. Eg. query with "knowledge management" NEAR "information retrieval" may be excluded from the documents, which contain one paragraph

discusses the problem of information retrieval and only until the next paragraph speaks of knowledge management. It is very difficult to predict which method of writing, dealing with concepts, etc. chosen author. NEAR operator is tightening the AND operator. NEAR operator can be recommended in situations where the user is sure of the distance of search terms in the sentence, respectively. paragraph.

BEFORE - is similar to NEAR, but the difference is that the first term (on the left) of this operator must appear to the left of the second term in the document

AFTER - is similar to NEAR, but the difference is that the first term (on the left) of this operator must be present in the document right from the second term

Brackets - are used to influence the order in which they are evaluated the operators. Typically, first evaluate the nested operators and operators on the same level are evaluated from left to right, each term enclosed in brackets is known as a Boolean expression. (The entire query is always understood as a Boolean expression, irrespective of whether or not they are placed brackets around it). This approach is the same as applies in mathematics to evaluate algebraic expressions.

Using filters

Filters allow you to refine search results. Filter function is to limit the space for documents in which the search query. Filters are supported differently by different search engines. Very well thought out system of filters to AltaVista, see the following examples.

Filter node

Filter node allows you to limit the search scope to one specific node or a specific subset of nodes with the structure of the URL. Is the URL address of the location of any information resource on the Internet.

Example: at AltaVista a filter node can be set as follows:

Host: vodafone uk ..

Data filter

The filter data is useful in finding information that was published in a certain period. Entering data filter it is at most search engines usually one part of the advanced search form.

Filter Anchor - (anchor) is seeking documents that contain the specified word or phrase in the text hyperlink.

Filter Applet - documents are found at the specified applet in the Java programming language. For example: applet: Butterfly looks for pages containing the name of the Java applet butterfly, which appears in the applet tag. Applet = a program written in Java. JAVA - A programming language developed by Sun. Hl. Java difference is that a program written in Java (called an applet) running directly on a real computer, but the virtual machine, which of course is applied as a specialized computer program for real.

Filter domain - limit search only to documents of the said domain. For example domain: uk - will be found only documents from the Czech Republic

Filter guest - designed to find documents on a server: For example, typing typing: host: www.oavm.cz.cz are searched only documents that are stored on the server and the Academy of Business College Meziříčí

Filter image - the search for documents that contain the image file name. For example: image: spider.gif

Filter Link - meet documents, which are linked to the specified URL, for example, by entering link: www.oavm.cz will be found that somewhere in your body have a link to the server PC VM - that way you can identify citations.

Filter text - The text entered must appear in the text document (excluding images, links and URL)

Filter title - the word or phrase must appear in the document title.

URL filter - looking to the specification documents whose URL contains the specified text. On entering the following URL: trade documents are found on different servers that contain the word business in the name of the server, directory path, file name, or anywhere in the URL.

It is very important to realize that each search engine has its own search strategy, which since the days of AltaVista users greatly simplified.

Literature:

1. PETERKA, J. *Budoucnost práce s informacemi v Internetu. PragueInternet World, 1998* [online]. [cit. 2010-02-22]. Dostupné z: <http://www.earchiv.cz/l311/index.php3>
2. *Search Strategies for Search Engines* [online]. Poslední aktualizace 29. 3. 2008 [cit. 2010-02-22]. Available at: <http://webliminal.com/search/search-web05.html>

5. CLASSIFICATION OF SEARCHING SERVICES

On the Internet there is a huge amount of search services. Processing retrieval request it is necessary to take into account which service is the best. Below is a basic classification.

5.1 Resource List

Resource lists are single topic-oriented records. Currently there is no list of resource lists, it is very difficult to be well versed in them.

For example:

Daily summary of articles about digital broadcasting

<http://www.digimedia.cz/zdroje/>

Daily summary of articles from the world of mobile technology

<http://www.mobilar.cz/zdroje/>

Summary of John December's resources

<http://www.valeriemates.com/december/resources.html>

<http://www.december.com/john/>

5.2 Subject catalogues (Web portals)

Subject catalogues are huge polytematicly-oriented databases. They used to be maintained solely manually. Due to meeting information needs two targets are set [SKLENÁK, 2001, s. 254]:

- it tries to achieve a structured view of information resources on the Internet
- it tries to eliminate the disadvantages of robots that work mechanically – perceiving documents as sets of individual words robots do not understand the contents and they do not know what they are indexing.

Advantages:

- general Search
- quality assurance

Disadvantages:

- range of catalogues
- generality of descriptions
- structure of categories
- time consuming
- the validity of links

The most important Czech subject catalogues (horizontal, general)

[ATLAS.CZ](#)

[Caramba!](#)

[Centrum.cz](#)

[VOLNÝ – NAJDI.TO](#)

[Seznam](#)

[QUICK.CZ](#)

The most important international web portals (horizontal)

Yahoo www.Yahoo.com

Open Directories Project <http://www.dmoz.org>

Galaxy <http://www.galaxy.com>

LookSmart <http://www.looksmart.com/>

About.com <http://home.about.com/index.htm>

GoGuides <http://www.goguides.org>

JoeAnt <http://www.joeant.com>

Currently there is a growing number of so-called vertical portals (vertical portals, vortals), which are focused on a particular group of users (eg, "Community Portals serve users from a specific region) or thematically oriented, again with regard to the interests of certain groups of users.

Examples:

Chem Web - provides access to information in chemistry and related fields. There is an easy access to abstracts, reports from conferences, newsletters and books. Is it possible to register for free.

Medscape - information from the fields of medicine, review articles, book reviews, expert columns, patient education, etc. Free registration.

[CNET – the source for computing and technology](#)

[Linux Online](#)

[Minerals Engineering International Online](#) – The Internet Magazine for Mineral Processors and Extractive Metallurgists

[PHPDeveloper.org](#)

[PSIgate](#), Physical Sciences Information Gateway

5.3 Search Engines

Search engines are a great tool in locating information on the Internet. Their use is very simple, they offer “simple” or “advanced” searching mode.

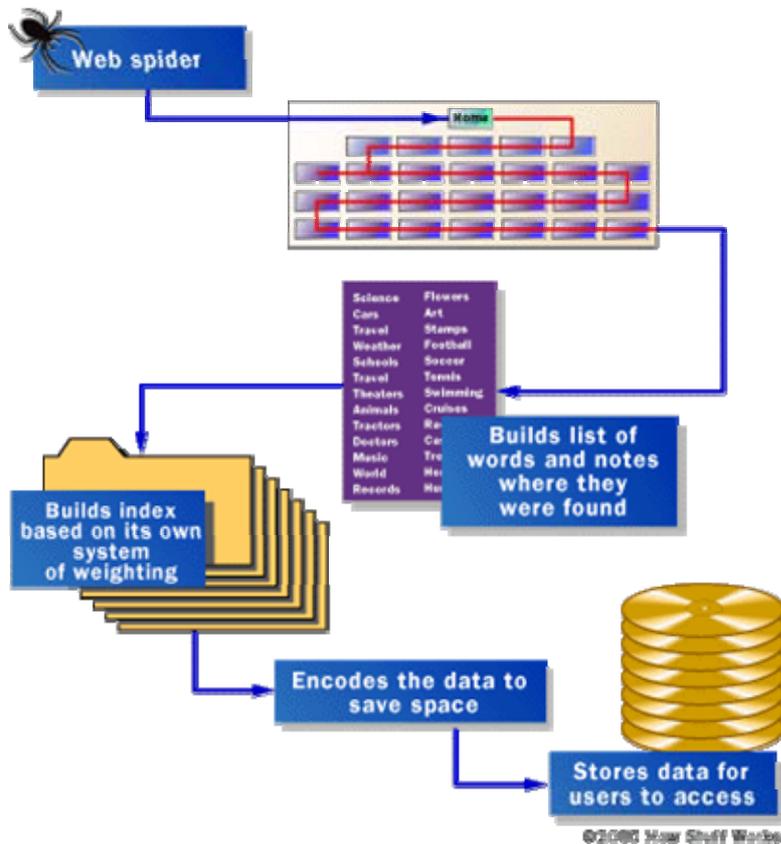
Search systems consist of three components:

- robot
- index
- own search engine

The robot - during milliseconds it scans Web pages depending on the keywords you enter, then it displays on the screen a list of sites which best correspond the conditions specified by the user.

Index - the results of the search spider are processed into the index page. It is a page that is created from individual Web pages that contain embedded keyword or search term. Search engines use an algorithm, which is essentially a particular set of math standards. Thus, each page is stored on the web world, tested in relation to key words or search terms that were used for searching. Until recently, the rule applied, the higher the keyword density in your page content, the higher the inclusion in search engine listings. But today is no longer valid (even though many SEO experts that still persists). Google (and most other search engines) now uses so-called semantic search, ie the evaluation of sites is directly dependent on how the page content matches the keyword (or term). For each word, which is part of the index, a list of links to Web documents. Structure of stored data stored in the index depends on the particular search engine, respectively. its query language.

Custom Search Engine - a user interface in which user queries.



Source: FRANKLIN, C. [2]

Examples of search engines:

Google <http://www.google.com>
 AltaVista <http://www.altavista.com>
 HotBot <http://www.hotbot.com>
 Lycos <http://www.lycos.com>
 Gigablast <http://gigablast.com>
 Nothern Lihgt <http://www.nothernlight.com>
 MSN <http://www.msn.com>
 BAIDU <http://www.baidu.com>

5.4 Meta-search Engines

Meta-search engines do not have their own databases but they send off requests to several other search engines, which means that searching is carried out automatically and simultaneously in several systems. Most meta-search engines combine the results, remove duplicates, and present more references than we would get from any individual server.

Meta-search can be sorted into two groups depending on where they are installed. The first group consists of meta-search engines installed on servers and constantly available to clients free of charge (MetaCrawler, Mamma, etc.). The second group consists of meta-search engines intended to be installed on client computers (desktops). These tools are provided for a fee. They are usually offered in two versions: a cheaper PLUS and more expensive and fully-equipped PRO. They enable query preparation off-line and they are equipped with tools for the management of ready retrieval (selection, sorting by other criteria, automatic updates, etc.). They have a predefined list of portals, catalogues and search engines, which the user can choose from and which can be involved in searching. For example Copernic belongs to the group of desktop meta-search engines.

Advantages:

- The user does not need to learn the search strategies of search engines
- He does not need to monitor the emergence of new systems
- Higher search completeness

Disadvantages:

- users are deprived of special search capabilities, which can not be generalized
- users lose information of advertising nature
- BE CAREFUL: meta-search engines take from each parallel system only 10-50 hits.

Examples:

Vivisimo <http://vivisimo.com>

Kartoo <http://www.kartoo.com>

Metacrawler <http://www.metacrawler.com>

Mamma <http://www.mamma.com>

Ixquick <http://www.ixquick.com>

Fazzle <http://www.fazzle.com>

5.5 Other types of search tools

Other search services specialize in search of a particular information space:

- according to the application or service
- according to the types of information
- according to geographic point of view - local services which collect for example only information about resources available in a particular geographical or linguistic area.

Search for people or personal home pages

<http://anywho.com>

<http://www.bigfoot.com>

<http://www.worldemail.com>

Corporate information

<http://centrumfirem.centrum.cz//search/index.php>

<http://yp.yahoo.com/>

Software, IT

<http://www.download.com/?legacy=cnet>

<http://www.jumbo.com/>

<http://www.pegasoft.cz/index.php>

Search for pictures

<http://gallery.yahoo.com/>

<http://www.picsearch.com>

Journals and journal articles

<http://journalseek.net/index.htm>

Library catalogues

<http://www.libdex.com/>

5.6 Virtual Libraries

Databases of these services are created by specialists by whom information is not only collected, but also selected, described and evaluated on the basis of certain criteria.

Virtual libraries are useful for those who look for high-quality sources on a certain topic.

Examples of virtual libraries

About.com

[Internet Public Library](#)

[World Wide Web Virtual Library](#)

5.7 Invisible web

As Mgr. Tkačíková quotes [6], excluding publicly indexable Web documents the Internet provides access to vast amounts of additional information which is called invisible web, or sometimes deep web.

- such information which is stored in databases which are generated dynamically on the basis of user interactions with the system (for example online library catalogues or bibliographic databases)

- such information to which an access is protected with a password and which is available only to authorized users often on a commercial basis (for example electronic version of magazines), <http://www.sciencedirect.com/>)

The difference between “visible” and “invisible” web are illustrated by the following two examples of information sources (the first aims at civil engineering matters, the other on education and training matters)

„visible Web“: iCivilEngineer.com

„invisible Web“: [Civil Engineering Database](#)

„visible Web“: [Educator's Reference Desk Resource Guides](#)

„invisible Web“: [ERIC Database](#)

The number of documents which can be searched by traditional search engines is up to five times higher than that of the Web information space which is accessible through search engines.

Such types of sources (particularly assorted databases) are registered by some specialized search services, for example [Complete Planet](#) or [Direct Search](#)

Elsevier Science Service Scirus [Scirus](#) – search for information both on “visible” and “invisible” web

Among other things, it allows searching for information on published journal articles and other professional publications from various sources (including full text searching of articles in [Science Direct](#)).

5.8 Social Networks

Systems of social networks are in fact a combination of specialized webhosting services and specialized search engines. A user fills his/her structured profile and immediately after that he/she can search and be searched, she can have the use of structured data in a precise format and items categorized into specific classifications

The social networks in the CR include Lidé.cz, Spolužáci.cz, Líbím se tri.cz, etc. Facebook and Twitter belong to the best known social networks in the world.

http://en.wikipedia.org/wiki/List_of_social_networking_websites

Social networks are becoming increasingly popular. Their myriad can be seen at:

http://en.wikipedia.org/wiki/List_of_social_networking_websites

Literature:

1. *European Search Engines, Directories and Lists* [online]. c1990 – 2010 Searchenginesoftheworld.com. [cit. 2010-03-12]. Dostupné z: http://www.netmasters.co.uk/european_search_engines/
2. FRANKLIN, C. *How Internet Search Engines Work* [online]. c1998 – 2010 HowStuffWorks, Inc. [cit. 2010-03-21]. Dostupné z: <http://computer.howstuffworks.com/internet/basics/search-engine1.htm>
3. GREŠKOVÁ, M. *Vyhľadavanie v Internete* [online]. Blog o vyhľadávani. [cit. 2010-03-12]. Dostupné z: <http://vyhladavanie.blogspot.com/>
4. *Search Engine Guide: the small business guide to search marketing* [online]. c 1998 – 2010 K. Clough, Inc. Dostupné z: <http://www.searchengineguide.com/searchengines.html>
5. *SearchEngineWatch.com* [online]. c2010 Incisive Interactive Marketing. Dostupné z: <http://searchenginewatch.com/>
6. SKLENÁK, V. *Data, informace, znalosti a Internet*. 1. vyd. Praha: C.H. Beck, 2001. 507 s. ISBN 80-7179-109-0.
7. TKAČÍKOVÁ, D. *Co je neviditelný web* [online]. Ústřední knihovna Vysoké školy báňské. [cit. 2010-03-21]. Dostupné z: <http://knihovna.vsb.cz/kurzy/hledani/04.html>

6. GOOGLE.COM

The search engine Google is without any doubts the most widely used web-based search engine (see the statistics <http://searchenginewatch.com/>).

Thinking up the title for their search engine Larry Page and Sergey Brin were inspired by the English term “Googol” which was made up by Milton Sirott.

The term was popularized in the book „Mathematics and the Imagination“ by Edwarda Kasnera and James Newtona. „Now her eis the name of a wery large numer „Googol“. *Most people say: „ A Googol is so large, that you cannot namet or talk about it, it so large that i tis infinite.* [KASNER, 2001, s. 20]. It is a mathematical term for the number represented by the numeral 1 followed by 100 zeros. The use of the term reflects their mission to organize a seemingly infinite amount of information on the web.

6.1 Chronology of crucial events

March 1995 – first meeting of Sergey Brin and Larry

January 1996 - December 1997 Larry Page and Sergey Brin began collaborating on a search engine called BackRub.

7. September 1998 – Getting the basic capital of million dollars from parents and “marvellous” investors (Andy Bechtolsheim Sun Microsystems Inc.) Brin and Page stopped studying and set up the company Google located in Mountain View in Silicon Valley south of San Francisco in California. The areay is nicknamed Googleplex.

February – June 1999 – Google got the investment worth \$ 25 million from Sequoia Capital and Kleiner, Perkins, Caufield&Byers, venture capital funds. Representatives of both funds headed for the Google management.

May 2000 - The search engine carried 18 million operations a day and became the most widely used service on the Internet.

6. August 2001 – Eric Schmidt (the former president of Novell and the technological manager of Sun Microsystem) was appointed the chief executive officer of Google. Brin and Page became presidents for technology or rather for products.

January 2002 – The firm started selling searching applications which enabled searching information without own intranet.

March 2002 – Google launched the betaversion of the program Google News

May – August 2002 – the company opens an office in Paris, expands its existing representation, which is in London, Tokio, Hamburg and Toronto. Google signs a contract for technological and advertising cooperation.

September 2002 – Google expanded advertising sale according to key words to Great Britain, France, Germany and Japan.

January 2003 – Google agreed on acquisition of the company Pyra labs, the creator of the successful web system Blogger.

March 2004 – Google launched Gmail

6. January 2004 . Yahoo does not want to use Google as its main search engine and they are planning to develop their own system.

29. April 2004 – Google asked the SEC for primary share issue worth about 3.3 billion USD

13. August 2004 Google closed to receiving orders for shares and their auction starts

2006 Google is officially represented by a subsidiary in the Czech Republic

2006 Google buyes youTube.com for 1.65 billion US dollars

2008 youTube.cz (the Czech Republic is in the elite group of 22 countries)

September 2008 – new browser Google Chrome

6.2 Search Strategy Google.com

- Default operator: AND
- Google does not distinguish small and capital letters
- Google recognizes phrases in quotation marks
- OR – the only Boolean Operator which Google detects

The plus sign +

(Google has the default operator AND. In case that some words cannot be marked as keywords, the system ignores them automatically, so the plus sign + is necessary. For example when we want to search the play Romeo and Juliet, Google removes the word “and”, so the results cannot be good. You can get for example Romeo Og Julie which is in Danish language. So we use the operator AND Romeo + Juliet.)

- The sign minus –
Google does not use the operator NOT but the sign minus. The sign minus must be written close in front of the word. There cannot be a blank space.

Results

The list usually displays parts of pages which contain three lines.

- If there are three dots (...) the text lies somewhere within the page, not at the beginning. In this case, you need to use keyboard shortcuts Ctl + F and enter again one of the search words, the cursor is transferred onto the search word.

- Sometimes we see in the link line indented to the right. Both items indented and indented above are similar, have similar primary domain. Google will link this to imply and say, "This is not just a single aspect of the subject searched, but probably a larger site, where about what you are looking for deals with multiple pages. Moreover indented under reference see the link "**More pages from xxx.xxx.cz**" click on this link page you can ride. After opening this link in the search box we can see **site: normale.buty.cz** is a filter that says that the following words will be found only on the Web site

Filters in the text entry

Filters: **cache, link, related, info, spell** cannot be used in combination with standard search commands, they are specific Google functions.

Filters: **site, allintitle, intitle, allinurl, inurl** are so called **modifiers of inquiries**

cache: URL

We use the command if we want to find a saved version of some Web sites. It is very handy when we, for example, display a page which is unavailable.

Example: cache: www.oavm.cz

link: URL

It finds out on which web pages can be found a link to the wanted address.

Example: link: www.oavm.cz

A very good filter for detecting citation

related: URL

The command will display pages that are similar to the site, the results are not usually very effective, sometimes rather curious. Not much used function.

Info: URL

This command displays information about the address, but very poorly, the command is almost unusable.

site: URL

If you enter this command in the search term, search results are limited to a given address, and address embedded in it. This term is used also in reducing the need to search only one domain, such as com

allintitle: search term

It searches all words of the search term in the title of a web page

Example: allintitle: obchodní akademie

intitle: search term

Finding at least one search word in the title of a web page

Example: intitle: obchodní škola

allinurl: search term

It searches all words search term in the address of a Web page

Example: allinurl: secondary school entrance exams

inurl: search query

Finding at least one search word in the website address

Example: inurl: secondary school entrance exams

Filetype: key word document format

Searches for the specified keyword in a document format

Example: filetype: ppt Internet

(Note: Gmail has it well-worked out searching strategy – see <http://mail.google.com/support/bin/answer.py?hl=en&answer=7190>)

Google services in Czech:

[AdWords Editor](#), [Blogger](#), [Dokumenty Google](#), [Gmail](#), [Google AdSense](#), [Google AdWords](#), [Google Analytics](#), [Google Apps](#), [Google Chrome](#), [Google Desktop](#), [Google Earth](#), [Google Pack](#), [Google Reader](#), [Google Scholar](#), [Google Sky](#), [Google Toolbar](#), [Hlášky Google](#), [iGoogle](#), [Kalendář](#) , [Google](#), [Mapy Google](#), [Nástroje pro webmastery](#), [Picasa - obrázkový editor](#), [Překladač Google](#), [Skupiny Google](#), [Slovník Google](#), [Vlastní vyhledávání Google](#), [Webová alba Picasa](#), [Weby Google](#), [YouTube](#), [Zprávy Google](#)

Source: *Official Google Blog Česká publika* [online], [4]

Literatura

1. *Google Guide: Making Searching Even Easier* [online]. [cit. 2010-04-21]. Dostupné z: <http://www.googleguide.com/>
2. ISKRA, J. *Google:vyhledávání, Gmail, Google Talk a další*. 1. vyd. Praha: Computer Press, 2006. 143 s. ISBN 80-251-1043-5.
3. KASNER, E., Newton, J. *Mathematics and the Imagination*“. Mineola, N.Y. : Dover Publications, 2001. 380 s. ISBN 0486417034.
4. *Official Google Blog Česká publika* [online]. c2010. [cit. 2010-04-21]. Dostupné z: <http://google-cz.blogspot.com/>.