

# History of the Internet

by Richard T. Griffiths



## Development of Computers till the 1960's

### How do we explain inventions and innovations?

The internet is an innovation (or rather a series of innovations) that enables communication and transmission of data between computers at different locations. It is an extremely new scientific development, but that does not mean that we cannot analyze it historically, using the concepts that we apply to other innovations in the past. Basically the discussion among historians about the causation of inventions tends to boil down to two main approaches:

**are inventions 'science' led, or are they determined by the material environment.**

The second approach that of the material environment, also usually devolves into two sub-questions:

**are invention determined by supply constraints, or are they called forth by demand.**

To illustrate this, let us turn to the early eighteenth century and the start of the industrial revolution.

The 'science school' faces a difficulty that many of the early inventions were not really scientifically based. The earliest textile machinery were made of largely of wood and depended for their success on various combinations of levels, pulleys and spindles. The steam-engine was admittedly more intricate, but it was not until the early 19th century that its principles were correctly described. And the series of anonymous advances in the size and shape of blast furnaces, their linings and the various mixes of fuels and ores all took place within the existing corpus of knowledge among iron-masters themselves.

Nonetheless, the science school emphasized the growth of learned societies, the rise of a new environment of experimentation and the growth of a corpus of experts working in these new areas (and the over-representation of non-conformist Scots among them is explained by the fact that because of their religion they were excluded from higher echelons of traditional careers). They emphasize the rise of a 'scientific method' rather than the role of formal science... the asking of new questions, the methodological pursuit of experiments and the insistence on scientific measurement. The material school that emphasizes supply constraints have a much easier time of it. The earliest textile innovations came in weaving which created a bottleneck in the spinning of yarn.

**History of the Internet**  
by Richard T. Griffiths



**Fig One: Flying Shuttle**

This, in its turn, led to a cluster of innovations in spinning (and later the introduction of steam-engines).

History of the Internet  
by Richard T. Griffiths

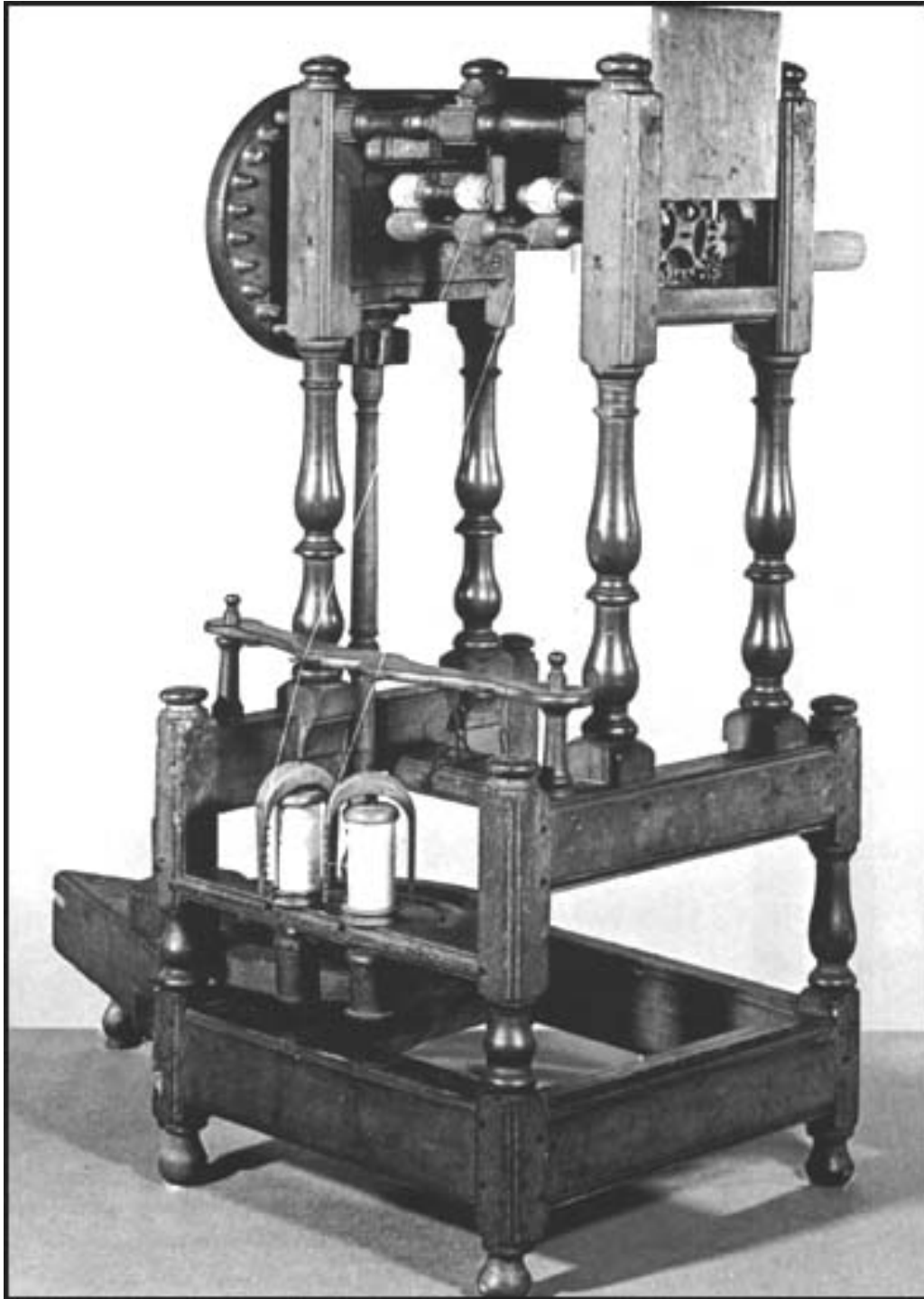
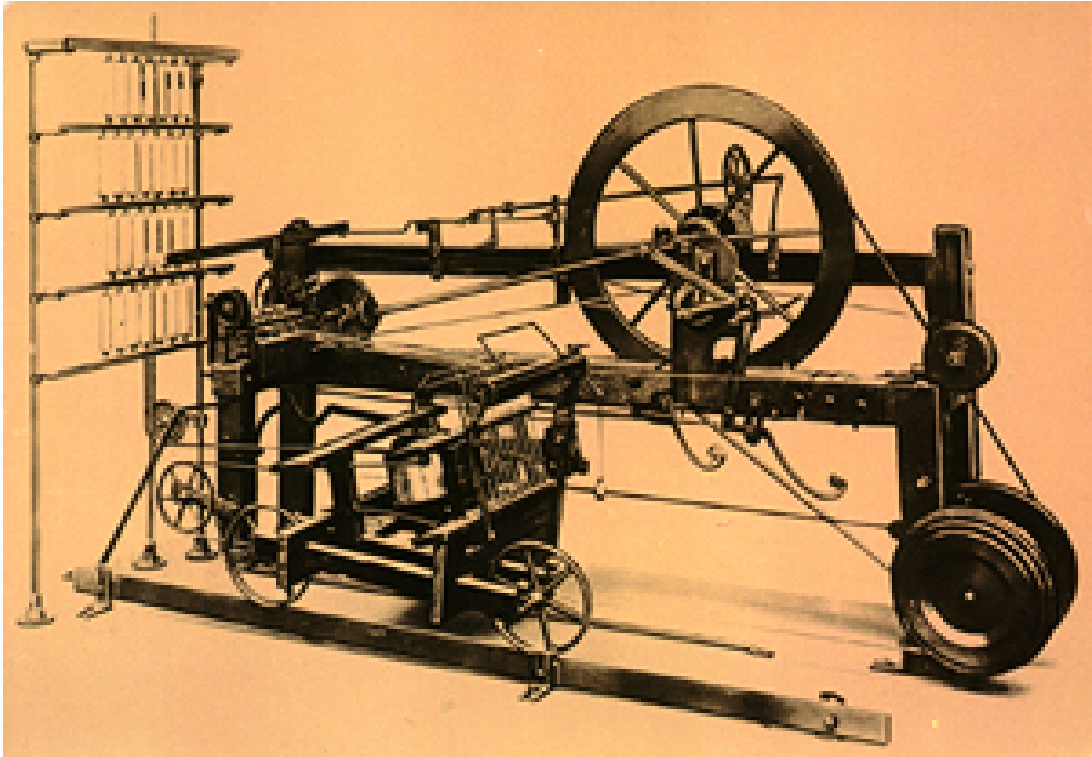
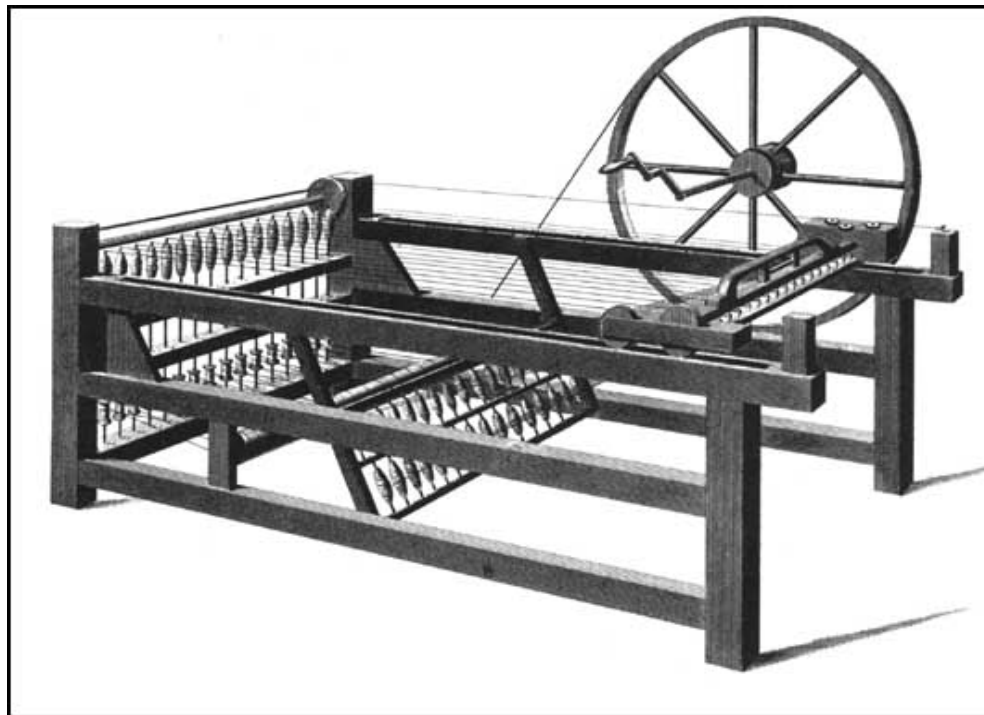


Figure 2: Arkwright's Water Frame (1769)

**History of the Internet**  
by Richard T. Griffiths



**Figure 3: Crompton's Mule (1779)**



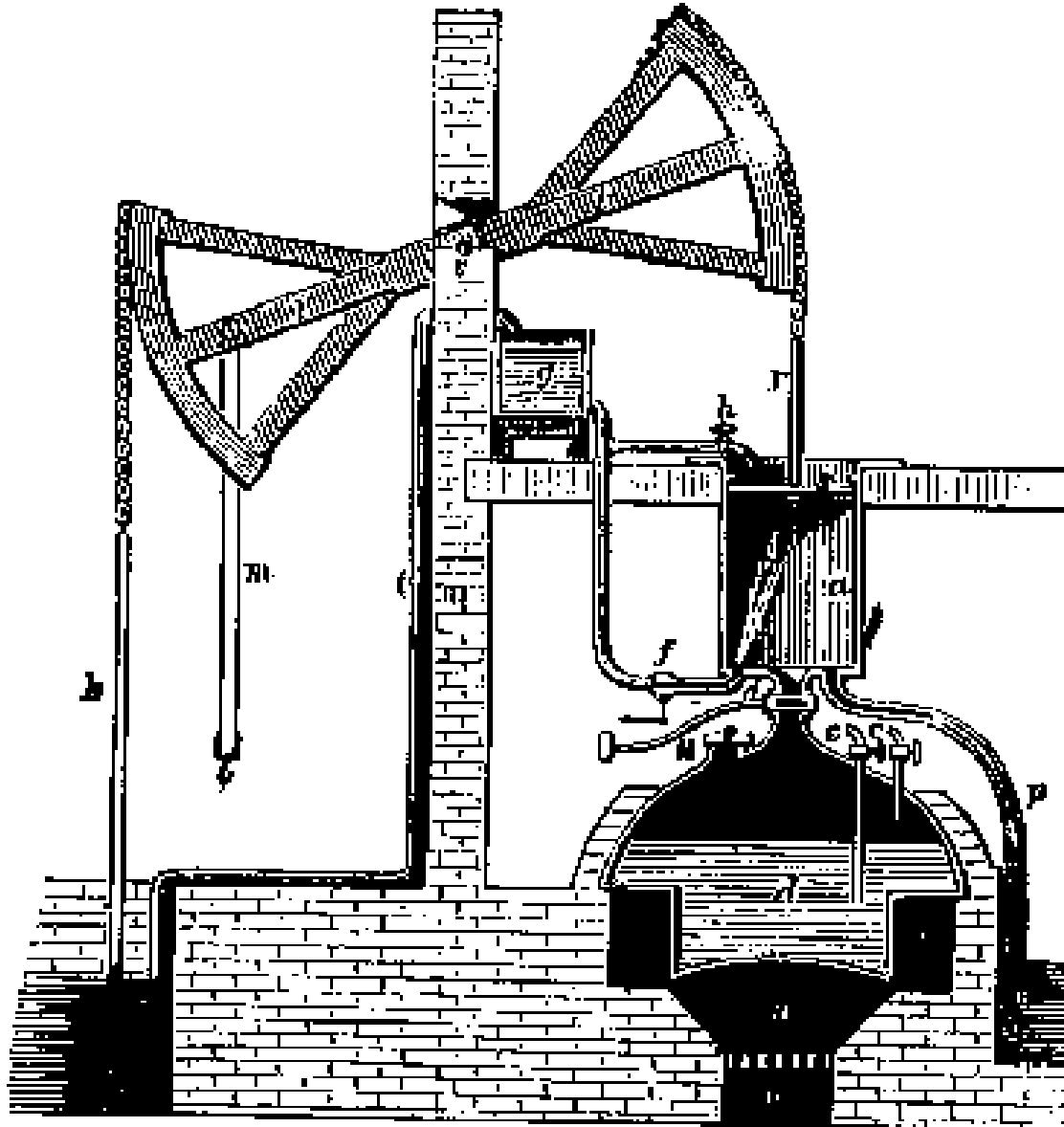
**Figure 4: Hargreaves' Spinning Jenny (1765)**

# History of the Internet

by Richard T. Griffiths

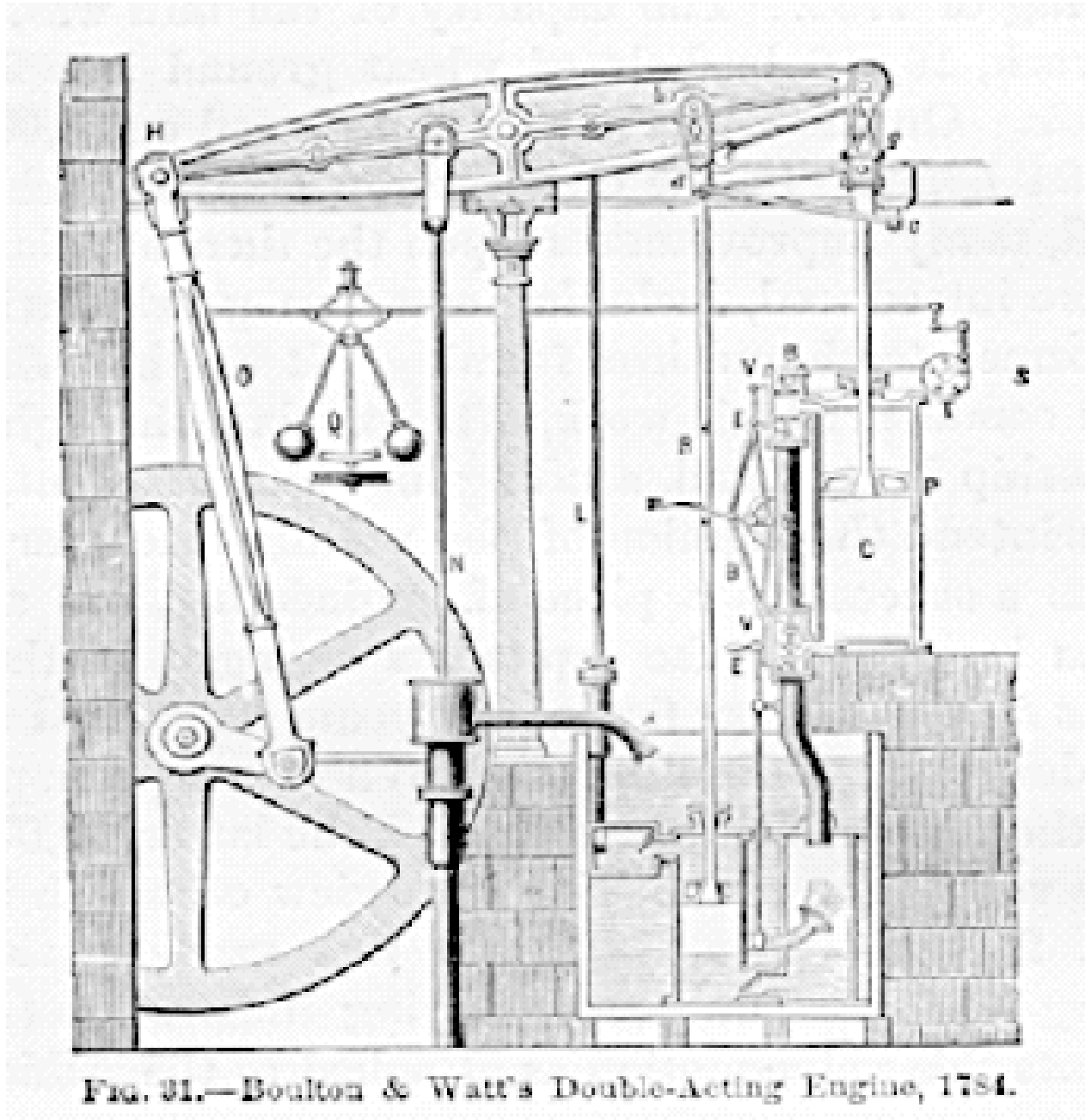
which created a new bottleneck in the weaving sector. And this led, in the 1830s to the invention, and continuous improvement, of power-looms.

The first steam-engines were pumping engines (water and air) introduced because the need for fuel meant the sinking of ever deeper mine shafts which had to be kept ventilated and unflooded. They were highly inefficient in terms of fuel consumption, which didn't really matter as long as they were situated at mine-heads.



Figures 5: Newcomen engine (1705)

History of the Internet  
by Richard T. Griffiths



**Figure 6: Boulton-Watt engine (1776)**

It was only when their efficiency was improved and adapted for a rotary motion, that they could be applied elsewhere.

The experimentation in blast-furnaces was equally dictated fuel shortages (caused by the disappearance of forests) and exhaustion of supplies of rich iron-ore. Thus, the argument goes, in each case shortages and bottlenecks drove up the prices of critical materials and stimulated the search for improvements.

# History of the Internet

by Richard T. Griffiths



**Figure 7: Abraham Derby's iron works at Coalbrookdale (1709)**

Before leaving the 'supply school' it is worth adding a caveat - some of these inventions depend on innovations elsewhere before they can be effectuated. For example, it is nice to invent a steam engine, but if you cannot make pistons and shafts to the right tolerances, or cannot make sheet metal of equal consistency or do not have appropriate welding techniques, you cannot build it - the works will jam and the boiler will buckle and explode.

The 'demand-siders' would argue for the importance of new markets. They would point to the population growth and urbanization in the United Kingdom in this period, to improvements in transportation and to development of overseas markets. This works best, of course, when dealing with consumer goods. For textiles they would point to the greater number of British inhabitants to clothe and the reduced opportunities for household production. They would observe the shift in preferences from woolen textiles to cotton and to lighter fabrics in general and they would certainly emphasize the coincidence of the development of spinning technology and the introduction of steam with the fact that Napoleonic Wars had given the British not only unfettered access to their own colonial markets, but those of France and the Netherlands as well.

# History of the Internet

by Richard T. Griffiths

The Industrial Revolution illustrates the problem for the historical analysis of inventions. Someone has to have the idea and unless we are going to be satisfied with a 'heroic inventors' school of explanation, we will be driven to searching for the social and scientific context. Yet an invention has to be feasible and it has to be applied if it is to have any effect. This means that it has to be worthwhile either because it allows one to do something better or cheaper (supply-side) or because it allows one to make more of something, or something entirely new (demand-side). It is impossible to isolate one factor, though the balance in the explanation varies according to the circumstance. While the supply-side explanation seems to have the stronger claims to 'importance' in the so-called 'First Industrial Revolution', this is far less in the 'Second Industrial Revolution', dating from the 1870s, based on chemicals and electronics and associated with not only with new production processes but with a whole new range of consumer goods, from motor vehicles to consumer durables, associated with higher real incomes.

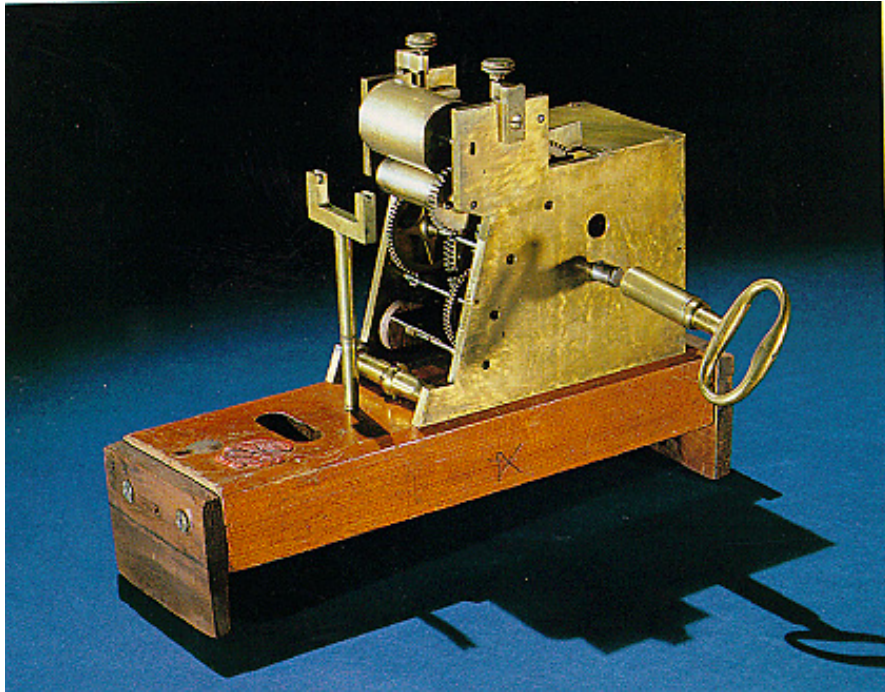
## The Development of Computers

The Internet is a system for allowing computers to communicate with each other. It goes without saying, that before we get the Internet we have to have computers. Much of the information for this section was derived from The Virtual Computing History Museum. The first step towards the modern computer was Samuel Morse's invention in 1844 of communication using electronic impulses, a key and a special code that sequences of pulses to letters of the alphabet. We won't get bogged down in whether Morse was actually the inventor of the telegraph (or his partner Alfred Vail in 1837) since you can find all you ever wanted to know about the topic at The Telegraph Office page.

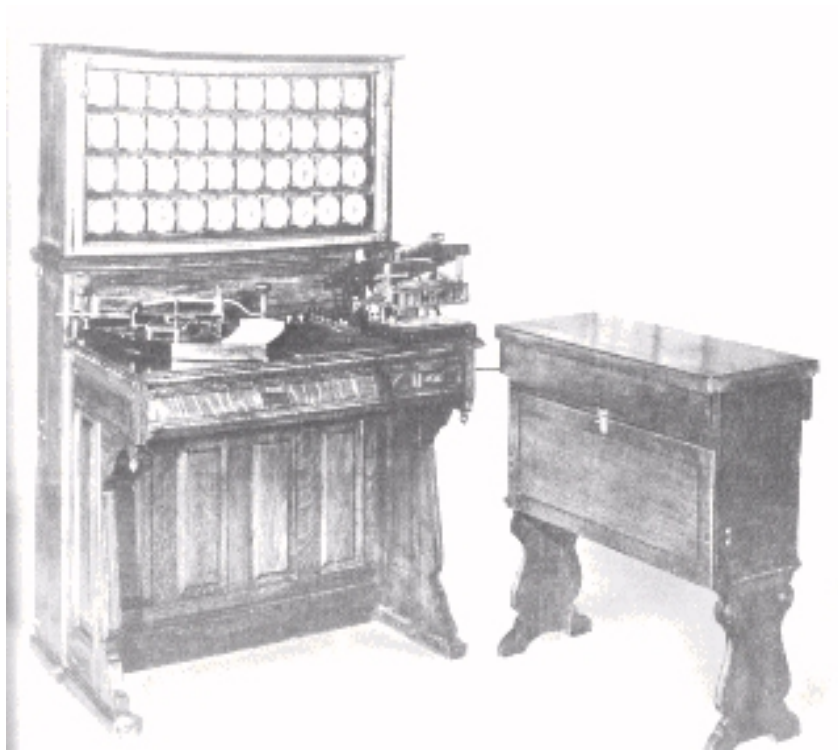


Figure 8: Babbage's Difference Engine (eventually built 1991)

**History of the Internet**  
by Richard T. Griffiths



**Figure 9: Morse's telegraph (1849)**



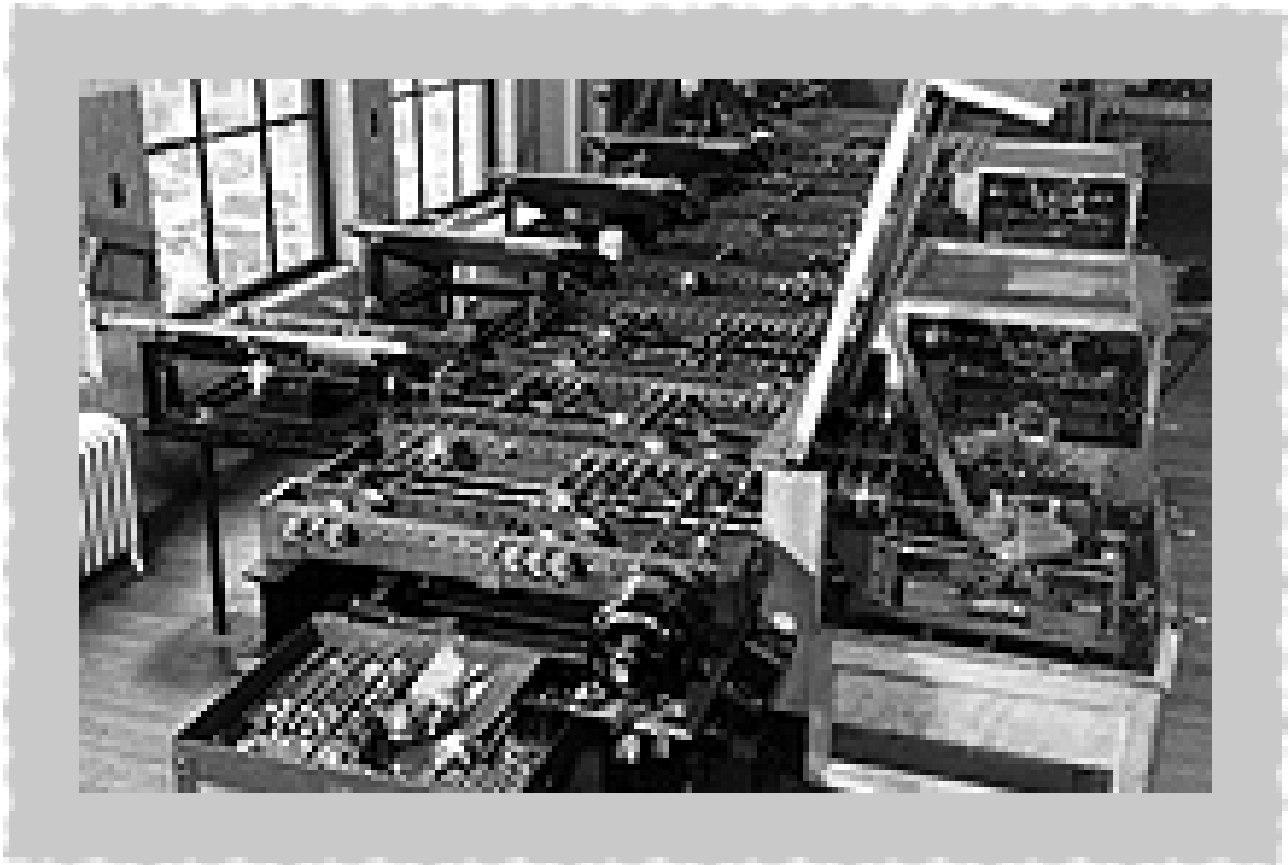
**Figure 10: Hollerith's Tabulating machine**

# History of the Internet

by Richard T. Griffiths

The next step is to link this particular invention to another of man's perennial strivings, the creation of a calculating machine. Although calculators have existed since the wire and bead abacus was first discovered in Egypt around 500 BC, one could say that the first main step towards the modern computer was Charles Babbage's experiments in the 1820s-1840s to build a "Difference Engine". I used the words 'one could say' deliberately because small errors in his calculations meant that Babbage never actually managed to build his engine - the Science Museum in Kensington built a copy of the Difference Engine in 1991 to celebrate the bi-centenary of his birth.

The idea of digital calculation was taken a step further by Herman Hollerith who developed digital processing machines to assist in compiling the 1890 US Census. Hollerith went on to found the Calculating-Tabulating-Recording (C-T-C) company in 1914, a company renamed IBM (International Business Machines) in 1924. Babbage's and Hollerith's ideas for digital computing, however, seemed to have led to a dead-end, with most scientists preferring to develop techniques for analog devices, based on slide-rule principles.



**Figure 11: Differential Analyzer**

These, too, could get pretty big as this Differential Analyzer built at the Massachusetts Institute of Technology (MIT) in the 1930s reveals. But machines of this size were also running up against the frontiers of their capabilities and, on the eve of the 1930s, new interest was being shown in digital devices. By now a whole host of devices associated with the development of the telephone (switches, relays etc) and radio (cathode tubes) would extend the possibilities of any solution. But what accelerated developments was the outbreak of World War II.

# History of the Internet

by Richard T. Griffiths

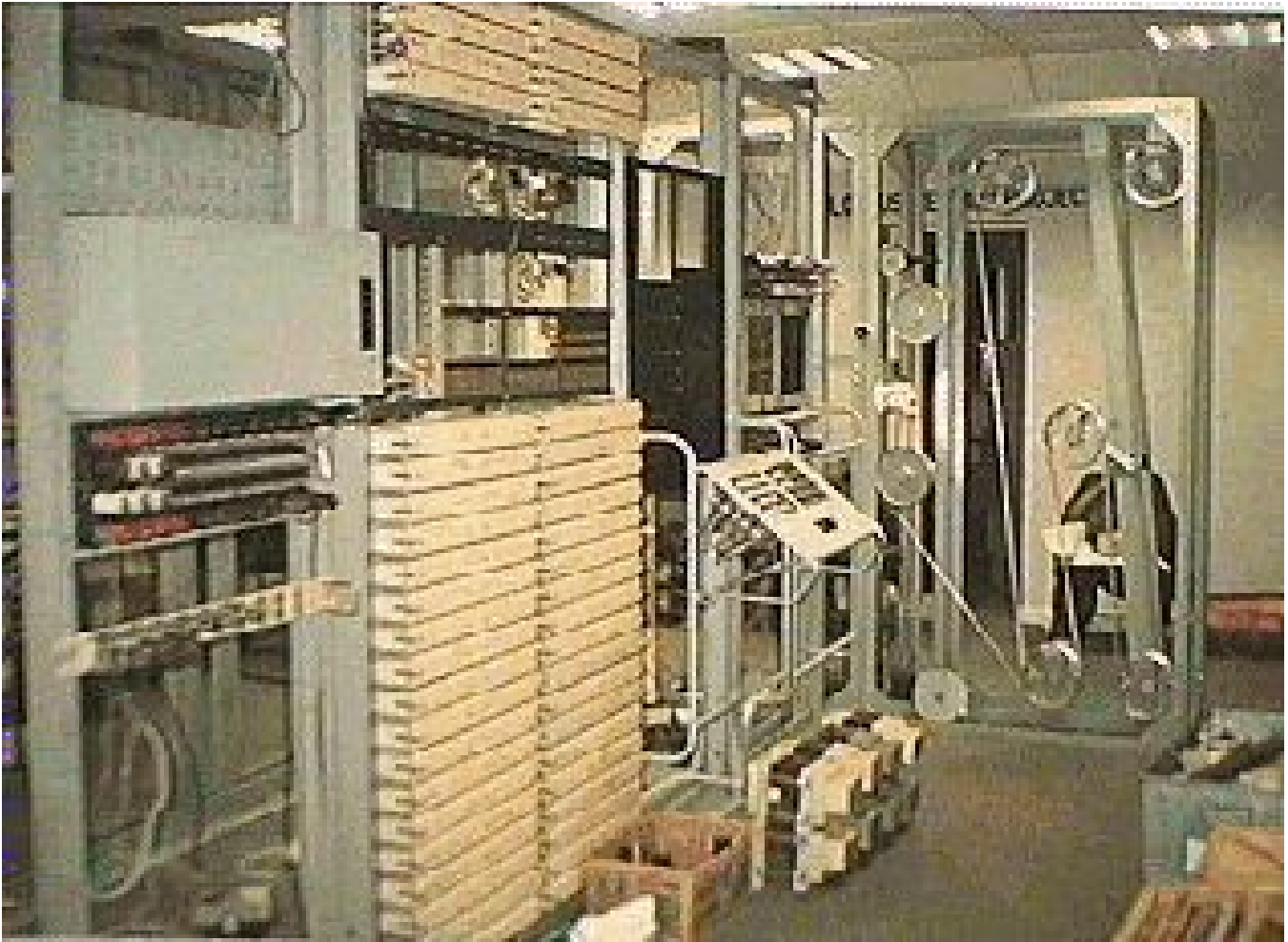
The war produced two major bottlenecks that were solved by digital machines. In the US, the need for gun-firing tables, navigational tables and tracking and aiming devices for anti-aircraft guns resulted in 1944 in the development of the first large scale automatic electromechanical calculator, the Harvard Mark I built by IBM. Note that it did not have an inbuilt program, the operating instructions were driven by a paper tape. A second crying need was to break the German (and Japanese) codes quickly enough to be useful. This work was undertaken by British scientists at Bletchley, and it culminated in the construction of the Colossus which became operational in 1944. This was more advanced than the Harvard Mark I, but its subsequent impact was limited by the fact that its very existence was a classified secret until 1970.



Figure 12 Harvard Mark I

# History of the Internet

by Richard T. Griffiths



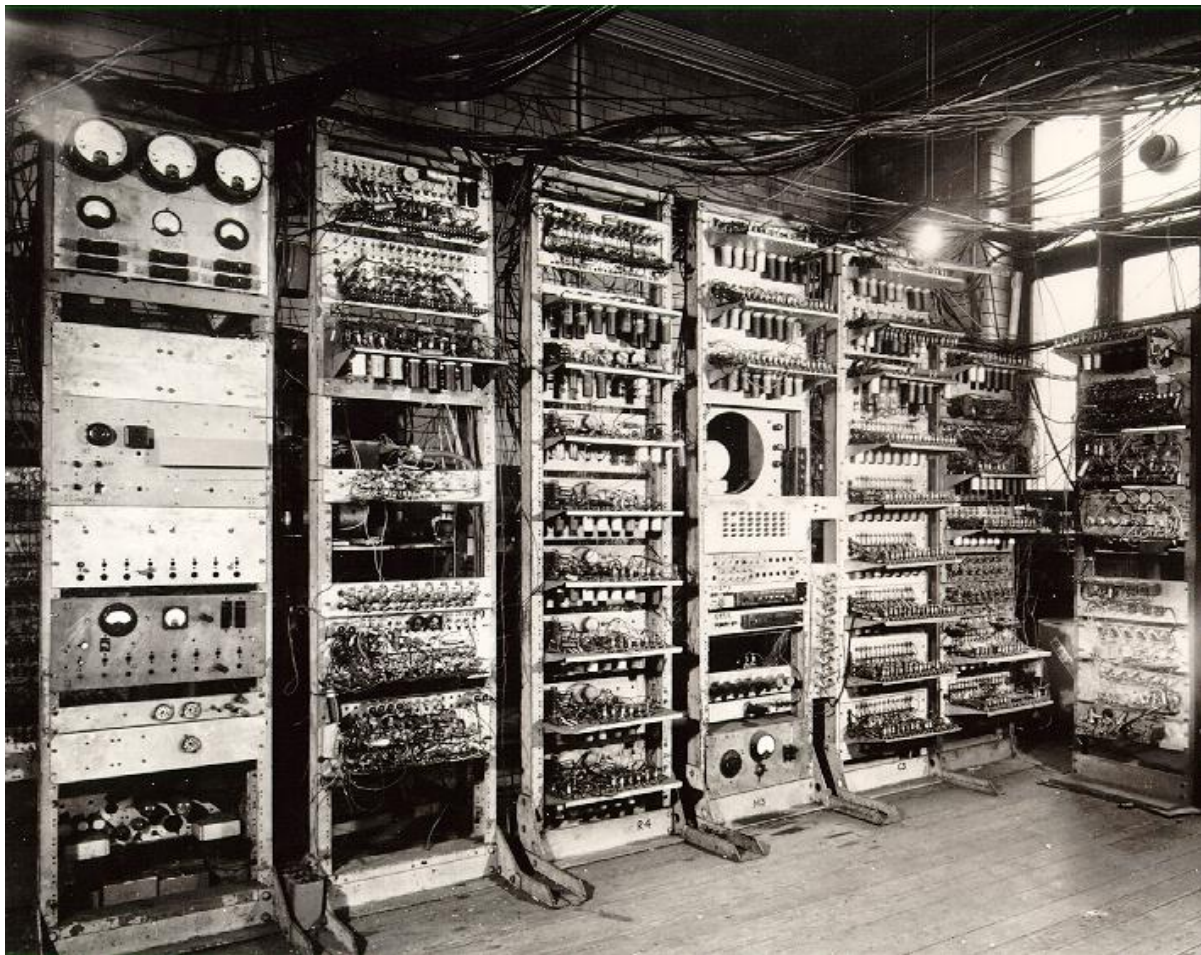
**Figure 13 The Colossus (1944 rebuilt 1994)**

The War had produced a considerable advance in design technology, but basically we were still at the stage of large and complex calculating machines. The challenge was to produce a device with an internal stored memory, a leap that would take us from calculators to computers proper. The war had also created a pool of scientists with experience in digital computing, and work in advancing technology proceeded rapidly on both sides of the Atlantic. If we are looking for the first modern computer, the credit should go to the Manchester University whose prototype, Baby, became operational in June 1948, followed soon by a full scale operational model, Manchester Mark I. The next major step, the incorporation of a Random Access Memory came three years later with the Whirlwind constructed at MIT.

**History of the Internet**  
by Richard T. Griffiths



**Figure 14 Manchester Mark I (1949)**



**Figure 15 Whirlwind (1951)**

# History of the Internet

by Richard T. Griffiths

Until now computer advances had been developed either for various branches of government or as prototype units within universities. In 1951 Remington-Rand entered the market with the UNIVAC computer, largely in an effort to recoup the cost over-run in its contract with the US government which had originally ordered the device for the census. A year later, it started producing ready-made software (although the term did not come into use until a decade later). IBM, which had previously specialized in punch-card systems, entered the market with its 700 series in 1953. Offering 60 per cent discount for educational uses, IBM quickly came to dominate the university market. Computers were now spreading quickly through the business and scientific communities, becoming ever faster and ever more user-friendly. They were also becoming smaller. By the end of the 1950s, transistors were beginning to oust cumbersome vacuum tubes and in 1958/59, the first 'integrated circuit' on a piece of silicon produced - five components on a piece 1 cm long. The 'chip' is born and entered into commercial production in 1961.



**Figure 16 IBM 7090 (1961)**

In 1961 IBM introduced a 'Compatible Time Sharing System' into its 7090/94 series which allowed separate terminals in different offices to access the same hardware. The concept of "remote access" to a "host" computer had become reality. And if you could link to one computer from a desktop terminal, why not to another.... why not to all?

# History of the Internet

by Richard T. Griffiths

## The Creation of ARPANET

To get to the origins of the Internet, we have to go back in time to 1957. You probably have no cause to remember, but it was International Geophysical Year, a year dedicated to gathering information about the upper atmosphere during a period of intense solar activity. Eisenhower announced in 1955 that, as part of the activities, the USA hoped to launch a small Earth orbiting satellite. The Kremlin announced that it hoped to do likewise. Planning in America focused on a sophisticated three stage rocket, but in Russia they took a more direct approach. Strapping four military rockets together, on 4 October 1957 the USSR launched Sputnik I (a 70 kgs beeping sphere the size of a medicine ball) into Earth orbit.

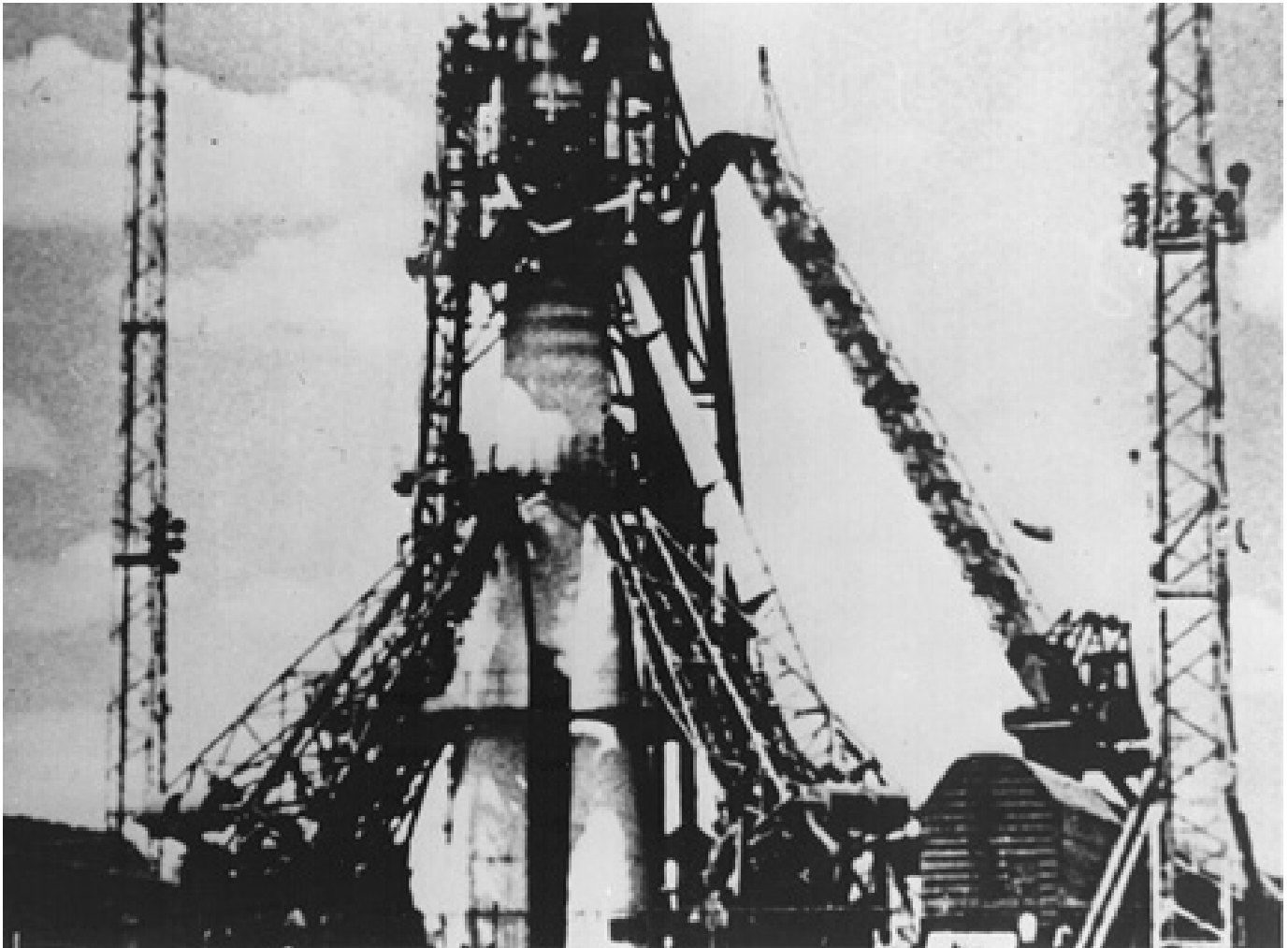
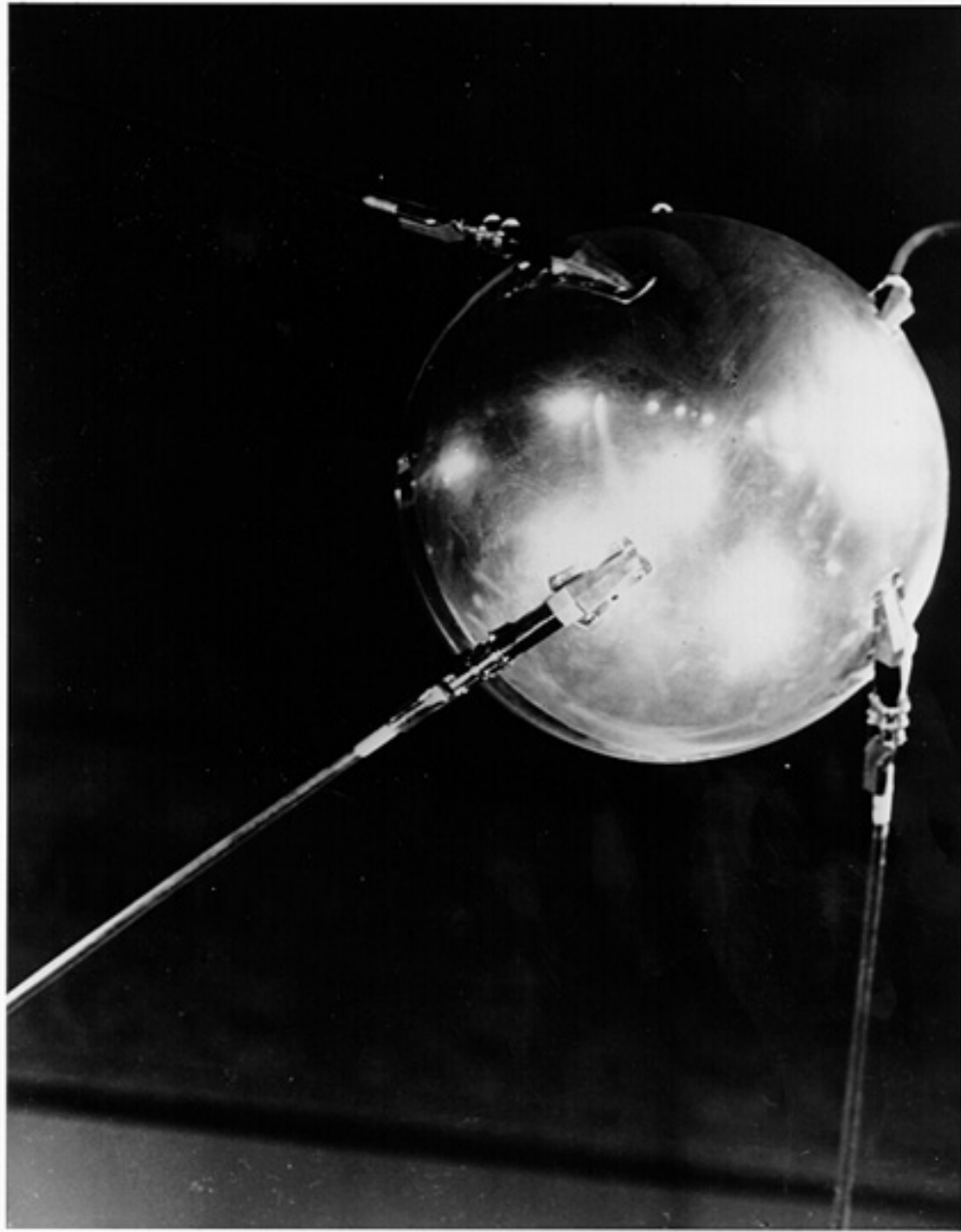


Figure 1: Sputnik launch (1957)

# History of the Internet

by Richard T. Griffiths



**Figure 2: Sputnik I**

The effect in the United States was electrifying, since it seemed overnight to wipe out the feeling on invulnerability the country had enjoyed since the explosion of the first nuclear bomb thirteen years before. One of the immediate reactions was the creation of the Advanced Research Projects Agency within the Ministry of Defense. Its mission was to apply state-of-the-art technology to US defense and to avoid being surprised (again!) by technological advances of the enemy. It was also given interim control of the US satellite program until the creation of NASA in October 1958.

ARPA became the technological think-tank of the American defense effort, employing directly a couple of hundred top scientists and with a budget sufficient for sub-contracting research to other top

# History of the Internet

by Richard T. Griffiths

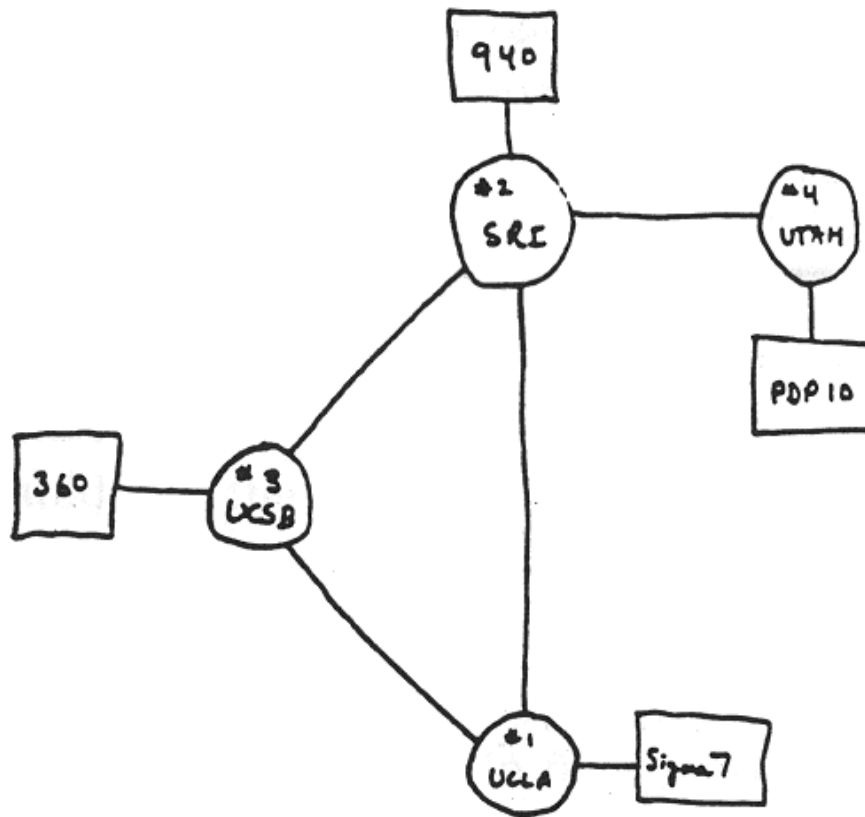
American institutions. Although the advanced computing would come to dominate its work, the initial focus of ARPA's activities were on space, ballistic missiles and nuclear test monitoring. Even so, from the start ARPA was interested in communicating between its operational base and its sub-contractors, preferably through direct links between its various computers.

In 1962 ARPA opened a computer research program and appointed to its head an MIT scientist John Licklider to lead it. Licklider had just published his first memorandum on the "Galactic Network" concept... a futuristic vision where computers would be networked together and would be accessible to everyone. Within ARPA, Leonard Klienrock was already developing ideas for sending information by breaking a message up into 'packages', sending them separately to their destination and reassembling them at the other end. This would give more flexibility than opening one line and sending the information through that alone. For example, the system would not be reliant on a single routing and, if files were broken-up before transfer, it would be more difficult to eavesdrop... both useful security advantages. The inadequacy of the telephone network for running programs and transferring data was revealed in 1965 when, as an experiment, computers in Berkeley and MIT were linked over a low speed dial-up telephone-line to become the first "wide area network" (WAN) ever created.

By 1966/67 research had developed sufficiently for the new head of computer research, Leonard Roberts, to publish a plan for computer network system called ARPANET\*\*. When these plans were published it became clear that independently of each other, and in ignorance of each other's work, teams at MIT, the National Physics Laboratory (UK) and by RAND Corporation had all been working on the feasibility of wide area networks, and their best ideas were incorporated into the ARPANET design. The final requirement was to design a protocol to allow the computers to send and receive messages and data, known as an interface message processor (IMPs). Work on this was completed in 1968, and the time was ready to put the theory to the test. In October 1969, IMPs installed in computers at both UCLA and Stanford. UCLA students would 'login' to Stanford's computer, access its databases and try to send data. The experiment was successful and the fledgling network had come into being. By December 1969 ARPANET comprised four host computers as with the addition of research centers in Santa Barbara and Utah. In the months that followed, scientists worked on refining the software that would expand the network's capabilities. At the same time, ever more computers were linked to the net. By December 1971 ARPANET linked 23 host computers to each other.

# History of the Internet

by Richard T. Griffiths



THE ARPA NETWORK

DEC 1969

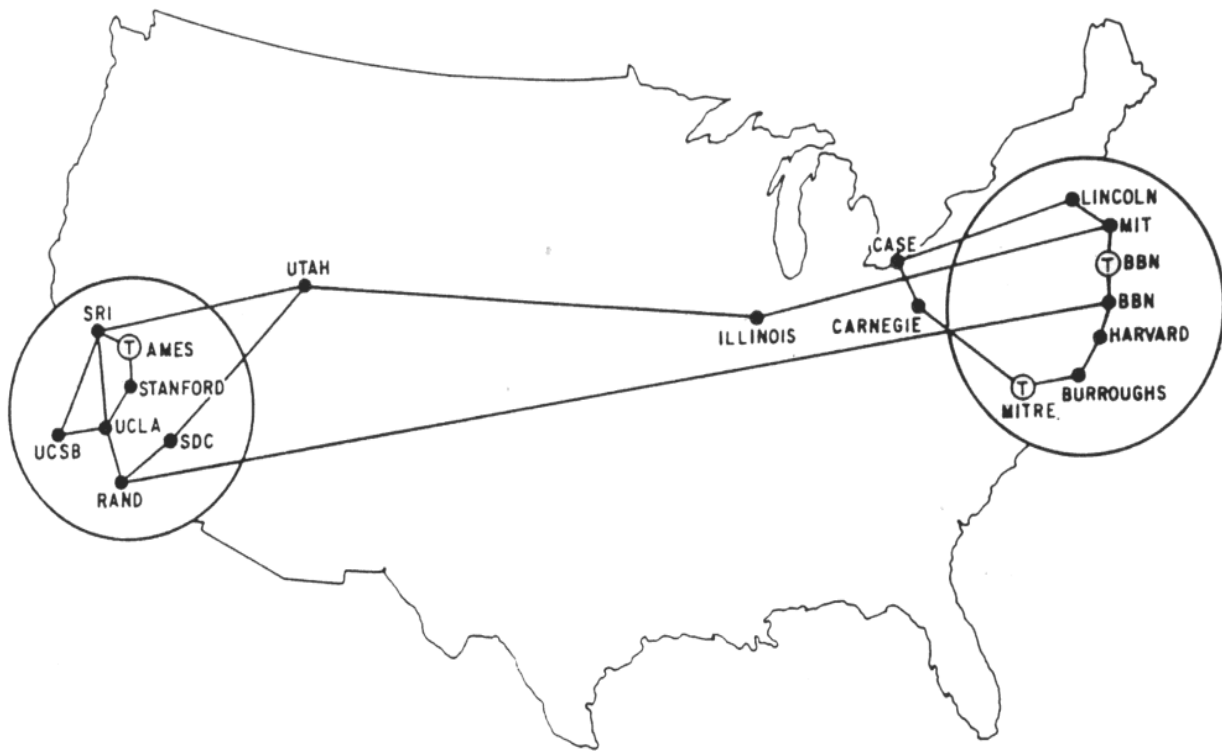
4 NODES

FIGURE 6.2 Drawing of 4 Node Network  
(Courtesy of Alex McKenzie)

Figure 3 ARPANET (1969)

# History of the Internet

by Richard T. Griffiths



MAP 4 September 1971

**Figure 4: ARPANET (1971)**

## Time for a Few Basics

Here we have the first true computer network. Since it is all still fairly basic, it is worth considering the underlying principles have basically remained the same (even if they, mercifully, operate far faster and look much prettier). We start off with a passive terminal and an active host, a keyboard and a computer. They are linked together by a cable. By typing in commands recognized by a computer, you can use the programs stored in its computer, access its files (and modify them and print them out as desired). Most people can envisage this arrangement within a single building, or complex of buildings.

In order to access another computer, at a completely different facility, we have first to reach it. This was usually done in these times over a (high speed) telephone line (or lines). Once you arrive at the new 'host' you have to convince it to treat you in the same way as someone behind a terminal within its own system. Hence the need of an interface message processor (IMP) and for the same IMP to be installed in both computers! Now you can access its files. Of course, order to preserve confidentiality, all computers differentiated between 'open' files and those that were password protected.

If you wanted to transfer a file or program to your own computer, the host computer uses a program to break it down into 'packages' attaching to each the address and its original position. It then sends them to your 'home' computer where a mirror program reassembles the message in the original order. In future, you could then access them from your home base. When dealing with a 'simple' network like

# History of the Internet

by Richard T. Griffiths

ARPANET it is difficult to see what the real advantage of this process was. But this would soon change...

## From ARPANET to Internet

In October 1972 ARPANET went 'public'. At the First International Conference on Computers and Communication, held in Washington DC, ARPA scientists demonstrated the system in operation, linking computers together from 40 different locations. This stimulated further research in scientific community throughout the Western World. Soon other networks would appear. The Washington conference also set up an Internetworking Working Group (IWG) to coordinate the research taking place. Meanwhile ARPA scientists worked on refining the system and expanding its capabilities:

In 1972, they successfully employed a new program to allow the sending of messages over the net, allowing direct person-to-person communication that we now refer to as e-mail. This development we will deal with at length in the next section.

Also in the early 70s, scientists developed host-to-host protocols. Before then the system only allowed a 'remote terminal' to access the files of each separate host. The new protocols allowed access to the hosts' programs (effectively merging the two host computers into one, for the duration of the link).

In 1974, ARPA scientists, working closely with experts in Stanford, developed a common language that would allow different networks to communicate with each other. This was known as a transmission control protocol/internet protocol (TCP/IP).

The development of TCP/IP marked a crucial stage in networking development, and it is important to reflect on the implications inherent in the design concepts... since it could all have turned out very differently. One crucial concept was that the system should have an 'open architecture', in fact implementing Licklider's original idea of a "Galactic Network":

Each network should be able to work on its own, developing its own applications without restraint and requiring no modification to participate in the Internet.

Within each network there would be a 'gateway', which would link it to the 'outside world'. This would be a larger computer (in order to handle the volume of traffic) with the necessary software to transmit and redirect any 'packages'.

This gateway software would retain no information about the traffic passing through. This was designed to cut-down workload and to speed up the traffic, but it also remove a possible means of censorship and control.

Packages would be routed through the fastest available route. If one computer was blocked or slow, the packages would be rerouted through the new until they eventually reached their destination. The gateways between the networks would always be open, and they would route the traffic without discrimination.

Also implicit in the development was that the operating principles would be freely available to all the networks. This freeing of design information was an early an integral part of the research environment, and greatly facilitated subsequent technological advance of the

It is worth remembering, at this stage, that we are still in a World where we are talking almost exclusively about large mainframe computers (owned only by large corporations, government institutions and universities). The system was therefore designed with the expectation that it would

# History of the Internet

by Richard T. Griffiths

work through a limited number of national (sub-) networks. Although 1974 marked the beginning of TCP/IP, it would take several years of modification and redesign before it was competed and universally adopted. One adaptation, for example, was that already in mid-1970s, a stripped-down version was designed that could be incorporated into the new micro-computers that were being developed. A second design challenge was to develop a version of the software that was compatible with each of the computer networks (including that of ARPANET itself!)

Meanwhile computer networking developed apace. In 1974 Stanford opened up Telenet, the first openly accessible public 'packet data service' (a commercial version of ARPANET). In the 1970s the US Department of Energy established MFENet for researchers into Magnetic Fusion Energy, which spawned HEPNet devoted to High Energy Physics. This inspired NASA physicists to establish SPAN for space physicists. In 1976 a Unix-to-Unix protocol was developed by AT&T Bell laboratories and was freely distributed to all UNIX computer users (since UNIX was the main operating system employed by universities, this opened up networking to the broader academic community). In 1979 Usenet was established, an open system focusing on e-mail communication and devoted to 'newsgroups' is opened, and still thriving today. In 1981 Bitnet (Because it's Time..) was developed City University New York to link university scientists using IBM computers, regardless of discipline, in the Eastern US. CSNet, funded by the US national Science Foundation was established to facilitate communication for Computer Scientists in universities, industry and government. In 1982 a European version of the Unix network, Eunet, was established, linking networks in the UK, Scandinavia and the Netherlands, followed in 1984 by a European version of Bitnet, known as EARN (European Academic and Research Network).

Throughout this period, the world is still fairly chaotic, with a plethora of competing techniques and protocols. ARPANET is still the backbone to the entire system. When, in 1982 it finally adopts the TCP/IP the Internet is born... a connected set of networks using the TCP/IP standard.

ARPANET GEOGRAPHIC MAP, OCTOBER 1980

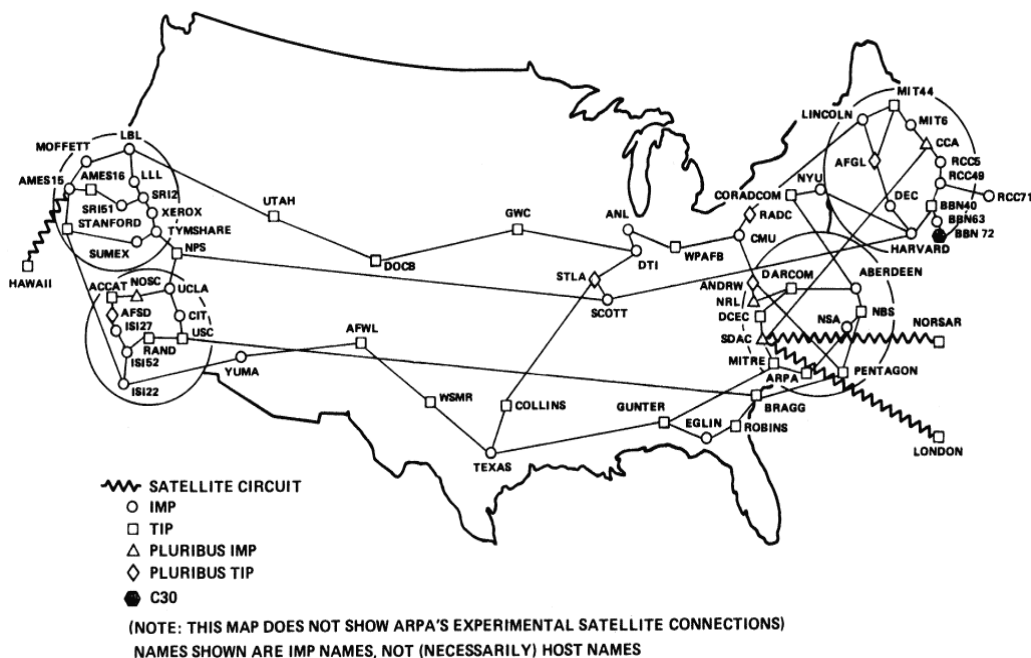


Figure 5: ARPANET (1980)

# History of the Internet

by Richard T. Griffiths

## From Internet to World Wide Web

So far, the net's development was almost entirely 'science-led'. All this time, however, we must remember that parallel advances in computer capacities and speeds (not to mention the introduction of glass-fiber cables into communications networks) were enabling the system to expand. This expansion, in its turn, started to produce supply constraints, which stimulated further advances. By the early 1980s, when the internet proper started operation, it was already beginning to face problems created by its own success. First, there were more computer 'hosts' linked to the net than had originally been envisaged (in 1984 the number of hosts topped 1000 for the first time) and, second, the volume of traffic per host was much larger (mainly because of the phenomenal success of e-mail). Increasingly predictions were voiced that the entire system would eventually grind to a halt.

One early, and essential development, was the introduction in 1984 of Domain Name Servers (DNS). Until then each host computer had been assigned a name, and there was a single integrated list of names and addresses that could easily be consulted. The new system introduced some tiering into US internet addresses such as edu. (educational), com. (commercial), gov. (governmental) in addition to org. (international organization) and a series of country codes. This made the names of host computers easier to remember (eg. our own address www.leidenuniv.nl), but the system is even cleverer because when we type in these addresses, the computer is sending/receiving a coded sequence of numbers as 132.229.XX.XX (which the address of the Leiden University computer).

A second development was the decision by national governments to encourage the use of the internet throughout the higher educational system, regardless of discipline. In 1984 the British government announced the construction of JANET (Joint Academic Network) to serve British universities, but more important was the decision, the following year, of the US National Science Foundation to establish NSFNet for the same purpose (one explicit requirement for receiving funding was that access had to be for "all qualified users on campus"). The American program involved a number of decisions that were crucial for the further development of the Internet:

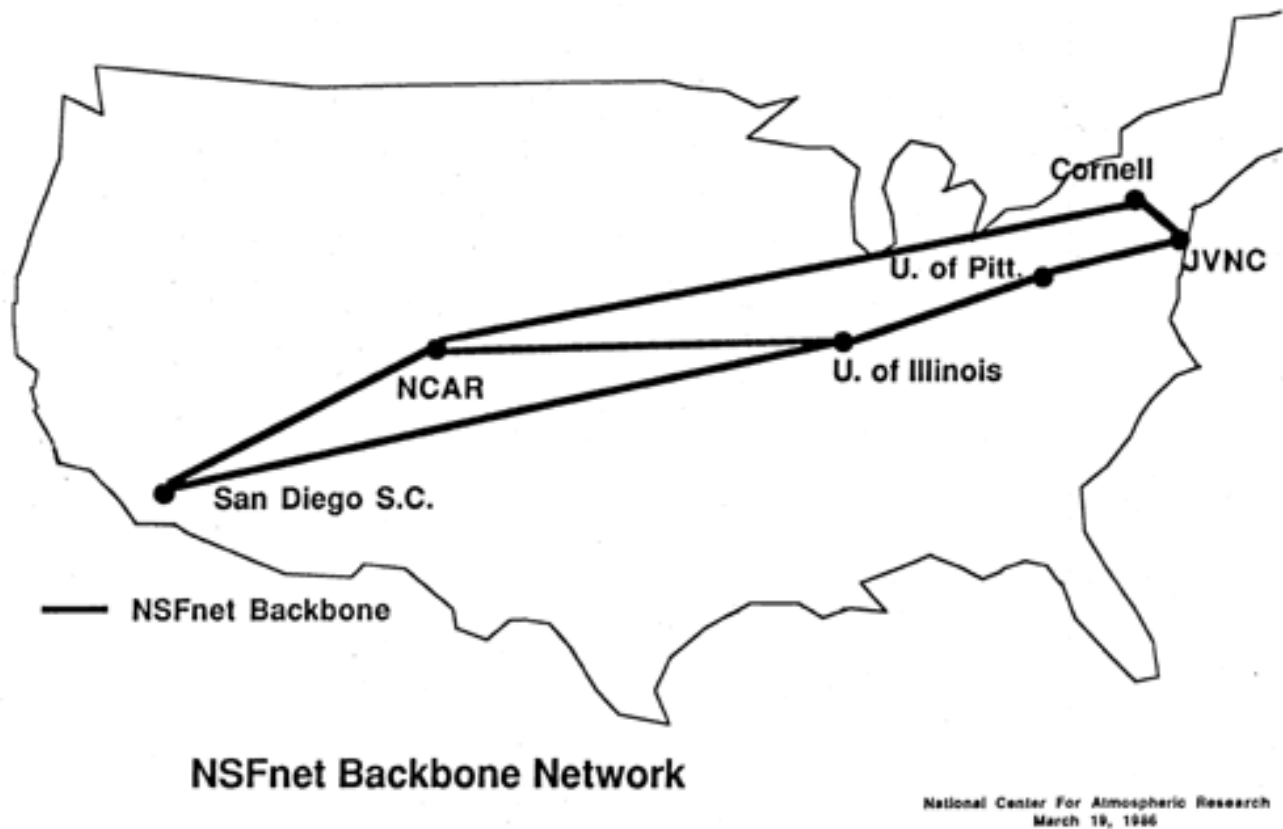
The use of TCP/IP protocols was mandatory for all participants in the program. Federal Agencies would share the cost of establishing common infrastructures (as trans-oceanic connections) and support the gateways. NSFNet signed shared infrastructure 'no-metered-cost' agreements with other scientific networks (including ARPANET), which formed the model for all subsequent agreements.

It threw its support behind the 'Internet Activities Board' (the direct descendent of the Internetworking Working Group established back in 1972) and encouraged international cooperation in further research.

Finally, NSFNet agreed to provide the 'backbone' for the US Internet service, and provided five 'supercomputers' to service the envisaged traffic. The first computers provided a network capacity of 56,000 bytes per second but the capacity was upgraded in 1988 to 1,544,000,000 bytes per second. There was one proviso.... this facility excluded "purposes not in support of research and education".

# History of the Internet

by Richard T. Griffiths



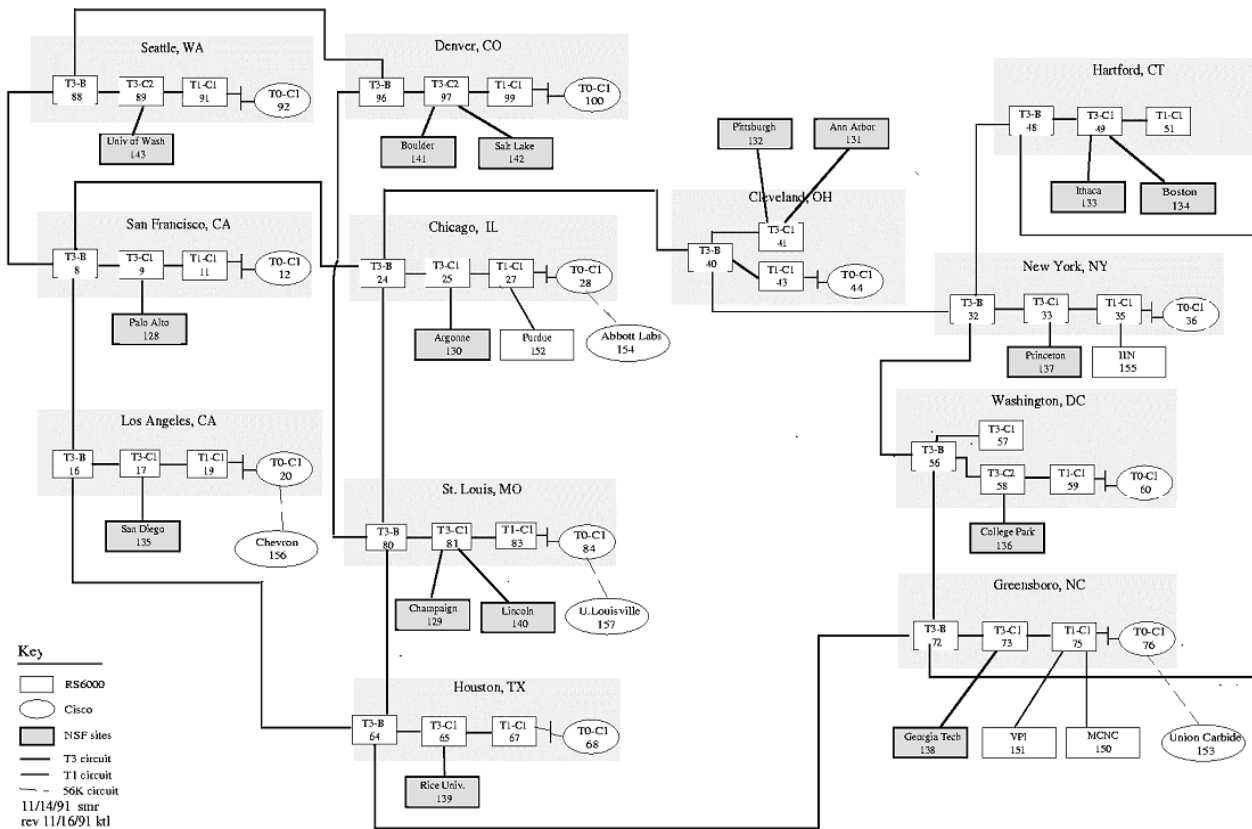
**Figure 6: NSF Backbone (March 1986)**

The effect of the creation of NSFNet was dramatic. In the first place it broke the capacity bottleneck in the system. Secondly, it encouraged a surge in Internet use. It had taken a decade for the number of computer hosts attached to 'the Net' to top the thousand mark. By 1986 the number of hosts had reached 5000 and a year later the figure had climbed to hosts 28,000. Thirdly, the exclusion of commercial users from the back-bone had had the (intended) consequence of encouraging the development of private Internet providers.

# History of the Internet

by Richard T. Griffiths

ANSNET/NSFNET T3 Topology as of 11/18/91



**Figure 7 NSFnet (1991)**

The exclusion of commercial users from the backbone did not mean that their interests had been neglected. For several years, hardware and software suppliers had been adding TCP/IP to their product packages, but they had little experience in how the products were supposed to work and therefore experienced difficulties in adapting it to their own needs. Part of the force behind the Internet's early development had been the open availability of information (since 1969 most of the key research memoranda, and the discussions they had generated, had been archived in downloadable on-line-files) but now the Internet Activities Board went a step further. In 1985 it organized the first workshop, specifically targeting the private sector, to discuss the potentials (and current limitations) of TCP/IP protocols... beginning a dialogue between government/academic scientists and the private sector, and among private entrepreneurs themselves (who, from the beginning were thus able to ensure the interoperability of their products). In 1987 the first subscription based commercial internet company, UUNET was founded. Others follow. At this stage, the Internet is still quite a forbidding place for the uninitiated. Access commands to find data range from the complicated to the impenetrable, the documentation available is mostly (highly) scientific and the presentation unattractive (courier script, no color), finding stuff is a pain in the neck and transfer times are relatively slow). The main attractions for the commercial sector are the e-mail facilities and access to e-mail, newsgroups, 'chat' facilities and computer games.

Although commercial exploitation of the net had started, the expansion of the Internet continued to be driven by the government and academic communities. It was also becoming ever more international.

# History of the Internet

by Richard T. Griffiths

By 1989 the number of hosts surpassed 100,000 for the first time and had climbed to 300,000 a year later. The end of the 1980s and the start of the 1990s provide a convenient cut-off point for several reasons:

In 1990 ARPANET (which had been stripped of its military research functions in 1983) became a victim of its own success. The network had been reduced to a pale shadow of its former self and was wound up.

In 1990, the first Internet search-engine for finding and retrieving computer files, Archie, was developed at McGill University, Montreal. The development of search-engines will be dealt with in the last lecture.

In 1991, the NSF removed its restriction on private access to its backbone computers

"Information superhighway" project came into being. This was the name given to popularize Al Gore's High Performance Computing Act which provided funds for further research into computing and improving the infrastructure of the Internet's (US) structure. Its largest provisions from 1992-96 were \$1,500 mln for the NSF, \$600 mln for NASA and \$660 for the Department of Energy.

And in 1991 the World Wide Web was released to the public and, on a personal note, Richard T. Griffiths (famous for his phrase 'a user friendly interface is a secretary') got kicked into Word Perfect and was launched into cyber-space.

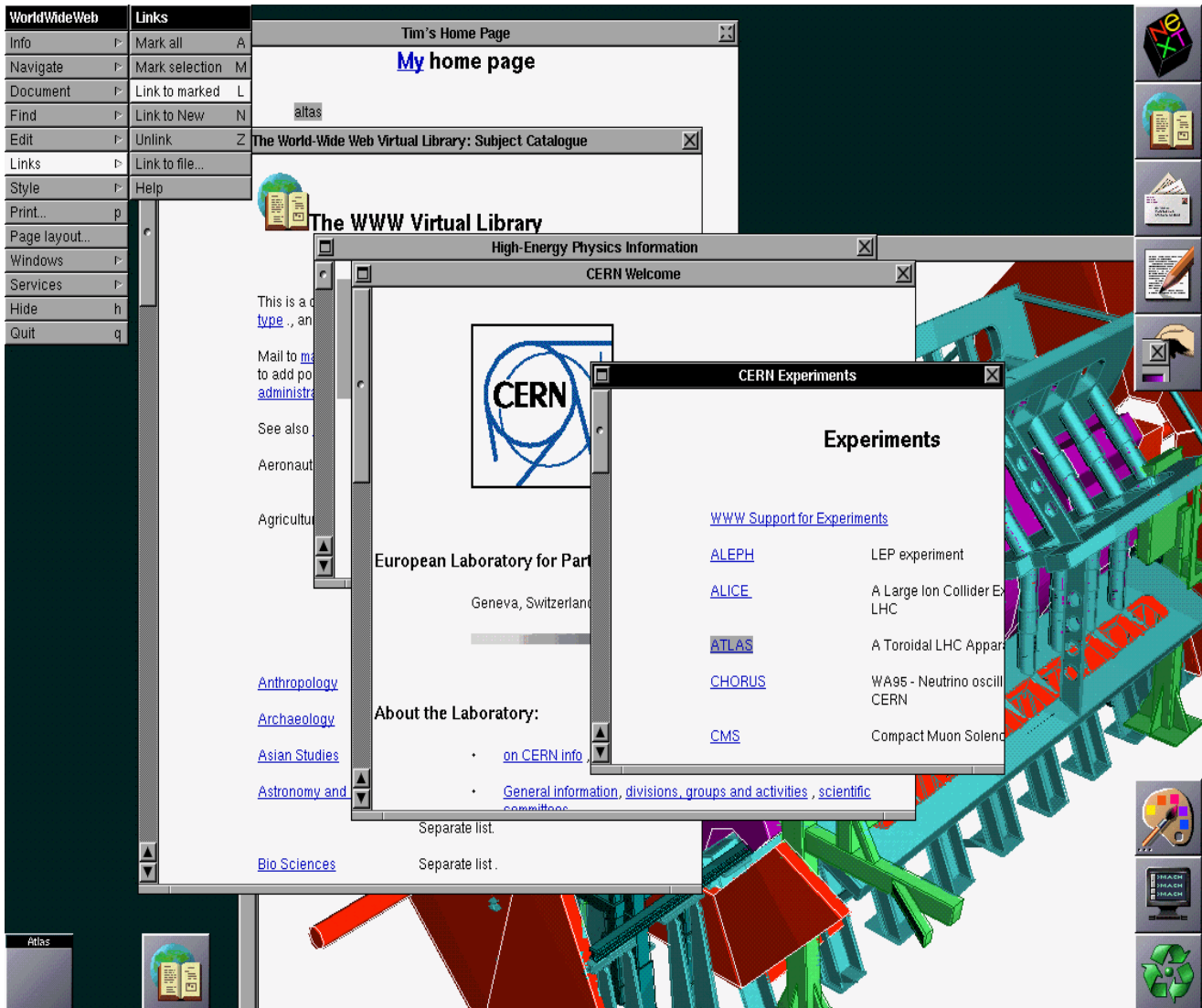
## **The World Wide Web (WWW)**

The World Wide Web is a network of sites that can be searched and retrieved by a special protocol known as a Hypertext Transfer protocol (HTTP). The protocol simplified the writing of addresses and automatically searched the internet for the address indicated and automatically called up the document for viewing.



# History of the Internet

by Richard T. Griffiths



**Figure 9: An early WWW screen (1994)**

The image above gives an idea of the 'state-of-the-art' in browser technology by 1994 (the earlier version would not have supported color, and the logos and diagrams would have been in separate windows as well). Until that occurred the transition to the new system was slow. By the end of 1992 there were only 50 web-sites in the World and a year later the number was still no more than 150.

What is the difference between the Internet as it has then existed and the Web? Tim Berners-Lee was often asked the same question and gave the following answer:

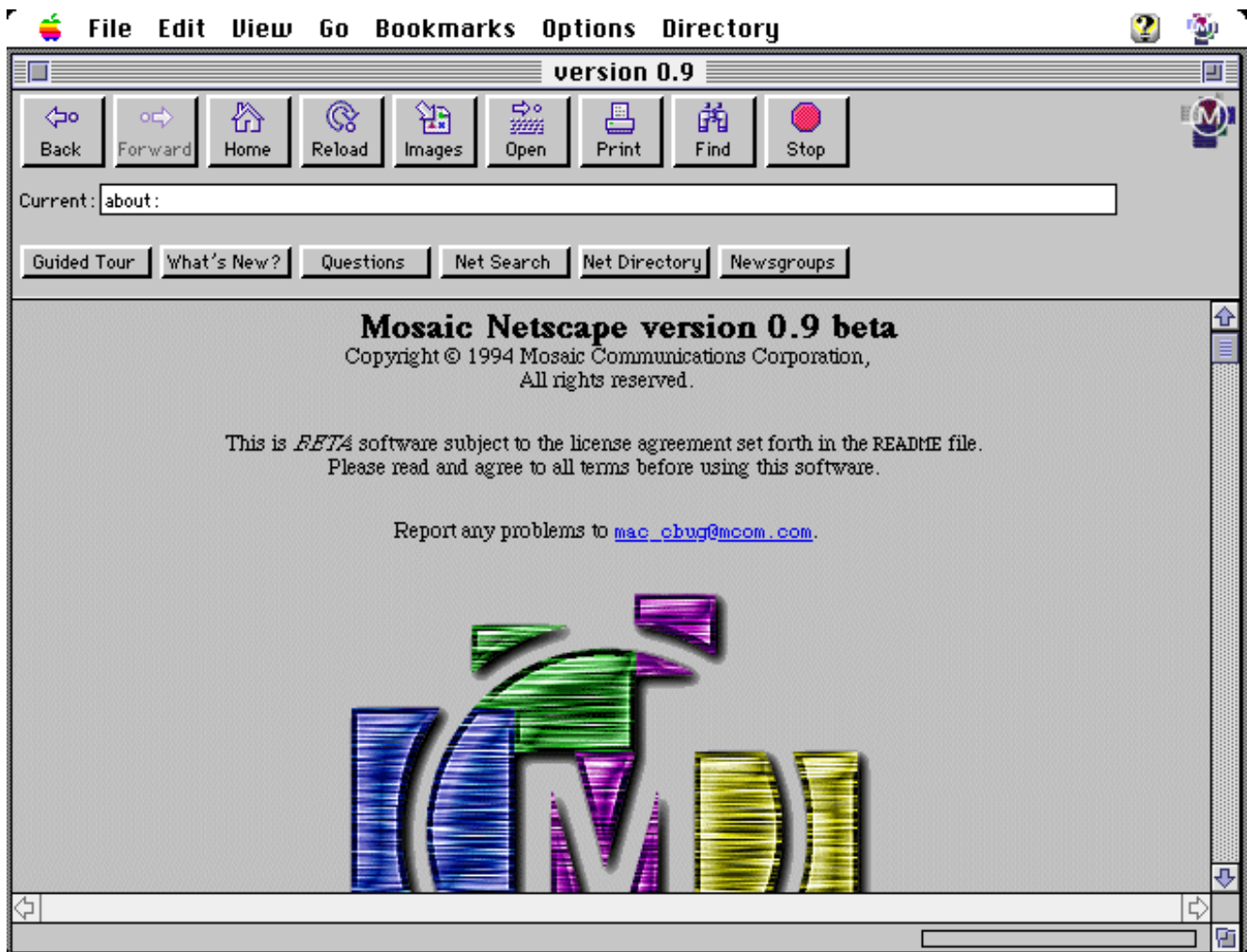
"The Internet ('Net) is a network of networks. Basically it is made from computers and cables. What Vint Cerf and Bob Khan did was to figure out how this could be used to send around little "packets" of information. As Vint points out, a packet is a bit like a postcard with a simple address on it. If you put the right address on a packet, and gave it to any computer which is connected as part of the Net, each computer would figure out which cable to send it down next so that it would get to its destination. That's what the Internet does. It delivers packets - anywhere in the world, normally well under a second.

# History of the Internet

by Richard T. Griffiths

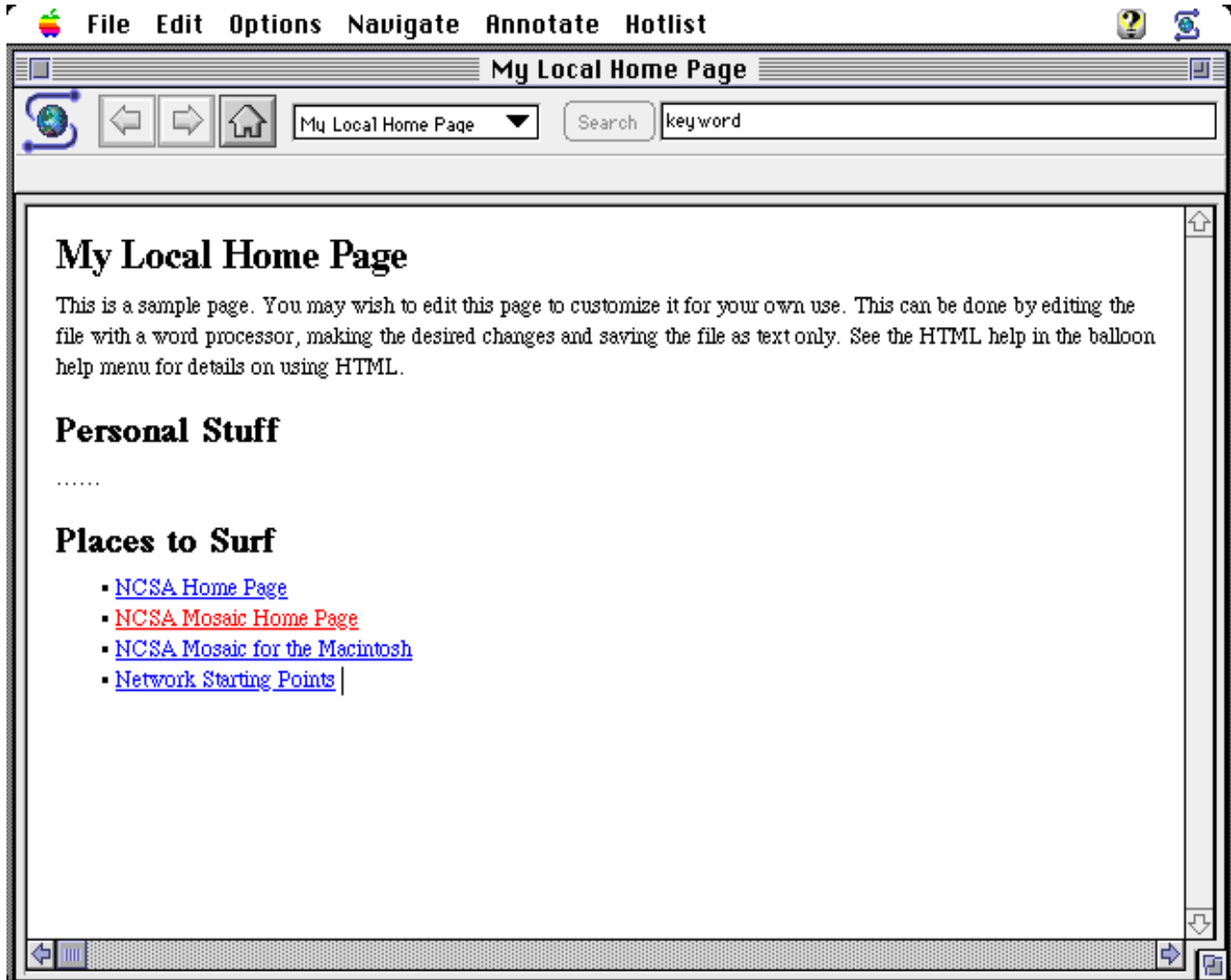
Lots of different sort of programs use the Internet: electronic mail, for example, was around long before the global hypertext system I invented and called the World Wide Web ("Web"). Now, videoconferencing and streamed audio channels are among other things which, like the Web, encode information in different ways and use different languages between computers ("protocols") to do provide a service.

The Web is an abstract (imaginary) space of information. On the Net, you find computers -- on the Web, you find document, sounds, videos,.... information. On the Net, the connections are cables between computers; on the Web, connections are hypertext links. The Web exists because of programs which communicate between computers on the Net. The Web could not be without the Net. The Web made the net useful because people are really interested in information (not to mention knowledge and wisdom!) and don't really want to have know about computers and cables."



# History of the Internet

by Richard T. Griffiths



**Figure 10: An early opening screen for Mosaic and Netscape**

In 1993 Mark Andreessen of NCSA (National Center for SuperComputing Applications, Illinois) launched Mosaic X. It was easy to install, easy to use and, significantly, backed by 24-hour customer support. It also enormously improved the graphic capabilities (by using 'in-line imaging' instead of separate boxes) and installed many of the features that are familiar to you through the browsers which are using to view these pages such as Netscape (which is the successor company established by Andreessen to exploit Mosaic) and Bill Gates' Internet Explorer. Like so many other Internet innovations, trial versions of Mosaic were made available free to the educational community. Mosaic soon became a runaway hit. By 1994 tens of thousands of versions had been installed on computers throughout the World. The potential of HTML to create graphically attractive web-sites and the ease with which these sites could be accessed through the new generations of web-browsers opened the Web to whole new groups. Until now, the Web had served two main communities - the scientific community (accessing on-line documentation) and a wider 'netizens' (net citizens) community (accessing e-mail and news-group facilities). Now commercial web-sites began their proliferation, followed at a short distance by local school/club/family sites. These developments were accelerated by the appearance of ever-more powerful (and cheap) personal computers (which increased both the number of netizens and the potential market for businesses) and by the increase in capacity of the communications infrastructure. The Web now exploded.

# History of the Internet

by Richard T. Griffiths

In 1994 there were 3,2 mln hosts and 3,000 web-sites. Twelve months later the number of hosts had doubled and the number of web-sites had climbed to 25,000. By the end of the next year the number of host computers had doubled again, and the number of web-sites had increased by more than ten-fold. In that year, by the way, the History Department of Leiden University established its own web presence, placing its site among the first 0.00005 per cent of web-sites ever constructed. The following year we started the course 'Internet for Historians' and, within the sections Economic and Social History, we began the development of 'course-based' web-sites. This all took place in 1997, by which time the number of host computers integrated into the Web had reached 19,5 mln hosts, and the number of web-sites had shot up to 1,2 million. By the last count, in January 2001, the number of hosts stood at 110 million and the number of web-sites had reached 30 million.

The history Department website of Leiden University has had several reincarnations since we started in 1995... the last one in the year 2000. You can see it here below. My colleague Marjana Rhebergen has converted the page into how it would have appeared in the original CERN in-line browser and an early Mosaic browser. It speaks volumes about how far we have come in a decade.... And why the WWW has become increasingly popular.



Figure 11: History Homepage (2000)

# History of the Internet

by Richard T. Griffiths



Figure 12: History Homepage in Mosaic (1994)

## History of e-mail

### ARPANET and the Invention of E-mail

ARPANET developers had always considered information and resource sharing as one of their primary goals. Indeed the construction of ARPANET had been undertaken to that very end (see Chapter One) but it would be fair to say that their horizons were limited to the exchange of scholarly, technical papers and programs via ftp. For example, direct human-human contact had little place in Licklider's original concept of a 'galactic network'... probably because it appeared too trivial an application in a world where machines still occupied whole rooms rather than the tops of desks. Yet, like the internet itself, it was invention waiting to happen. Almost as soon as terminals in different rooms could be linked to the same 'host' computer through 'time-share' operating systems, it became possible to leave messages for one another within the same system. Such applications began to appear from 1961 onwards and immediately proved popular among users. However their limitation was that their use was restricted to the users of a single computer. So why not between different computers in different locations? So, once ARPANET came into being (linking four centers in 1969, 15 by 1971) scientists began to consider sending direct messages over the same medium.

# History of the Internet

by Richard T. Griffiths

In 1971 Ray Tomlinson of ARPANET sent the World's first e-mail, by adapting an existing, popular, time-share internal mail program and linking it to the new network file transfer technology that underpinned ARPANET's further activities. The first message was simply addressed to himself, sent from one computer to another, with the text 'Testing 1-2-3'. The next thing he did was to address a message to all ARPANET users explaining the availability of 'electronic mail' and giving instructions on how to address mail to another user using the convention - users' log-in name @ host computer name - which is still the basis of e-mail today.

That was a start, but it was still unbelievably crude. The only way to know what was in a message was to open it in its entirety. The messages had to be read in the order they were received. They read as pieces of continuous text (rather like a teletext message today). The reading process and sending process took place via two completely separate programs. What was needed was a system to make the entire operation more user-friendly... a piece-of-cake given the computer expertise concentrated around ARPANET. Innovations followed thick and fast. More or less in chronological order there was developed:

- a listing of messages, indexed by subject and date
- the ability to selectively delete messages
- the ability to receive and send mail from the same program
- the ability to forward messages, to automatically include (sender) address with the message
- the ability to file and save messages, and
- a standard protocol to allow the exchange of messages between programs.

As a result of these developments, by the summer of 1972 (within twelve months of the first message) most of the facilities that we recognize in current e-mail programs were already in place.

It did not take long for e-mail to establish its own particular style. It was more like a post-card than a letter, or perhaps more accurately, more like an office memorandum (with its headings "To:", "From:", "Subject:" and "cc:". Anyway, whatever the analogy chosen, e-mail offered:

- a high level of informality (possible aided by the fact that all the original users were already in a 'community', albeit a scientific one)
- a tolerance of spelling mistakes and typo's (that would be totally unacceptable in a formal letter)
- a terseness of expression and relative brevity of message (because, all said and done, its still no joy to read screen after screen of message... and because the ability to give each message a 'subject' description, which could be read and replied to separately, made it logic to send several short messages instead of one long one).

The style of e-mail communication was one reason behind the success of the new medium. We are in the 1970s, when the stuffiness and hierarchy associated with society in the 1950s had already been swept away. The 'bluntness' of the medium was no longer seen as threatening but, instead conveyed a feeling of intimacy and immediacy. But there were other advantages. For example, as long as the cost of keeping computer links open were carried by the computing centre, it was a much cheaper medium than the telephone. Moreover, unlike a telephone, one could always keep a copy of the communication. And, finally, although communication was virtually simultaneous, the recipient did not actually have to be present to receive the message.

# History of the Internet

by Richard T. Griffiths

If you are not familiar with e-mail, the links below give a guide to some of the more common programs in use. Browsers & Mail

E-mail was an instant success.... although, for the first five years, its development was largely ignored, and left unrecognized, by the ARPANET administration. By that time various discussion groups devoted to specific topics (whereby you put yourself on a mailing-list to receive all posts on a particular topic) had sprung into existence... and electronic mail was set to overtake file transfer in the volume of traffic over ARPANET. By the end of the 70s there were 17 groups in existence, by 1982 there were 44. The most popular among the subjects of these early lists were science fiction and Human-nets. The latter, which appeared around 1975, was devoted to the social implications of the e-mail medium itself, and it helps define the moment when the e-mail users began to realize the full implications of the communication tool they were using. But ironically, while ARPANET scientists were beginning to philosophize about the brave new world they were about to enter, they were the only ones with access to it... a group of elite defense and communication scientists in institutes whose membership of the net was dependent their role as ARPA (sub-)contractors. Before it could reach the broader public, e-mail itself had to be re-invented.

## Usenet and the Reinvention of E-mail

As mentioned at the start of this section, time-sharing (and message exchange) on a single computer had been developing since the early 1960s - the bottleneck lay in repeating the trick between different computers. In 1964 collaborative project was called Multics [Multiplexed Information and Computing Service] started by General Electrics, MIT, and AT&T (<http://www.att.com/technology>), to construct a time-share program to be implemented on the GE 645 computer. In 1969 AT&T Bell Laboratories withdrew from the project and began work on their own. The operating system they developed was known as The Unix Programming Environment. From the start the system was made an "open" one, in the sense that all Bell employees had free access to it, and so too did universities. This made for several advantages - it guaranteed the spread of the system, it provided designers with more feedback on glitches in the system and possible applications, and it allowed university scientists (and their students) to design their own special applications. In a way, universities were forced to gain hand-on experience with the system since while Bell Labs provided the tapes and the manuals for the system, they offered no back-up or support services. The system was widely adopted throughout universities, popular because it was free and because it could be run on computers far smaller than the room-huggers that dominated the ARPA network. In 1981, the spread of Unix run software received a further flip when it was installed as the basic program environment on IBM's mainframe computers (starting with the 3033AP).

Given Unix's 'open' nature of the system it is not surprising that there also developed an equally wide community of Unix-users with expertise in running the system and developing new applications for it. In 1986, computers using the system were able to link-up to each other through 'dial-in' facilities, but the operation was still primitive and painfully slow. The system was never integrated into the 'Internet' since it was incompatible with the file transfer programs eventually adopted. Nevertheless, it did allow mutual accessing of files and programs and it offered facilities for 'computer conferencing'... and computer chess.

Two post-graduate students at Duke University, Tom Truscott and Jim Ellis, were responsible for the concept and early development of Usenet. They had met through a joint interest in computer chess and had designed their own program. They were regular participants at computer chess competitions, which utilized the 'conferencing' facility of the Unix system (with the competition organizers footing the huge telephone charges). It was through these competitions that they had met people from Bell Labs and had worked there in the summer of 1979. It was also in the summer of 1979 that they attended their first Unix-users conference. Returning to Duke they conceived the idea of creating a network for

# History of the Internet

by Richard T. Griffiths

Unix users and started to design the programming to do it. Their ambition was to make a little contribution to Unix software and "hopefully achieve some minor level of fame." In 1979 they had succeeded and sharing the software with friends at the University of North Carolina, they established the first 'Usenet'. Perhaps not surprisingly, one of the first groups established was NET.chess. In January 1980 Truscott and Ellis presented their system at a meeting of the Academic Unix Users Group and handed out a five page 'Invitation to a General Access UNIX Network'. It explained:

"The initially most significant service will be to provide a rapid access newsletter. Any node can submit an article, which will in due course propagate to all nodes. A 'news' program has been designed which can perform this service. The first articles will probably concern bug fixes, trouble reports, and general cries for help. Certain categories of news, such as 'have/want' articles, may become sufficiently popular as to warrant separate newsgroups. (The news program mentioned above supports newsgroups.)"

The Usenet concept differed from that used by Arpanet. Whereas to join an Arpanet group, you subscribed and were sent the e-mails, in Usenet the messages were (temporarily) stored in the computer and could be accessed at will by the reader - the difference between having a postcard land on your doormat or reading on a pin-board in a secretarial office.

One couldn't really say that after the 'Invitation' things took off with a bang. By mid-summer 1980 usenet like this:

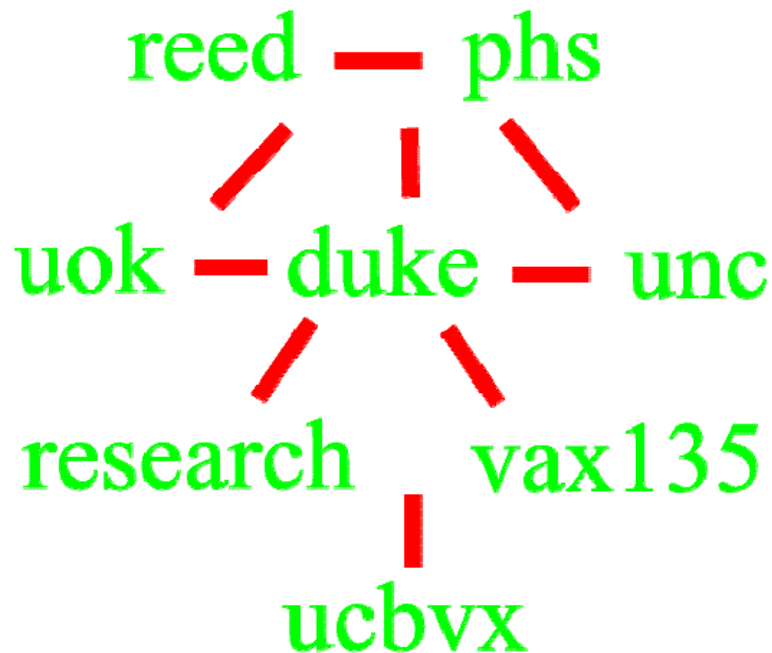


Figure One: Usenet Map, Summer 1980



# History of the Internet

by Richard T. Griffiths

By the end of 1981, 150 computers were linked to the network and within a year the number had swollen to 400. Whereas initially Truscott and Ellis had anticipated two or three messages being transmitted daily, the number had increased to fifty a day by the end of 1982. By that time there were more than fifty newsgroups in existence. The contrast in content with Arpanet was noticeable. Whereas 67 per cent of Arpanet's groups were devoted to computer/technical topics, this accounted for only 25 per cent of Usenet's groups. 21 per cent of Arpanet's groups dealt with 'serious' discussion topics (most of which were fed through to Usenet as well) leaving 10 per cent as 'fun' groups (including chess and science fiction). By contrast, fully half of the Usenet groups could be described as 'recreational' and some (net.jokes) downright frivolous.

Usenet now began an explosive growth, as shown in the following data:

Year	Number of Sites	Articles/Day	Number Newsgroups
1979	3	2	-
1980	15	10	-
1981	150	20	-
1982	400	50	50
1983	600	120	-
1984	900	225	-
1985	1300	375	-
1986	2500	500	240

Usenet operated as a series of (big) linked computers whose "wizards" (owners is wrong word since these were still universities or research institutes) agreed to pass on all 'post' sent to them to the next recipient (immediately) and so on through the 'network'. Obviously, they also agreed to make them accessible to its users. The Usenet network was subsidised (often unknowingly) by these larger machine owners, who picked up the phone-bill. Around 1983, within the network, there developed a 'backbone' of reliable, larger computers through which most of the traffic flowed (if only because it was likely to arrive earlier than through other routes). The wizards of the backbone sites - a group of people using the lists and developing software - were in regular contact and became known as the "Usenet Cabal". It was the cabal that decided what would be carried, and how it would be catalogued.... ie. what groups should be created. BUT for the first years, with a limited community, there was little need for censorship.

## Newsgroups and 'The Great Renaming'

In the history of newsgroups, the great renaming which took place in 1986/87 is the next main step. By now there were 241 news groups divided into three categories:

net.... unmoderated groups  
mod.... moderated groups  
fa..... ARPAnet groups

There were two problems.

# History of the Internet

by Richard T. Griffiths

First, there was a lot of unserious stuff around. There was also a need for some very serious localized groups (one woman in New Jersey posted an advert for a second-hand Dinette set which traveled all the way to Australia). Within the groups, the cabal had already in August 1985, banished the carrying to 'chain letters' (a ban that remained until 1993). In March 1986 the first 'netiquette' had appeared. Among the sins it condemned were the practices of SHOUTING (ie using capitals) 'flaming' (sending unnecessarily immoderate messages to a newsgroup) which could lead to 'flamewars' and 'mailbombing' (sending repeated messages to a group or person with the intention of making life unworkable). All of this increased the traffic carried. Moreover, there were also 'fluff' groups (net.jokes, net.rumour, net.bizarre, net.flame, net.wobegon etc)

Second, it cost increasing amounts of money to route these all through the Net. This applied especially to the overseas links from North America which were established in 1982/83. The link to the UK was carried by through Centre for Seismic Studies, North Virginia (vax135); that with mainland Europe through a Philips Digital Equipment Corporation (decvax) in the USA to the Mathematics Centre (mcvax) at the University of Amsterdam (at \$6 per minute) and that to Australia also from DEC to University of Melbourne (munnari).

The crux of the matter was one of cost. As one contributor expressed it already in 1982:

"I am beginning to wonder about USENET. I thought it was supposed to represent electronic mail and bulletins among a group of professionals with a common interest, thus representing fast communications about important technical topics. Instead it appears to be mutating into electronic graffiti. If the system did not cost anything, that would be fine, but for us here at Tektronix, at least, it is costing us better than \$200 a month for 300-baud long distance to copy lists of people's favorite movies, and recipes for goulash, and arguments about metaphysics and so on. Is this really appropriate to this type of system?"

Thus, the decision was taken to rationalize the information flows (although this did not always pass without resentment). New categories added between July 1986 and March 1987 were:

```
comp... computer oriented
misc... miscellaneous
rec.... recreation
sci.... science
soc.... social (this carried a lot of boy-meets-girl stuff)
talk... gossip
```

This enabled some universities to drop the 'talk' and 'social' groups and for several years these groups and 'alt' (see below) were not transmitted to Europe.

The renaming exercise had seen the gods mingle with the mortals. Accusations of censorship arose... and not only among the netusers. Some of the wizards also thought things had gotten a little out of hand and sought to reconstitute a 'Usenet' outside the 'backbone' for this they created the stem:

```
alt.... alternative
```

The word was intended to represent 'outside the system'. The first group was alt.gourmand, introduced because the creator didn't like mod.gourmand or rec.recipes. Alt.sex and alt.rock-n-roll soon followed.

# History of the Internet

by Richard T. Griffiths

The Great Renaming represented the height of the Cabal's power, and also marked the beginning of its decline. The cabal was forced by the 'great renaming' controversies to democratize their naming procedures, which loosened its control. Moreover, the availability of new routes (particularly NSFNET - named after US National Science Foundation - backed up by five super-computers) swiftly removed the bottleneck in newsgroup traffic. And finally, in 1987, UUnet was set up - a non-profit-making news carrier, which had every incentive to carry and distribute the alt. group and others.

## The Democratization of the Net

The Great renaming established the format of Usenet newsgroups that we know today. It was initially defined by the programming needed to send and receive it, but by the early 1990s this 'functional' definition ceased to be applicable and much of the news was transferred over the Web. It was carried to an ever wider public - first as internet access spread over wider sectors of the academic community and later as the Web opened up. By 1990 the number of newsgroups exceeded 1000 for the first time. It was also the year in which alt.sex became the most widely read group on the net. Four years later, the usenet exceeded the 10,000 mark. There was now a further fillip to the development. As the WWW became more user friendly, so commercial providers began to enter the scene in a big way. By 1993, the traffic on usenet spanned the globe (though heavily concentrated in the US and Europe).



**Figure Three: Usenet Traffic Flows 1993**

In 1995 Internet access providers as CompuServe, Aol and Prodigy all started business. The result was an explosion of news groups, not only into the absolute minutiae of utter trivia (always someone else's hobby) and, recently, by the proliferation of foreign language groups. Whilst the Web is supposed always to have been international, the fact is that for much of its history the lingua franca has been English. This now changed and virtually every language group has its own collection of newsgroups:

# History of the Internet

by Richard T. Griffiths

fr France  
it Italy  
nl Netherlands  
no Norway  
pl Poland

What a great aid to language learning! By now, the number of groups must be close to 100,000. And, on a personal note, on January 22nd 1998, much to my delight, some enterprising fan created the newsgroup it.sport.calcio.fiorentina.

E-mail and Usenet have sparked immense changes in the social space of those who have experienced it...and will continue to do so. Unfortunately it has little to offer the professional historian. There are some newsgroups, but the discussion is not always at a high level (well, they are open access groups!) and, if you are thinking of using any material you might locate, don't forget that any message is only posted for a limited duration. The following groups, however, might be worth a visit:

alt.culture.african.american-history  
alt.history.ancient-worlds  
alt.history.colonial  
alt.history.what-if  
sci.space.history  
soc.history.ancient  
soc.history.medieval  
soc.history.moderated  
soc.history.science  
soc.history.war.misc  
soc.history.war.us-civil-war  
soc.history.war.vietnam  
soc.history.what-if  
soc.history.world-war-ii

Most of these are 'fun' sites. For the serious stuff (perhaps too serious) you have to go to the relevant lists.

## Lists

As mentioned earlier, a newsgroup places all the news on a server and the reader selects what he or she wants to read. In a list, specific items of news are sent to all readers who have indicated previously that they want to receive it. From the moment that e-mail facilities allowed users to store addresses and to forward messages to everyone on a specific address list, it was possible to construct lists. Arpanet's 'newsgroups' were really lists and many other networks that emerged in the 1980s also had people that ran lists devoted to specific subjects. The drawbacks lay in the fact that all these operations had to be performed manually, making the success of a list contingent upon the diligence of the list's moderator. Added to the human factor, lists were prone to all the normal frustrations associated with slow/failed connections, especially on the oversaturated transatlantic cables. Improved infrastructure would remove the second of these problems, human ingenuity the first. In 1985 Eric Thomas (an American computer student working in Paris) designed the first automatic readdressing system but it was slow, primitive and allowed no interaction. Moreover, the initial contacts were all through human agencies. Within a year, however, he had added a user-interface system (including passwords and security checks) and could process all the requests that the previous machines were performing manually. Moreover, the interfaces were duplicated (so, with a machine in Europe, messages only had to cross the Atlantic once, to an American mirror). The

# History of the Internet

by Richard T. Griffiths

system became operational in June 1986 and by the end of the year it was serving 41 lists. Eighteen months later there were 1000 lists on the system.

The system was continuously modified as demand increased. One major blip worth mentioning as the 'IBM crisis' around early 1990 as universities and other research organizations began getting rid of their (large, unwieldy and increasingly out-of-date mainframes) and thus cutting off their list managers in the process. THUS decided to go commercial and develop new software that was not dependent on the IBM operating system. In June 1994 this was ready.... new version made available in May 1995 linked software to Windows environment. The result is that Listserv today supports 23,000 public lists and almost 100,000 public lists delivering in a day 22 million messages. There is a useful entrance to most of these through: <http://liszt.com/>

For us, the most interesting development is the decision in 1995 of the National Endowment for the Humanities to fund a listserv program (including free software) for the Humanities in the form of H-Net which now hosts over 100 history discussion groups with a total of 60,000 members (0,1 per cent of all list-serv members)! These are accessible through: H-Net

Note, by the way, that economic history lists are not taken up in alphabetical order but appear at the end of the list! Each list will give you instructions how to subscribe, and many allow you to do so from the H-Net site. You will often be sent an on-line form to fill in and it might take from a day to a week to join. And follow the instructions (including full stops and spacing) meticulously... the message is being read by a computer, not a human!

These lists cover a wide range of historical subjects, but they vary in their usefulness and relevance. What most of them contain are:

- announcements (of jobs, conferences etc.)
- requests for information
- book reviews
- discussion 'chains'

It is the last two that makes them most interesting for history students. The reviews are far more descriptive and detailed than those found in academic journals and since debates are conducted among academics many are held at quite a high intellectual level, although often on esoteric subjects. But you don't have to read them all. Most discussion chains carry the same subject descriptor and so you can always delete the less interesting ones. As in most lists (academic or otherwise), the best advice is to 'lurk' for a while before participating actively. You can also get an indication of the type of discussions by visiting the group's archives. Get a feel for the 'style' of the group and the expected level of response... and don't get upset if some of the replies are not exactly sympathetic. If you decide to leave a group, don't forget to 'unsubscribe', or else your mail-box will soon overflow. And don't forget your netiquette.

## Netiquette

Netiquette is nothing more than etiquette for the net. It and its first appearance in 1986 and attempted to set out the rules of acceptable behavior. There are plenty of netiquette sites to be found via Yahoo but they are all typical to the one linked above... and they all read a bit like advice columns in the pages of the more staid women's magazines. Still, some of the tips are not immediately obvious:

- Don't SHOUT (by using capitals)
- If you \*must\* emphasize, use asterisks (but we don't like preachers!)
- If you want to show emotion, use an 'emoticon' to laugh :-), or frown :-(, or joke ;-)

# History of the Internet

by Richard T. Griffiths

- Don't spew (repeat your opinion ad nauseam, usually to people who don't want to listen - eg. don't cite bible quotes in alt.satanism)
- Don't blather (go on, and on and on.... a screen, at the most two, should really be the limit)
- And be moderate in your opinions, or you might be guilty of FLAMING.

To flame is to express an immoderate opinion and it can easily lead to flame-wars... which degenerate into a slanging-match, much to the disgruntlement of other readers of the group. And they have ways of getting even. You don't want to get bombed (have your private e-mail address deluged by repeated messages until your box is full).

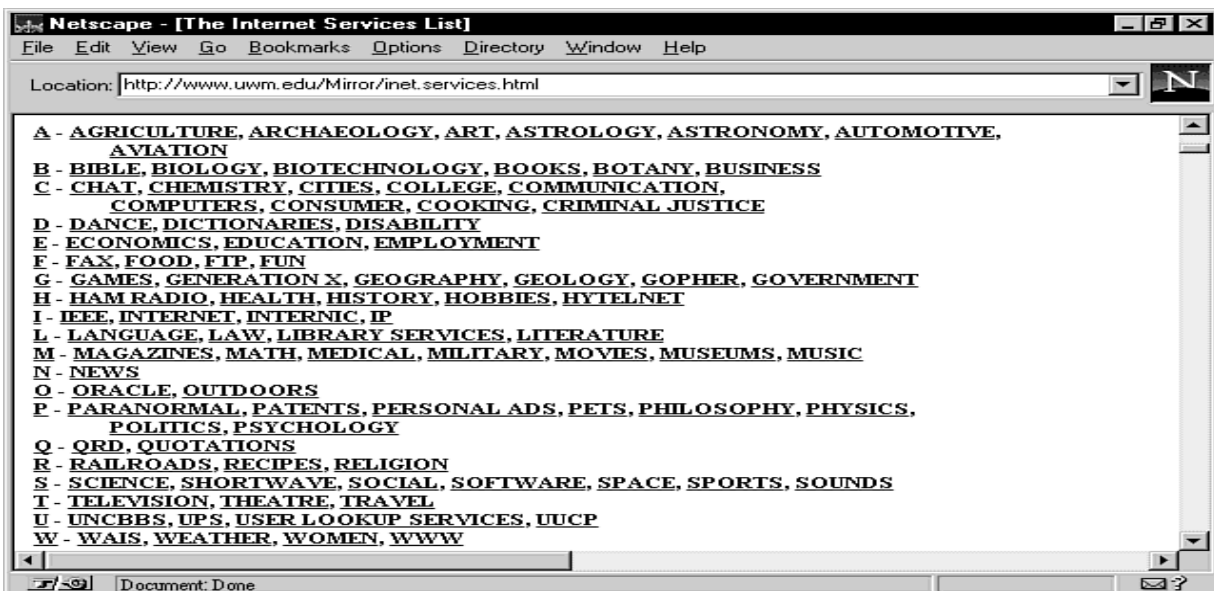
So it might makes sense to read one of those on-line manuals after all.

- Search Engines
- In the beginning
- How do we find information on the net?
- Directories
- Search Engines
- Databases
- Library Catalogues

Much of the information for this paper was taken from the superb Search Engine Watch, which is the best place for further information. Much of the rest came from the companies themselves. All the data was current in September 2001.

"There is nothing worthwhile on the web, and you will never find it anyway".

This complaint has been voiced since the beginning of the web... and usually by people who had never even tried. From the start of the internet, there have been directories and lists on various topics maintained by enthusiasts (who were often experts in their fields). My own favorite used to be a list maintained by Scott Yanoff. He started a small list for his personal use in 1991 and found himself being snowballed by e-mail suggestions from grateful users until he became almost an institution, and author of a couple of internet guides. He seems to have given up in 1995. This is what it looked like then.



# History of the Internet

by Richard T. Griffiths

By the late 1980s, however, the amount of data was getting too large to rely on helpful hints from other users. From its start in 1983, the internet had grown to 1000 hosts in 1984, to 10,000 in 1987, to 100,000 in 1990 and to 1,000,000 in 1992. Information retrieval was becoming a bottleneck and a clustering of innovations took place to resolve the problem.

## In the beginning

1990 Archie, developed at McGill University (Montreal) first search engine for finding and retrieving computer files. At the time these large institutional computers placed their data and program files into two categories: open and closed. When you 'logged-in' to another computer, you could access the 'open' files by identifying yourself as "anonymous" and using your e-mail address as the password. Then you could browse through their archive and download any files you wanted.

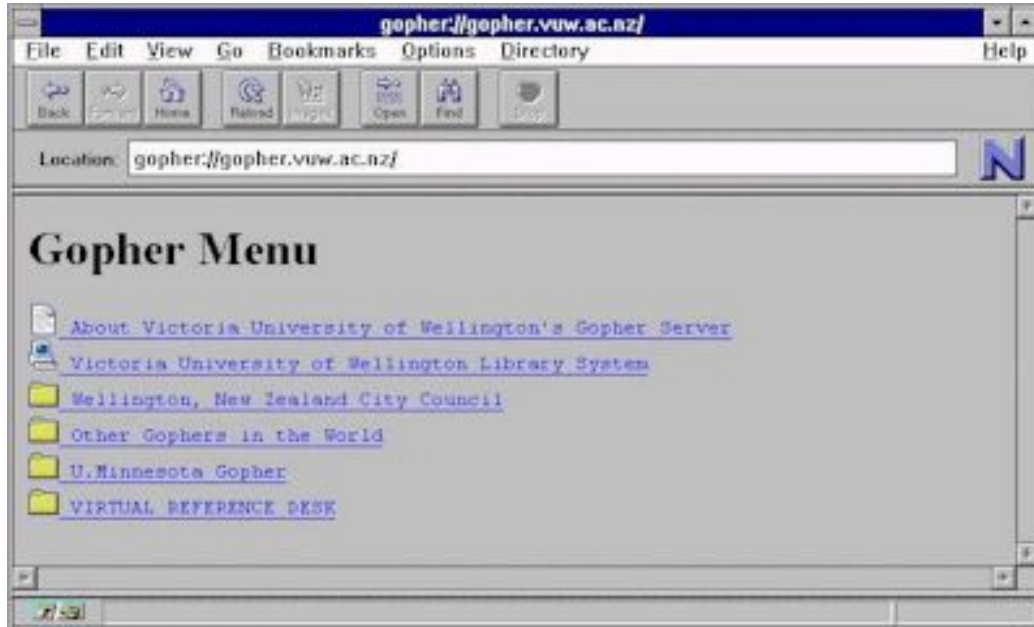
What Archie did was automatically at night (when the traffic was less) to visit all the archives they knew about and to copy the list into a searchable database (this piece of software was known as a spider). When you logged into an Archie site (by telnet) it would tell you where any file was and you could view and e-mail the results to yourself..... and you could go through the entire 'log-in and retrieve' procedures for each computer yourself. It was a striking comment on the state-of-the-art at the time that McGill soon discovered that half of the whole US-Canada traffic was running through its Archie server, so it shut down public access. By then, however, there were many alternative sites hosting the service.



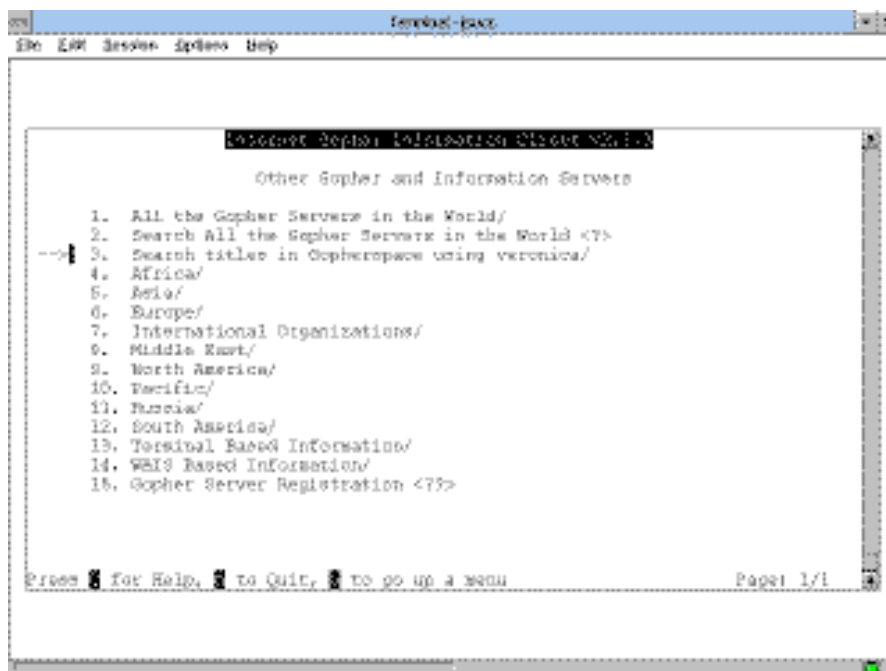
1991 The Gopher system represented an improvement on ftp retrieval developed at the University of Minnesota (whose mascot was a golden gopher).

# History of the Internet

by Richard T. Griffiths



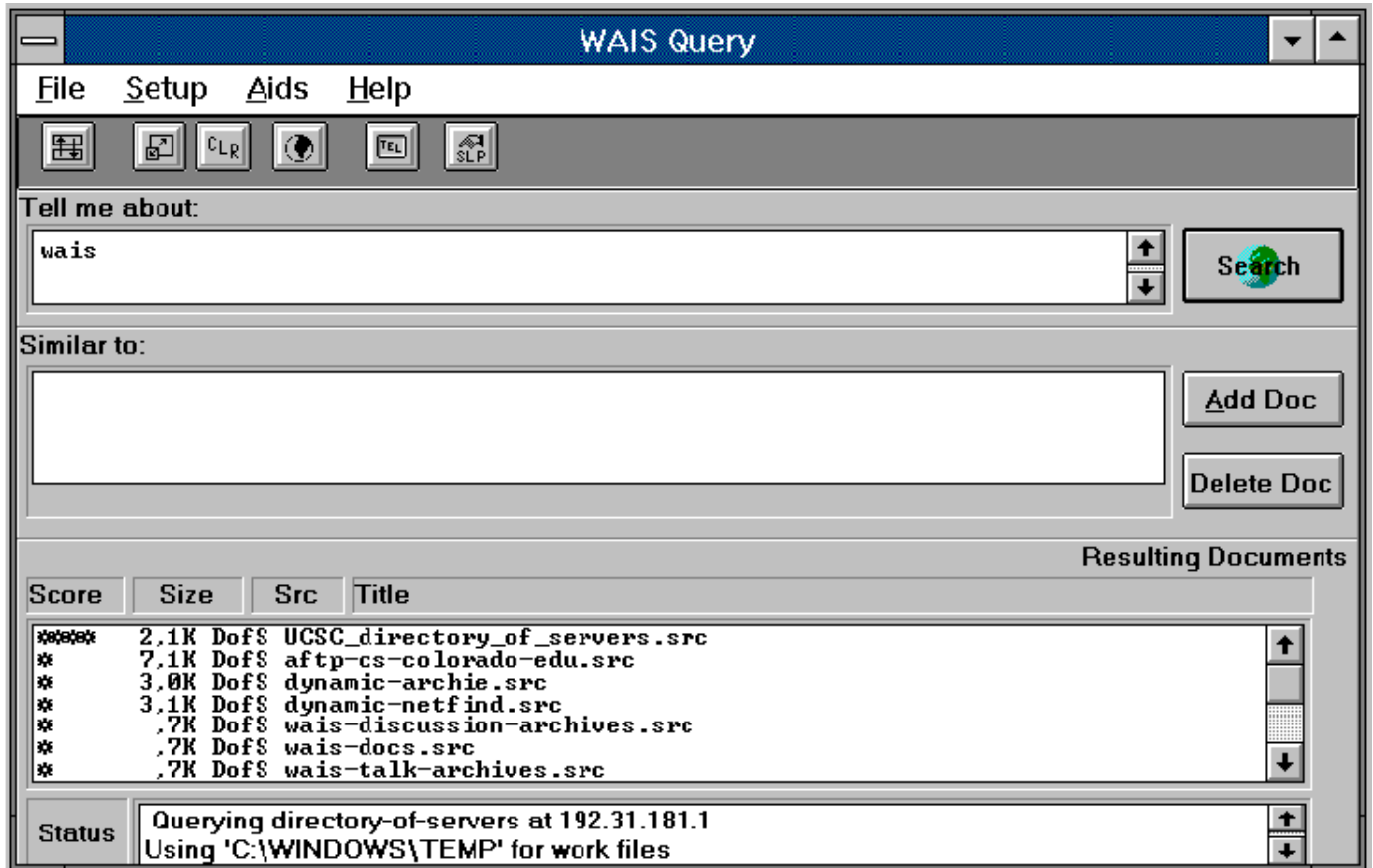
The host computers (servers) put their files in a 'menu' form and the menus of the different servers were merged. Now you logged into any gopher server and you could query it for information by typing in keywords and, again like Archie, you would get a list of items. But now, instead of sending yourself the list and individually looking up the items, you scrolled down the list, pressed 'enter' and you were transferred directly to the relevant 'gopher' address, where you could read the contents. Then, if you wanted, you sent the file to yourself via e-mail. Since 'gopher' was a useful way for storing data, the system caught-on very rapidly. And within 'gopherspace' search-engine called Veronica (supposedly Very Easy Rodent-Oriented Network Index to Computerised Archives), developed at the University of Nevada operated on the same principle as Archie but it also allowed you to distinguish between a search for 'directories' and an undifferentiated search combining directories and files (the latter was much larger and time-consuming). Again, having located something, you e-mailed it to yourself.



# History of the Internet

by Richard T. Griffiths

1991 also saw the birth of WAIS (Wide Area Information Server) developed by Thinking Machines Corp.



Wais was also logged into separately. Wais searched through information on the basis of the contents. So, if using Archie and Veronica was like searching through a card index of book titles, WAIS was like using an book index. WAIS's data base was smaller than the other two but, even so, searching through the lot was daunting and time-consuming. So WAIS broke down its databases into separate subject indices and the researcher could then restrict the word search within the relevant category. At its peak WAIS linked up 600 databases around the World. WAIS ordered the results in rank order of the frequency they appeared and since it was gopher-based, you could click to the document and read its contents (and e-mail it to yourself if so desired).

Thus, these early search-engines had a spider built up databases either of directories or web-pages or built-up directories (specifically limited in ambition and range, but supposedly limited to better sites) and they could also rank by terms within a document.

None of the first generation search-engines mentioned above have survived... but, like dinosaurs, they live on in better adapted versions. These principles they developed, refined and made more powerful, underpin the design of almost all subsequent search-engines.

# History of the Internet

by Richard T. Griffiths

This was situation in the early 1990s. Then in 1991 the WWW was developed and two years later the Mosaic graphics browser. These contributed to an enormous expansion of the net, but they also offered the development of a new generation of user-friendly search-engines. If, in 1992 the number of hosts had reached 1,000,000, by 1996 the number had surpassed 10,000,000. Moreover, the number of web-sites was beginning to increase exponentially. Two years later there were 36 million hosts and 4 million web-sites.

There is more information of the web than ever before, but in many ways it is easier to locate the information we want... if we work systematically and intelligently, and have a little patience.

How do we find information on the net?

## Directories and Search Engines

With the exception of the WWW Virtual Library, which is really a set of clickable categories, all the following directories and search engines have a SEARCH function

If you type in a single word there are no problems, other than the fact that single words, such as 'history' will probably produce an unmanageable number of hits. So try to be more specific.

NOTE many search engines are sensitive to CAPITAL LETTERS. If you use lower case letters, all the search engines below will also match it to capitals as well. If you use capitals, Alta Vista and Infoseek will not bother looking for lower case matches (this could be useful, for example if you are looking for WHO as in World Health Organization). As a rule, except for names, use lower case only (and if something that should be there doesn't show up, try a capital letter).

All the search engines mentioned here recognize the following 'search engine maths'. NOTE do not leave spaces

**use +history+medieval for documents mentioning both**

**use +history+medieval+women for documents mentioning all three**

**use + history+medieval-women for documents mentioning the first two but excluding the third**

**use +history-women for documents including the first but excluding the second**

**use "medieval history" for documents containing the exact phrase, or words in that order**

Some of the search engines (in advanced search mode) use 'Boolean' logic. NOTE not yahoo!, Infoseek and Google, and in Alta Vista only in 'advanced mode'. The principle is the same as above, but it is a little more powerful. For example:

**use history AND medieval for documents mentioning both**

**use history NOT women for documents mentioning the first but excluding the second**

**use history OR geschiedenis for documents mentioning either word**

**use history NEAR medieval to stipulate that the words present need to be close to each other**

**use history AND (medieval OR renaissance) for building up more complex searches.**

NOTE: you must use CAPITALS for these instructions. Check the 'detailed instructions' option for the detail in each case.

Some search engines allow you to search within a category. Use this facility. If you get down to 'history' through categories as education/research/academic etc you will obviously escape the life story of the family's pet hamster. And then follow the steps above.

# History of the Internet

by Richard T. Griffiths

## Directories

Directories are lists of sites, chosen by human-beings (they still exist).



WWW Virtual Library was set-up by Tim Berners-Lee who was the founder of the WWW. It is non-commercial and is run by a federation of volunteer institutions which follow certain rules and which try to ensure that the links are relevant and up-to-date. It might be worth looking at the home-page if you want a theme which might equally fall under economics or law etc., but it does have a history library, most of which is hosted by the University of Kansas. This can be accessed two ways;

alphabetically

thematically and geographically [<http://www.ukans.edu/history/VL/>]

I recommend the latter and you can see which bits are done by Kansas and which bits by other organizations (eg 'labor and business history' is coordinated by the IISG, Amsterdam). The central index has grown enormously from

July 1998: 2500 sites

July 1999: 4000 sites (plus about 1000 more in sub-sites)

January 2000: 5800 sites

The library is good... in parts. Peter Doorn, writing the notes for this course four years ago was scathing: links were not kept up-to-date, the classification was often misleading, and some of the collections were pathetic. The situation has now improved, but there are still areas where you wonder why you bothered. On the other hand, some of the sections are really excellent

# YAHOO!

Yahoo! (supposedly an anachronym for 'Yet Another Hierarchical Official Oracle' but this is now denied by its creators) is a commercial directory established in late 1994 by two PhD students at Stanford University (David Filo and Jerry Yang, who also developed the software). In 1995 Marc Andreessen, invited them to use the more powerful computers at Netscape, but it maintains a separate commercial identity. The principle is one of user registration, but it also uses an advanced 'spider' engine.

It is the largest of the directories, employing 80 editors and manipulating a database of over 1 million links. It also has a history section which has grown from:

# History of the Internet

by Richard T. Griffiths

July 1998: 10,500 sites

July 1999: 17,800 sites (directories) linked into almost 1000 categories

January 2000: 20,000 sites

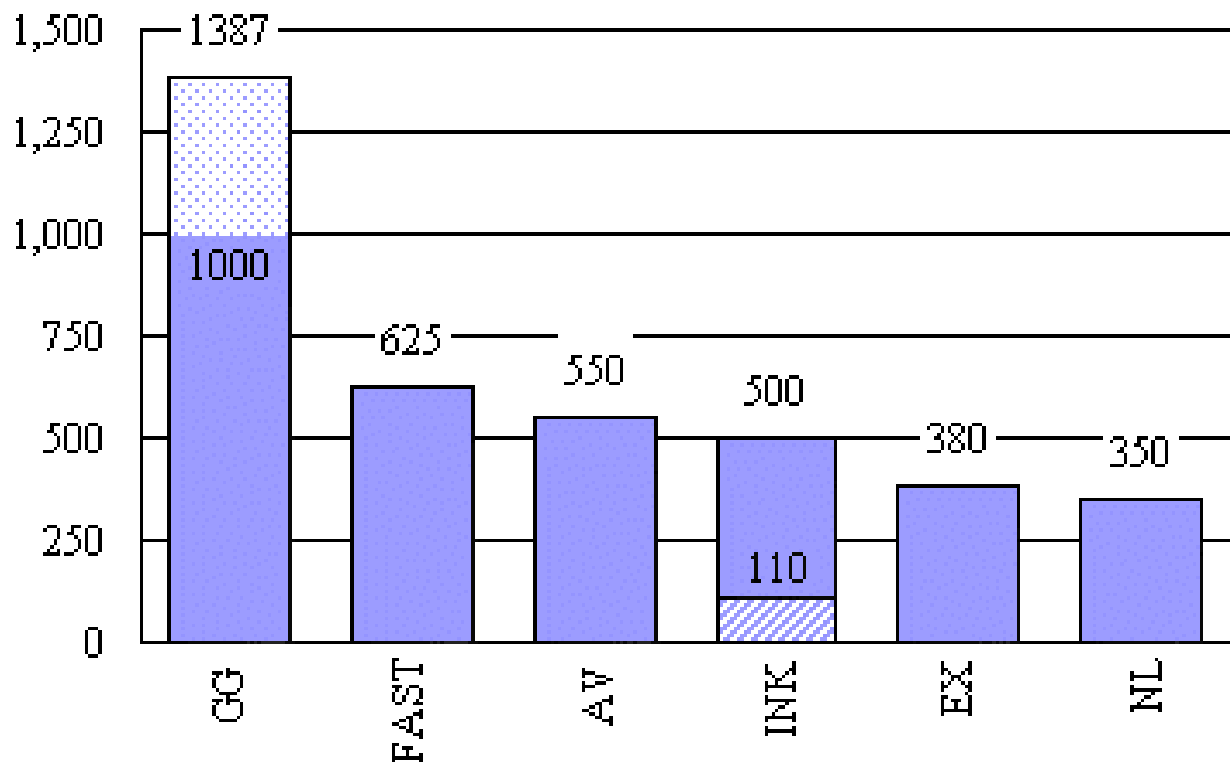
July 2001: 27,000 sites

It also has a 'search within category' function.

## Search Engines

Directories are rather like a library book catalogue, telling you the titles available. At the end of each book, there is often an index, telling you exactly where to look within the book to find a mention of a particular name or topic. Imagine if, instead of looking through each book, someone had torn out all the indexes from all the books and rearranged them so that the names and topics were put together. And then imagine that it can sort through all those pages.... a stack one hundred miles high... and give you the results (ranked in an approximate order of potential relevance) and give you the answer in under one second. You've got the picture. This is the phenomenal power behind today's modern search engines. This is the power of Goggle. One suggestion: although there are several meta-search-engines available (which simultaneously submit your research term to different search engines) you are better off looking through the first ten pages on one top-quality search engine than looking though the first page on ten separate engines.

## Millions of Web Pages Indexed



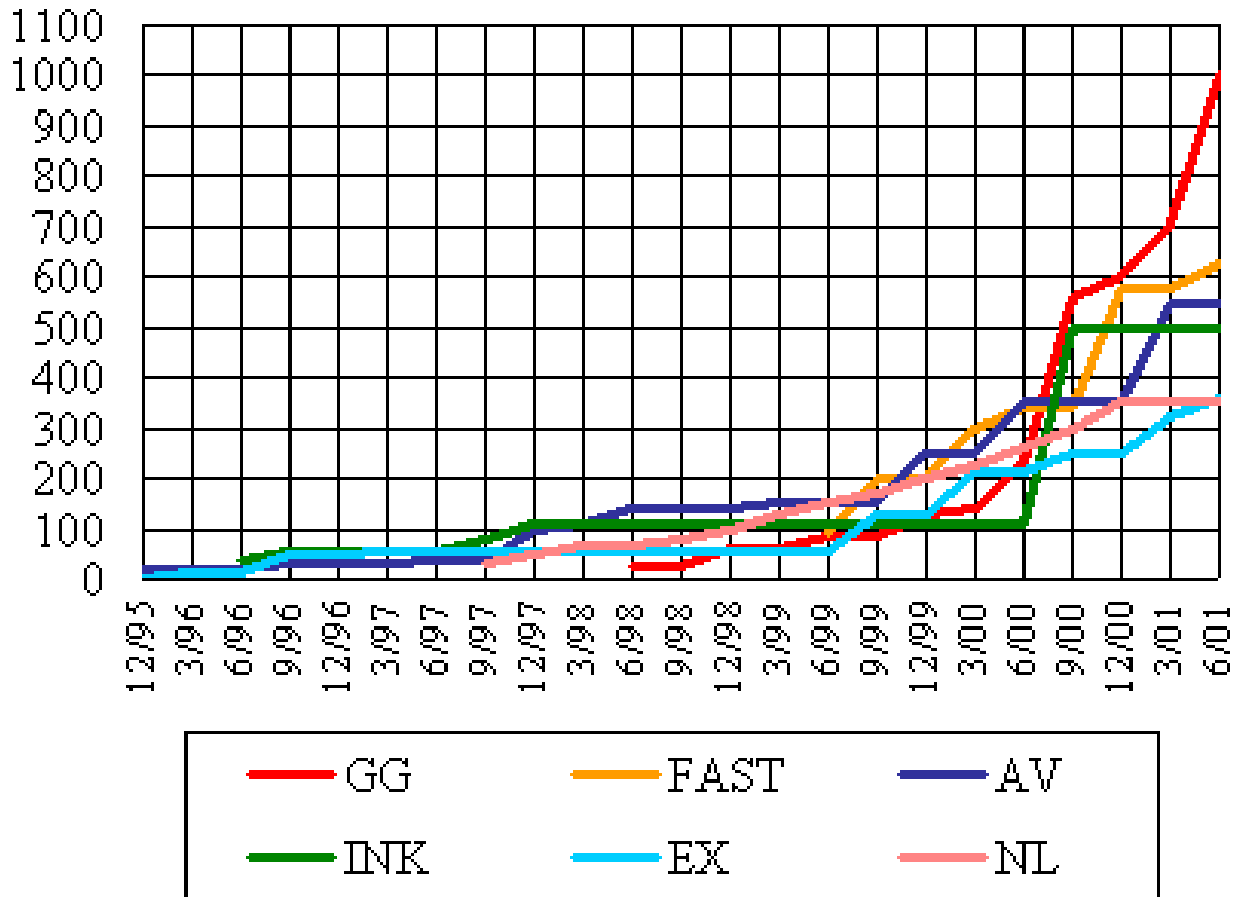
KEY: GG=Google, FAST=FAST, AV=AltaVista, INK=Inktomi, WT=WebTop.com, NL=Northern Light, EX=Excite. Also use this key for charts below.

# History of the Internet

by Richard T. Griffiths

## Search Engine Sizes

(millions of web pages)



There are thousands of search-engines, but we have selected three of the largest and most flexible. Search-engines operate by selecting individual web-pages or documents. Although some give you the option of selecting sites, their coverage is far smaller than the main directories. Many search engines use the same 'spiders' to compile their indices, so the difference lies in the way they interpret the data and how they allow you to manipulate the results. Keep in mind that when a search-engine gives you a couple of hundred of hits, on twenty pages, many people do not bother beyond the first three pages. It is worth pausing to reflect what these 'spiders' are looking for. Web-page makers offer the spiders four information sources:

- Page title or, by default, the first words (like 'welcome to my page')
- Description (written in meta-text, which means you can't see it on the screen) which tries to pack in as many keywords as possible
- Keywords (again in meta-text)
- The text of the document itself

Moreover, most engines allow page-makers to submit their locations for inclusion. Of these, the title is the most important determinant of content. Some search engines (including Alta Vista and Info seek) incorporate the meta-text when making rankings (looking at the frequency a word appears or how near the top) but others use the first paragraphs of the document text itself. So, even if they use the same spider, they will not necessarily give you the same documents in the same order. Moreover,

## History of the Internet

by Richard T. Griffiths

they way they categories their information and the search functions the provide also influences their usefulness.



Google I tipped this search-engine when it first started as one to watch. Google is a deliberate misspelling of googol - 10 to the power of 100.... but really chosen because the name sounded 'cool.' It was formed by two Stanford graduates in April 1998. The principle behind it was that it monitors other indices to see who links to what (and to rank these) in order to locate the real 'authorities' on a topic and it uses this to rank the results. It had superb, clean looks and it now also has the largest coverage of any search engine (1.6 thousand million web-pages) and it still delivers plenty of useable references within the first hundred or so results (it depends... but you can easily see when the quality tails off). If you use the "cached" pages (kept on its own computers) you can see you search terms highlighted in the document. It has recently introduced an "advanced search" category and an image search. In total, a fabulous resource.



AltaVista (meaning 'view from above') opened in December 1995 as an offshoot of Digital Computers. It has an index to 550 million web-pages, divided into 24,000 categories. It was also the first site to include a translation service (the same one behind some of our history web-sites) and a search facility for images and sound files. I used to rely a great deal on the image search function, but I fear that it has now been surpassed by Goggle, which gives more returns and which also allows you easily to see the images in their text context.



Northern Light (named after a record breaking 19th century American schooner) is a private company, established in 1995 employing 40 researchers. It has a reported 350 million web-pages in its index. It

# History of the Internet

by Richard T. Griffiths

is particularly proud of its 'special collection' documents which mostly comprise journal articles that usually evade search-engine spiders (and some can be ordered for a small charge).

## Databases

We saw that Northern Light was very proud of the fact that it integrated journal contents into its indices, but it is far from complete. Yet, Northern Light has highlighted a long-standing problem. We can hope to find books in library catalogues, but where we find recent journal articles? There was always one solution, and that was to locate journals that published indices (and abstracts) of scientific journals and (after locating the relevant sub-categories) to plough through them by hand. The most popular among historians was Historical Abstracts (available on diskette in the Leiden University's library) whilst economic historians could rely on Journal of Economic Literature. Some of these are now available on-line, and all with search functions, but most are only available within Local Area Networks (and you should check your own library's resources). For staff and students of Leiden University, you should go here and search further in ERL (Silver Platter).

A final route through to journal publications is through on-line bibliographies. The best way to access these is to enter the name of your field of interest into a search engine and add the term "bibliography" to your search. It might be intellectually satisfying to rediscover the wheel, but most of us have better things to do with our time.

## Library Catalogues

Finally we come to good, old-fashioned library catalogues. These are still the best ways for finding printed information which is classified and catalogued by the data on its cover. In other words, we can use library catalogues to locate books, journals and working papers.... but not to locate chapters or articles. Most major libraries began computerizing their catalogues in the 1970s and it was only a small step to making them available to the on-line community. However, this is not always possible through the WWW... many sites are reached through telnetting (not that you need to do anything) and they have an old-fashioned feel about them. But, especially for historians, there is still far more information available behind these catalogues than there is on the internet (and there will be for some considerable time to come). We assume that you are already familiar with our own university catalogue but most national (copyright) libraries have now put their catalogues on-line. You can find the list of addresses via Yahoo here.

