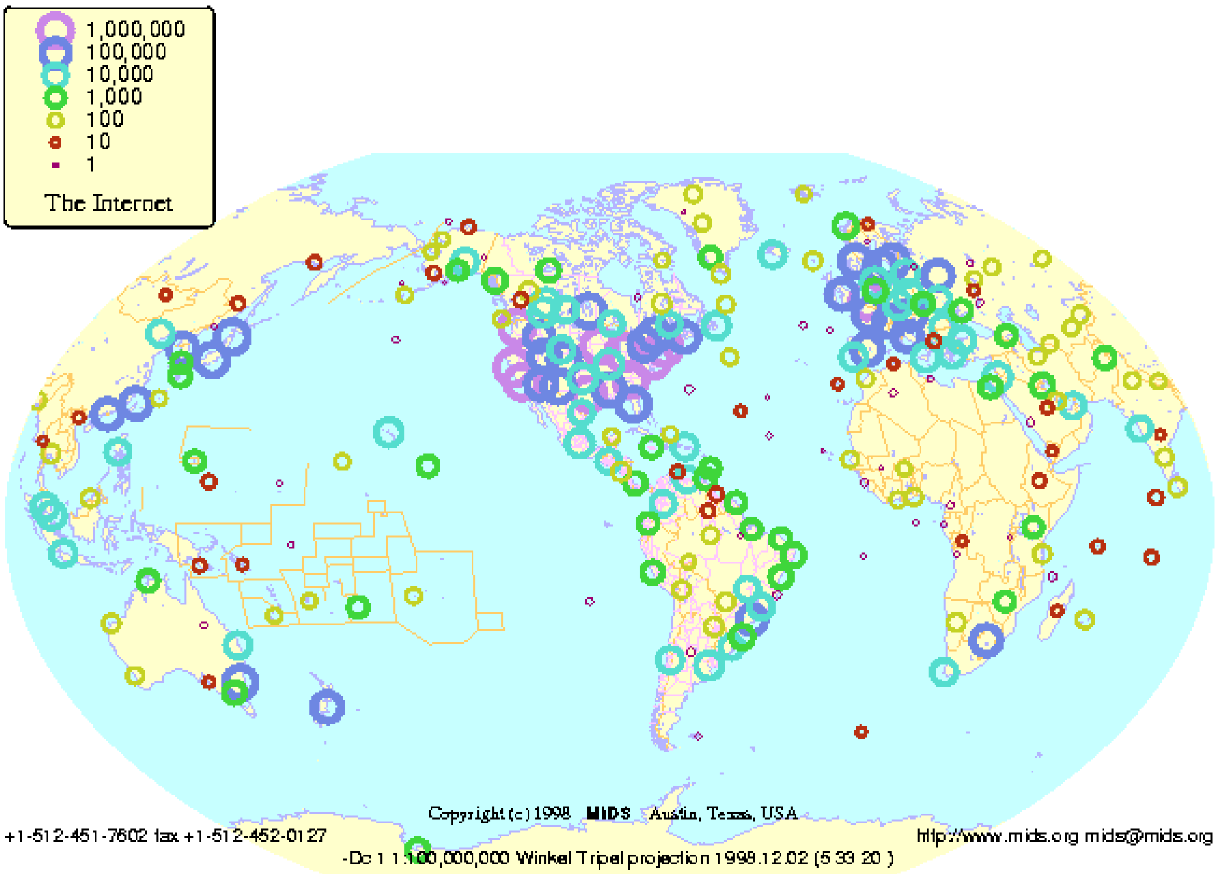
Planning for the Continued Rapid Change of the Internet

Hosts on the Internet in July 1998   Each circle represents the hosts in a major city of the world.



The Internet Map

This map shows the 40 million hosts around the world in the Internet matrix as of July 1998. See http://www.mids.org for a current version of this map image in color and for additional connectivity maps. (Map used with permission.)

At the start of 1999, 206 of the world's 246 countries and territories were on the Internet; represented by an estimated 50 million hosts, 700 million web pages, and 140 million users. In mid 1998 an estimated 1 PetaByte (1 quadrillion, 10 ** 15 bytes) per week was sent over the Internet (The then MCI Internet Backbone, now Cable & Wireless Backbone, alone was carrying 200 TeraBytes (trillion bytes) per week in mid-1998 (from Vinton G. Cerf, Sr. Vice President, Internet Architecture and Technology, MCI WorldCom)). The number of bytes sent over the Internet each week is currently doubling about every six to twelve months.

[This map was provided by Matrix Information and Directory Services, Inc. (MIDS) http://www.MIDS.org. This map was suggested by the Internet Society which serves as the international organization for global coordination and cooperation on the Internet (ISOC) http://www.ISOC.org. An Internet host is defined as a computer that is connected to the Internet. MIDS has researched the size, capacity, and growth of the Internet since 1993 and is the oldest company with such data available. MIDS organizes the information textually, graphically, and geographically. Matrix IQ at http://www.MIQ.net uses this data to provide an Internet monitoring service for ISPs (Internet Service Providers) and Corporate Networks. See the MIDS website.]

Introduction

The local loop of the telephone line defines the speed of the Internet for most people. The local loop has lacked the speed to make flipping through pages in books or full motion video available over the Internet. Storage systems for electronic documents have defined large-scale document storage and video storage as too slow, too unreliable, and too expensive to be consistent with the free paradigm of the Internet. The demands for processing power by electronic gaming are opening the way to publish interactive simulations as a means of expression. In the next three years, huge increases in speed, capacity, and processing power on the Internet, and on the local loop, will add new document types to the meaning of electronic publishing on the Internet.

Publishing has been defined as a way to make information widely available to the public. Our formerly mechanical world required many preliminary actions, such a typesetting and proofing, which became associated with the meaning of publishing. Electronic publishing makes is possible to instantly make anything available to the world with no preliminary preparation. Under these circumstances, materials can be divided between text posted to the Internet and works published with preliminary preparation.

THE INTERNET

The Next Three Years on the Internet

As the cost of switching and transmission over the Internet continue to drop, more and more types of documents can be delivered over it. Email is currently a given. Color pictures from web sites are expected to be high quality art. Next to come is audio and free phone calls over the Internet. When phone calls become free, telecommuting will become commonplace and travel-challenged individuals will be able to work at many more types of jobs from home. Ultimately, any job that is limited to answering a phone and viewing a computer screen can be moved to any location in the world. Adding video for free internet-based videophone calls will even further reduce the need for the physical meetings that are seen as essential for conducting business today.

Two processes are occurring simultaneously. First, enthusiasts are buying low cost digital video cameras for low resolution, low frame rate, and intermittent (choppy transmission) videophones. The second process is commercial. Telephone equipment vendors such as Nortel Networks and Lucent Technologies are moving to design and manufacture switches and transmission equipment to provide phone service over the Internet. The latest step in this commercial process was Lucent Technologies' US$ 20 billion purchase of the computer-networking vendor Ascend in January 1999. [4] This process will introduce 'quality of service' protocols on the Internet. The quality of service protocols will eliminate the intermittence of current Internet phone calls. Quality of service protocols will pave the way for broadcast TV, broadcast HDTV (High Definition TV), and finally by on-demand TV and on-demand HDTV. The introduction of commercial quality on-demand TV is likely to pay for great technical advances on the Internet without requiring additional payments by users for the technology. Intellectual property owners can continue to be compensated by advertising or through rental payments, but the public will rent movies over the Internet rather than going to video stores.

A longer distance into the future, wallpaper may become something that one unrolls, cuts, and sticks on one's walls (computer display material that one unrolls, cuts, and sticks on one's walls, ceiling, and floor which then become the display system.). One's PC may figure out where the wallpaper is on the wall, assemble a seamless display system from the glued strips of display material, and use it to display live, prerecorded, or simulated scenes of where one would rather be. Wallpaper may evolve, in its historic role of surrounding one with images, from a static to a dynamic medium.

In the case of the Internet's near-term evolution, the minutiae (technical details that are difficult to make use of) are in the reports that today's commercial fiber optic transmitters can send 320 Gigabits per second [5], that transmitters under development in March 1998 by Lucent technologies can send 1 Terabit (1 trillion bits) per second [6], and that Internet switches are being designed that can switch 1 Terabit per second (announced by Nortel Networks in October 1998). [7] On October 12, 1999, Nortel Networks announced that it plans to offer transmission and reception systems capable of 6.4 Terabits per second (80 OC1536 channels of 80 Gigabits per second each.) in 2001 [7a].

The changes in transmission and reception systems are very significant, because no change in the buried fibers is necessary to increase the speed of the network, only the transmission and reception systems need be changed.

The Local Loop

The Internet switching and transmission infrastructure is almost ready for all the video we could want. The telephone local loop is the last obstacle.

The local loop is the copper wire from the telephone central office to your home or office. (The central office is the nondescript, clean, windowless building, with the telephone company logo high on the wall, within 3 or 4 miles of your home or office.)

It is likely that DSL (Digital Subscriber Line) equipment and cable modems will increase the speed of local loop Internet connections available to most users by a factor of 10 to 100 in the next three years. These improvements are possible using the physical cable and telephone plant that is currently in place. However, these changes depend on decisions made by the users' cable and telephone vendors.

DSL is very much better than ISDN (Integrated Services Digital Network) because the data completely bypasses the entire telephone company switching system, and is therefore not affected by telephone company decisions. The data travels over the copper wire from your computer's DSL modem to the DSL modem in the central office. From the DSL modem in the central office, your data is placed directly on a packet multiplexed link to the Internet, bypassing the circuit switched telephone company infrastructure that is so inefficient for Internet based data communications.

Physically Replacing the Copper Local Loop with a Fiber Optic Link

The very slowest rate at which the local loop enhancement could occur is if it becomes necessary to build a completely new infrastructure to do an end run around the political obstacles currently emplaced by the incumbent local loop vendors. This would be equivalent to the laying of cable TV cables in urban areas, a process that took about a decade in the United States. The difference in cost between laying a copper and optical fiber link from the street (public right-of-way) to your home and laying a copper link without fiber is about 10 US dollars.

Fiber optic local loops have the additional advantage that they have the physical capacity to be 10 thousand to 1 million times faster than most users' existing local loop Internet connections. The actual speed of the fiber optic connections depends on the transmitters and receivers used, just as the speed of copper wire connections depends on the transmitters and receivers used. Thus, as in the case of copper links, the speed of fiber optic links can be increased by upgrading the transmitters and receivers, which will continue to drop in price because of their similarity to computer electronics, which are dropping in price rapidly.

DSL is an example of upgrading the transmitters and receivers on existing copper wire links. [4] For optical transmitters and receivers, there is very substantial headroom in the physical optical fibers for increasing data rates by improving optical transmitters and receivers. Even the 6.4 Terabit per second data rate mentioned above is still far below the theoretical capacity of the optical fiber carrier frequency (baud rate) of 230 THz (TeraHertz, trillion cycles per second) for fibers designed to carry 1,300 nm (nanometer) light.

Also waiting in the wings, and which might add a measure of destabilization to the market, is Teledesic, backed in part by investor Bill Gates. By 2003, Teledesic plans to provide up to 64-megabit signals to laptop sized roof mounted antennas and up to 2-megabit up-link signals from the antennas, worldwide, using 288 low earth orbit (850 miles) satellites. The pricing is expected to compare favorably with both local loop and cellular service. Teledesic also plans to make 64-megabit per second up-links available at a higher cost. [8]

The PC/TV/VCR Merger

Microprocessors have a transistor budget, the number of transistors that can be placed on a chip for the price at which the chip can be sold. The current microprocessor transistor budgets are between 5 and 10 million transistors. This has exceeded the number of transistors needed to run even the most complex microprocessor designs used in PCs. (Servers and workstations are still demanding more complex designs.) To sop up the remaining transistors in the budget, transistor based functions are being vacuumed up from the other chips and boards in the PC.

In the next generations of PCs, the microprocessor chip may absorb all the transistor functions in the PC leaving the PC with no adapter cards and no motherboard, just the microprocessor. [9] Now that PCs must play DVD video, the microprocessor chip may also absorb DVD, video, TV, and VCR functions, and still the transistor budget of the next generation of microprocessors may not be used up. Having run out of things for the ever-increasing number of transistors to do, chipmakers may place fewer transistors on the microprocessor chip than the transistor budget would otherwise allow. The microprocessor chip may then start to become less expensive, rather than maintaining a constant cost with an increasing number of transistors, as manufacturing processes improve. This reduction in cost is in marked contrast to the past twenty years when every increase in transistor budget was met with an increased demand for transistors (caused by an increased demand for PC functions) which kept the price of the microprocessors and PCs constant.

Our whole experience with PC prices is that the prices have stayed constant at about the two thousand five hundred US dollars that most people could afford. This is about to change radically. (There is speculation that PCs may become completely free. [10] See also the National Semiconductor - Cyrix Oct 13, 1998 Jalapeno chip press release [11], the National Semiconductor - Cyrix October 19, 1997 Cayenne chip press release [12], and a description of single chip PCs. [13])

At most, chips cost two or three US dollars to manufacture in terms of marginal cost. Marginal cost is the cost of producing one more chip, and does not include the cost of setting up the chip production line in the first place. In a few chip generations, the microprocessor chips may cost two or three US dollars and the PCs may cost ten US dollars (one hundred US dollars for skeptics).

This progression was born out at a recent HDTV public sales event held in Newport Beach, California, only weeks before the start of HDTV broadcasts in the Los Angeles area. Most HDTV demonstrations were playing DVDs or VCRs and using the HDTV sets' compatibility modes to display the pictures, essentially showing 50-year-old NTSC (National Television Standards Committee) signals on HDTV sets. A few of the vendors had what they called 'HDTV generators' that produced real HDTV signals for the sets. The generators were PCs that had the HDTV programming recorded on their magnetic disk drives.

The HDTV sets sold for around ten thousand US dollars. The vendors were waiting for full time HDTV broadcasts and for HDTV VCRs, which were expected to arrive in a few years. The vendors saw the video industry as profiting from the high prices on the new HDTV's and HDTV VCR's. The HDTV video equipment was high priced because it was controlled by the home video industry. At the same time, stock PCs were playing real HDTV programming and some of the vendors were showing the HDTV images on stock PC projectors. The home video industry is likely to lose control of its industry and its customers to PCs (and to Microsoft) at the same time the local loop vendors are loosing control of the physical transmission medium.

Has It Happened Before?

Yes, with PC memory prices.

For years, memory prices declined by half every one and one half years. Then, with the advent of Microsoft Windows 3.0, the demand for memory exceeded the supply, world wide, for three or four years. As the cost of production dropped with advances in technology, the gap between the market-scarcity driven price and the production cost continued to widen. This gap, the profit, was invested in new manufacturing facilities for memory. New, billion US dollar semiconductor plants were opening at the rate of one per week. Finally, in 1997, supply caught up with demand, and prices dropped to a little above production costs, a decline of 40 to 1, from US 40 per MegaByte to US$ 1 per MegaByte, a decline 97.5 percent in less than a year.

The demand for memory stemmed from the desire of consumers to use Microsoft Windows. When this need was met, consumers did not see a need for more memory. In 1999 most users cannot see a need for more than 32 MegaBytes of RAM in their PCs. They then buy 64 or 128 MegaBytes of RAM, secure in the knowledge that they will never need it. This has placed a cap on the demand for memory.

Similarly, while workstation and gaming users will continue to desire more computing power, most consumers think that their spreadsheets, word processor documents, and Internet pages display wonderfully fast on a 450 MHz (MegaHertz) PC, and in fact, would not see any difference in speed on a faster PC. It is the demands of the huge numbers of average PC users that establishes the limits on PC capabilities, for both memory and speed.

HISTORY OF COMPUTER TECHNOLOGY AND THE INTERNET

Minutiae

When the word minutiae was coined, the idea that a two-hour movie would be stored as 4 GigaBytes or 32 billion bits on a DVD (Digital Video Disc in common usage) was inconceivable. Also inconceivable was the 450 MHz (MegaHertz) PC that could execute the 450 million instructions per second required to decode the DVD in software and play the movie or run the simulation of a DVD based game.

When overwhelmed by minutiae, one asks: "Is it big, or is it small?" Big and bigger become 'big', fast and faster merge into 'fast' and descriptions of future change become meaningless and unavailable. MegaBytes, GigaBytes, and TeraBytes become 'Whatever. . . '.

Digital Watches

Products are of interest when their price becomes accessible. It has always been possible to build digital watches; they were just too expensive to sell. When digital watch prices declined to three thousand US dollars in the early 1970's people could choose a digital watch over a diamond encrusted solid gold watch. We became interested and we watched digital watch prices drop to three dollars in the 1980's.

Digital watches are of interest here because they are electronic, like the Internet, like PCs.

Digital watches have continued to drop in price in the 1990's but we have been uninterested. Who cares whether the electronics in the watch cost thirty-five or thirty-six cents? The watches must continue to cost three dollars to pay for the marketing expenses to advertise and sell the watch. The only way to eliminate the marketing expenses from the cost of the watch is to give it away, which is why so many electronic watches are given away.

The moral here is that all costs in electronics continue to drop forever. It is just that we do not care about the drop before the drop makes a given product affordable, and we do not care about the drop after the cost of electronics becomes insignificant in the cost of a product.

(In the extreme, the promise of nanotechnology is to reduce the cost of all physical things to zero. [1,2,3] (Drexler, 1986, 1987) (Drexler 1991) (Drexler, 1992) This may seem outlandish, but so is talking over a wire or free worldwide email and hyperlinked webpages.)

The Spread of New Technologies

When the price drop of technology enters the window of economic viability or interest (becomes affordable), the technology spreads rapidly. Printing spread rapidly in the fifteenth century. The telephone spread very rapidly in the 1880's. Television spread rapidly in the early 1950's. The Internet became ubiquitous in the three years preceding 1999. The DVD promises to replace CD drives in PCs, with DVD drives that can read CDs, in one year, following Microsoft's introduction of Windows 98 in mid-1998. Who would buy an obsolete technology (a CD reader) when one could buy a similarly priced equivalent new technology (a DVD/CD reader)?

The spread of new technologies has always been blindingly fast, and it is getting faster. This accelerating rate of change is another problem in minutiae: we have had enough change on the Internet, and we are not ready for any more. Many people are already planning their systems as though the Internet will stay forever the way it is today.

The Internet is going to change more in the next three years than it has changed in the last three years, and for most people the Internet did not exist three years ago.

PRESERVATION

Cost of Digital Storage

The cost to purchase one GigaByte of online magnetic disk storage is about 40 US dollars today. The cost is dropping at about forty percent per year and has been for the last ten years. This explains why computer shoppers are now looking for GigaBytes rather than MegaBytes. It also explains why, when a given configuration is costed out for an optical disc jukebox and for magnetic disk storage with off-line optical backup, the magnetic disk solution is often more cost effective.

This estimated decline in storage costs is based on IBM's predictions for its magnetoresistive (MR) head technology. IBM invented the MR technology, and MR is currently the technological basis for advancements in the magnetic disk industry. (See Appendix).

Writeable DVD disc blanks are expected to drop to the current one US dollar cost of writeable CD disc blanks in a few years. This will allow the duplication of 4 GigaBytes for one US dollar. These estimates are based on the assumption that DVD disc blanks will follow the price trends of CD disc blanks. For a history and current figures see [14] and (Taylor, 1997).

In the not too distant future, the cost of storing movies online, and the cost of offline DVD backup may become inconsequential. Libraries may be able to digitally store everything the libraries wish to store. The cost of digital storage may cease to be a gating issue in collection development.

Security, Integrity, and Preservation

In computing, security is defined as the process of preventing others from modifying library and other materials that are stored digitally. Integrity is defined the process of preventing one's own actions from unintentionally modifying library and other materials. Preservation includes security and integrity and covers several time scales. Preserving the availability of library materials means providing perpetual instant access. The ability to reload a backup copy in an hour, day, or week is a behind-the-scenes activity that preserves perpetual instant access. Preserving information forever is a goal that makes preserving information for a century, for a millennium, or for all of recorded history more likely.

For reliability purposes, if the cost of magnetic disks is inconsequential, then the cost of having two duplicates of the magnetic disks is also inconsequential. Having documents on magnetic disks in one city or region mirrored on duplicate magnetic disks in another city or region may become commonplace. In fact, it may become a professional responsibility of librarians and archivists to ensure that materials on the Internet have a permanent logical location name and a mirrored physical location so that the library materials will be available even during a disaster.

Security and integrity are not technical issues. One does not have to be technical to say: "Show me how you would restore the backup tape." One could use the installation of a new system to test the restoration of the latest backup of the old system. This is not an unreasonable scenario; if the old system was physically destroyed, one would be restoring the backup to a greenfield (new) system.

Encrypted checksums (Garrett, 1996) can attest to the integrity of recorded information. To create an encrypted checksum, one adds up the bits that make up a digitized document (the bits can be interpreted as numbers for this process) and creates a checksum. (Addition is a simplification of the actual process.) This checksum can only be produced from the configuration of bits that represent the current version of the digitized document. This checksum can then be encrypted so that an intruder wishing to change the document could not surreptitiously create a new checksum from the bits in the altered document. In the not too distant future, the first task of the day for librarians and archivists may be to check the system of checksums that protect their holdings.

Because the encrypted checksum process only encrypts the checksum, the document remains in clear text, in readable form. This allows the document to survive the demise of the civilization which holds the encryption key for the checksum.

The gold discs affixed to the Voyager spacecraft twenty years ago were designed to last 1 billion years. Now that the Voyager discs are outside the solar system (and the spacecraft has sent signals indicating it was beyond the solar system) the discs are immune to anything that might affect planet earth (errant asteroids, etc.). Thus, there is historical precedent (twenty-year-old historical precedent) for preserving information forever. Norsam Technologies is developing nickel computer discs that will last several thousand with very little environmental protection. Norsam is also planning iridium discs that may last billions of years. Norsam's discs can have raster images of documents recorded on them so that the images can be viewed optically with a microscope, avoiding the problem of obsolete disc readers. [15]

NEW PUBLISHING FORMATS

Runtime Editing

Many of the newer publishing mediums provide a means of modifying the content, or the presentation of the content, of a work while the works is being used. Some hyperlinked books have multiple story threads. Readers can choose one of multiple outcomes of events in the story. DVDs provide runtime editing to transparently delete the R rated parts of a movie when one sets the DVD player to display the G rated version of the movie. XML (eXtensible Markup Language, see Typesetting and Unicode section below.) has been extended to facilitate the conversion of textual displays to speech for vision impaired users. With a simple and private self-administered color-blindness test, the degree of one's color-blindness could be added to one's personal and confidential display profile. Currently, one's personal display profile contains one's personal preferences, such as icon placement and background colors, for one's computer based desktop display. The color-blindness profile would allow the computer display to shift all displayed colors into a color-space that could be seen by the user.

Shifting the display color-space for color-blind individuals and transparently deleting the R rated portion of a movie can be applied after documents are published. Over time, authors may create works to facilitate the use of dynamic runtime editing, as the authors of hyperlinked books, with multiple story threads, have already done. Authors might create multiple versions of the content of books to facilitate multiple alternate book designs to accommodate many different groups of readers, including persons with limited visual acuity. For persons with limited visual acuity, this might include the incorporation of large type fonts and the modification of tables and other graphics.

Runtime editing is still a static, predefined process. In a simulation, an author creates a program that actively tracks a user's interaction with the author's dynamic published work. This tracking once had to be left to the library patron's imagination, now an author can take a more active role in the reader's experience.

Simulations

Published works have always been the authors' simulation of reality. Readers of books have always been able to project themselves into the authors' virtual realities and take part using the readers' imagination. Dreams are a form of this and represent an area that is clearly beyond our current technology.

Electronic gaming is just a new form of what publishing has always been. Simulations can be interfaced to each other creating something like an encyclopedia for an imaginary world, just as a current encyclopedia is a group effort at imaging or describing our world. Multiple people can take part in a simulation, just as people can have a literary discussion or enjoy a movie or play together as a group activity. Disney's "Honey, I Shrunk the Audience" attraction continues the effort to involve the audience in the presentation of a literary work.

Facilities to support simulations may be expanding even faster than the Internet is growing.

Microsoft Windows 98 comes with support for eight computer screens, more than enough for a full cubic-surround (top, bottom, and four sides) visual experience in simulations with 6-track surround-sound. PCs are soon to enter the GHz (GigaHertz) or billions of instructions per second range. IBM announced GHz copper wired integrated circuits in 1997 [16] and licensed the technology to Sun Microsystems. Intel then announced similar products that will be available very early in the next millennium. Hand in hand with faster processors, new algorithms are being developed to reduce the computational load of simulations: to do more with less.

While microprocessors are getting faster, the amount of data being processed is also increasing. Intel's Merced and McKinley chips do 64-bit processing [17]; simultaneously processing twice as many bits as current 32-bit PCs. At the same time, multiple processors are being incorporated in single PC's creating multiprocessor configurations. Multiple PC's can be configured in clusters, linked by very high speed SANs (System or Storage Area networks) further increasing the speed available for simulations. In November of 1998, Unisys announced public availability of a 16-node cluster of 10 processor PCs that have 160 of the latest Intel microprocessors working together as a single system and announced plans for a 512-processor system. [18]

In the wings are plans to use all the computers on the Internet in a cooperative way.

The demands of simulation bring into sharp focus the fact that a significant portion of the world's economic resources are now being concentrated on advancements in computing and networking. This investment is driving hard to push the envelope for every aspect of software and hardware design and implementation. At the same time, advances in science and fundamental technologies are independently delivering huge increases in capabilities. This flow of money and technical advancement is what is catapulting libraries, archives, records centers, and document management systems into a new world of speed and capacity that must be molded to serve the needs of the library, archives, records center, and document management system users.

Real-Time Publications

Live television combines the spontaneity and reality of the moment with the high production value of multiple camera angles, managed audio, simultaneous commentary and translation, an array of cuts (from dissolves to warps) and even the inclusion of pre-recorded material. In live television, publishing occurs in real time.

There are many Internet cameras that show the same scene seven days per week, twenty-four hours per day. These cameras broadcast raw data with no production value. There is not publishing, it is just broadcasting, and perhaps recording.

The excitement of giving instant access to real data has given the Internet a long history of carrying real time data. Broadcasting real time data is not new in this millennium. Clock towers, church bells, and Mullahs on minarets have called the faithful to prayer for centuries. Samuel Clemens got his pen name by continuously reporting that his sounding stick showed a depth of over two fathoms, more than the draft of a riverboat. A sounding beyond the second fathom mark was reported as a loud 'Mark Twain', signifying a safe water depth.

As production values increase, real time data broadcasting will become real time publishing. The depth of water in California reservoirs, along with inflow, outflow, special notes, histories, and explanations is available, hour by hour, at: [19] and at [20]. For a specific reservoir, see Shasta Lake at: [21].

Better know are the Internet weather sites, such as [22] and [23], where many search techniques and well designed graphics are used to create a customized view of current weather information for each user.

The hundreds of earthquakes that occur in Southern California each day are measured, their epicenter is plotted, and a symbol showing the earthquakes' magnitude, location, and hours since occurrence is shown on a real-time map, along with major fault lines, at [24]. See [25] for an overview and [26] for the authors. In this case of real time publishing, almost all of the production value was added for the first geographic areas implemented. Very little additional production effort would be required to extend this type of seismic real time publishing to the entire world.

An example of real time publishing with a highly finished product is the Los Angeles area freeway traffic display at [27]. The display is updated every 30 seconds using data from over one thousand sensors buried at approximately half-kilometer (quarter-mile) intervals along the freeways. The displayed freeway map shows traffic speed at each sensor point with a color: red for under 20 miles per hour, yellow for 20 to 35 miles per hour, and green for over 35 miles per hour. An entire freeway route can be checked for congestion at a glance. The real time map fulfills all the cartographic requirements of quickly conveying accurate geographic information.

The real time editing software for the traffic display is being extended to work with other traffic sensor arrays in the roadways of other cities: [28]. In many cases all that is necessary is to draw a new city map, identify the format of the sensor data, and input the sensor data to the editing software. Additional cities can be brought on the Internet for a very low cost. Conversely, future sensor array installations can be justified because their valuable information can be conveyed so effectively to a very large number of roadway users in real time.

The traceability to real data gives a grounding less frequently found in our lightly-researched world. The vague but common knowledge that Bill Gates is the world's richest person takes on a new, more galvanic meaning, with a visit to [29] which estimates Bill Gates' current wealth, minute by minute, to the nearest thousand US dollars, based on his Microsoft stock holding. By watching this web site, one can watch the world's first and only centi-billionaire (one hundred billion US dollars), Bill Gates, as he works on his second hundred billion.

Further into the future, the real time publications may be integrated. The moisture content of storms may be estimated from the radar images of the storms. The rate of precipitation over a reservoir basin may be measured and validated against the rate of inflow into the reservoir. The rate of precipitation may be mapped onto roadway displays to show current and projected effects on traffic.

The ubiquity of the Internet is still increasing. As new users come onboard, they find more and more features available at no cost. In the next few years, real time publishing may be extended to worldwide data sets. In many cases real time publishing may be free on the Internet. Like many Internet features, real time publishing is both a result of the Internet's existence, and a driving force behind the Internet's ubiquity and low cost.

THE BOOK AND THE INTERNET

Book Speed

Strangely enough, the Internet has not been capable of competing with the book for speed of access.

One of the preeminent access features of the book is the ability to flip through the pages. To see how speed is defined in this context, imaging looking for something in a dictionary. You flip through the pages of the dictionary until you get to the right page and then you look down the page. Notice how many pages you flip through before stopping. It is usually 50 to 100 pages. How long does it take? Answer: about 1 to 2 seconds.

The first and last words that are defined on a page in a dictionary are reproduced in bold type at the top of the page in the dictionary. It is these words that we look at as we flip through the pages. This is an example of book design, part of the preparation done in publishing. Book design makes it easy to find chapters, section heading, and illustrations, all at 50 to 60 pages per second.

Internet designers think that displaying one page per second is very fast. This definition of speed makes it impossible for the Internet, and most computer systems, to compete with the book. This also explains why computer systems designers, who by definition have the best access to computers, surround themselves with printouts and books. Computer system designers intuitively know that paper provides faster access to information than the systems they have designed. Their definition, and our definition, of page flipping speed are set by book design. And, book design evolved in the context of the definition of speed set by our visual system which can perceive changes at the rate of 50 or 60 per second.

In light of the fact that, after almost 20 years of document imaging, a one page per second flip rate is considered fast and a 10 page per second rate is considered overkill, it is unlikely that computer system designers will ever see that page flipping occurs at 50 to 60 pages per second. However, by implementing HDTV video in PCs, a page-flipping rate of 50 to 60 pages per second will automatically be provided and the paper surrounding the computer systems designers' PC development stations will magically disappear.

Resolution and Contrast

HDTV compares favorably with the resolution available in movies. PC monitors still struggle to reproduce the contrast in HDTV images. For text on paper, books still have a much better resolution and a much higher contrast than computer screens. For resolution and contrast, books are far ahead of PCs and the Internet and there is no current technology that will place PCs and the Internet ahead of books in this area. A few decades in the future nanotechnology is likely to sweep these problems aside, along with computers, as we know them. [1,2,3] (Drexler, 1986, 1987) (Drexler 1991) (Drexler, 1992)

Typesetting and Unicode

Books have been able to reproduce any written language. In computing, we have had ASCII (American Standard Code for Information Interchange) [30] to represent written languages. ASCII has emphasized English and has several somewhat compatible extensions to circumvent the limitation that ASCII can only represent a maximum of 256 characters. ASCII is limited to representing a maximum of 256 characters because it uses only 8 bits to represent each character. In Microsoft Windows 2000 and Microsoft Office 2000, and in Microsoft Windows NT, the predecessor to Windows 2000 [31], ASCII is replaced with Unicode [32]. Unicode is a compatible superset of ASCII that can represent up to 65,536 characters because Unicode uses 16 bits to represent each character. Other vendors are supporting the Unicode effort as well. Unicode is intended to represent the characters of most languages.

(Unicode is often called a two-byte code. This is a misnomer, because 'byte' is defined as the number of bits used to represent a character. Thus, an ASCII byte is 8 bits and a Unicode byte is 16 bits. The standard term used to define a group of 8 bits is 'octet', a term that appears often in data communications and other computer standards. When terms were less rigorously defined, and memory was a million times more expensive, half of an 8-bit byte was called a 'nybble'.)

The change from ASCII to Unicode will make many of the elements of the Internet publishing foundation, that are currently ASCII-based, more language independent. HTML (HyperText Markup Language) and the database metacodes used to describe and catalog documents are examples of Internet publishing elements that are currently ASCII-based. HTML is in the process of being replaced with XML (eXtensible Markup Language), which supports Unicode. (See [33] for XML and HTML. See [34] for XML, HTML, and SGML (Standard Generalized Markup Language), the basis for HTML and XML.) While the substitution of 16-bit Unicode characters for 8-bit ASCII characters is intended to be transparent to users, the change may present challenges to implementers under some circumstances.

BENEFITS OF TRADITIONAL PUBLISHING TECHNOLOGY

Publishing books was expensive. This had many important benefits. First, it could not be done often, so authors had to collect their thoughts and have them well organized. Any changes would have to wait for a second edition, which would not be published for many years. The second edition might never be published; this gave authors impetus to being done with their creative work before starting on the publishing trail.

The publishers who invested in publishing an author's work wanted assurance that their money would have a reasonable return. To ensure quality publications, publishers paid for pre-publication peer review of manuscripts. Publishers also did careful editing and proofing. Changes between editions were substantial, well thought out, and were often listed in a preface. An author's internal thought process was not laid out in a series of daily editions in which the ebb and flow of good and bad versions of their presentation washed over their hapless readers.

Hypertext has always had a micro-hypertext side in which every keystroke and mouse click is recorded as a reversible change which creates a new document version. Hypertext is also, at least conceptually, designed to stay out of the way when not wanted. This unobtrusiveness would allow the viewing of book editions without seeing the daily changes (or the dailies in movies, which are made from yesterday's takes). Hypertext carries with it the possibility of recording too much and the professional responsibility, on the part of librarians, to avoid preserving too much.

In the mechanical world, books were expensive and were well cared for. Libraries kept track of where books were shelved. Because the number of editions was severely limited, the fact that a book was in multiple libraries meant that there were multiple identical copies of a work to survive disaster. On the Internet, no one knows which of many versions of a work is available, where it is available, or when it will cease to be available at each of the locations where it is currently available.

When one referred to the edition of a book, others could find an exact copy of the work. Now, with electronic publishing on the Internet, even the identification of a revision is often missing. The emphasis is on the most recent copy, not on the version on which a referring author based observations or conclusions.

As a word, 'published' may be too worn or trammeled to use. 'Publishing', or its successor term, may be defined as a careful collection of thoughts, infrequently done, with careful review and editing, made available from a permanently available location.

CONCLUSION

Electronic publishing will bring new document forms and has already created a need to better define the term publishing. Speed is as important an aspect of documents as physical size and logical organization. Libraries will continue to preserver and make available authors' views of the world. The professional responsibilities of librarians will remain the same but the challenges of implementation will be greater as the rate of change increases. For more details, updates, and background information on this paper please see [35].

Sidebar: Billions and Billions

Transistors are so inexpensive that nobody cares much about them any more. But, how many can one buy for a dollar? Every bit in RAM (Random Access Memory) requires a transistor to hold the value of one or zero in an associated capacitor. There are eight bits in a byte and a million bytes in a MegaByte. The cost of a MegaByte of RAM is closing in on one dollar. At a dollar per MegaByte for RAM, one can buy eight million transistors for a dollar. Twenty-five years ago, RAM cost one dollar per byte, so the price reduction for RAM has been one-million-to-one over the past twenty-five years. As one can see from the table in this article, magnetic disk space will also drop by a factor of one-million-to-one over thirty years, so the cost of magnetic disk space is dropping in price at about the same rate at which RAM has dropped in price in the past.

For the nostalgia buffs, in the 1950's transistor radios were advertised based on the number of transistors they had. An eight transistor radio was said to be better than a six transistor radio. Then manufacturers started adding transistors for the sake of advertising rather than improving the radio, and the ten transistor radio was born. This eventually ran its course. But now, a 128-MegaByte PC has over one billion transistors. Soon school children will say: "*My* computer has a billion transistors.", "*My* computer is better, it has two billion transistors.", "*My* computer is the best, it has billions and billions of transistors!", ". . . or is that hamburgers . . .", ". . . or stars?"

Endnotes

[1] http://www.Foresight.org
[2] http://www.Foresight.org/EOC/index.html
[3] http://www.Foresight.org/UTF/Unbound_LBW/index.html
[4] http://www.Lucent.com
[5] http:// www.Nortel.com/broadband/news/1998c/9_15_9898438_Optical_Terabit.html
[6] http://www.Lucent.com/ideas2/perspectives/science/terabit.html
[7] http://www.Nortel.com/cool/navigator/navigate295/navigate1/terabit.html
[7a] http://www.nortelnetworks.com/corporate/news/newsreleases/1999d/10_12_9999633_80gigabit.html
[8] http://Microsoft.com/BillGates/interests/teledesic.htm
[9] http://www-CGI.CNN.com/TECH/computing/9810/07/chippower.idg/Computer world.
[10] http://www.cnn.com/TECH/computing/9810/15/freepc.idg/
[11] http://www.National.com/news/item/0,1567,329,00.html
[12] http://www.Cyrix.com/html/about/1997/cayenne.htm
[13] http://web-e7.ZDnet.com/pcmag/insites/inside_track/it980922.htm
[14] http://www.VideoDiscovery.com/vdyweb/dvd/dvdfaq.html

[15] http://www.Norsam.com
[16] http://www.chips.IBM.com/bluelogic/showcase/copper/
[17] http://www.Intel.com
[18] http:// www.marketplace.Unisys.com/ent/scalability.html
[19] http://cdec.water.CA.gov
[20] http://cdec.water.CA.gov/misc/hourly_res.html
[21] http://cdec.water.CA.gov/cgi-progs/queryF?sha
[22] http://www.AccuWeather.com/weatherf/index_corp
[23] http://banners.wunderground.com/
[24] http://www.scecdc.scec.org/recenteqs/Maps/Los_Angeles.html
[25] http://www.scecdc.scec.org/recenteqs/index.html
[26] http://www.scecdc.scec.org/recenteqs/credits.html
[27] http://traffic.Maxwell.com/la/index.html
[28] http://traffic.Maxwell.com/index.html
[29] http://www.webho.com/WealthClock/
[30] http://www.ANSI.org
[31] http://www.Microsoft.com
[32] http://www.Unicode.org
[33] http://www.w3.org
[34] http://www.OASIS-open.org
[35] http://www.ArchiveBuilders.com

Reference List

Bashe, C. (1986), IBM's Early Computers, MIT Press, Cambridge, MA.

Drexler, E. (1986, 1987), Engines of Creation, Anchor Books, New York. Available http://www.Foresight.org/EOC/index.html

Drexler, E. (1991), Unbounding the Future: the Nanotechnology Revolution, William Morrow and Company, New York. Available http://www.Foresight.org/UTF/Unbound_LBW/index.html

Drexler, E. (1992), Nanosystems, Molecular Machinery, Manufacturing and Computation, John Wiley & Sons, New York.

Garrett, J. and Waters, D (1996), Preserving Digital Information, the Report of the Task Force on Archiving of Digital Information, commissioned by the Commission on Preservation and Access and the Research Libraries Group, Research Libraries Group, Mountain View, CA.

Grochowski, E. (1994), The Era of Magnetoresistive Heads, IBM Research Division, Almaden Research Center, San Jose, CA.

Taylor, J. (1997), DVD Demystified: The Guidebook for DVD-Video and DVD-ROM, McGraw Hill, New York.

Keywords

Internet, publishing, flip rate, Unicode, history of technology, storage cost

Updates and More Detailed Descriptions

When using the information in this article, please check the website http://www.ArchiveBuilders.com for updates. The version number of this article is just before the page number below. The website also has articles that provide more details on some of the terms and concepts in this article.

Comments

Please let us know how you like this paper, or if you had any questions. What would you like to see in the future? Also, please let us know where you saw this paper. For more, and the most recent version of this article, please visit our web site at http://www.ArchiveBuilders.com

Please send your comments via email to SteveGilheany@ArchiveBuilders.com. Tel: +1 310-937-7000. Fax: +1 310-937-7001. Also, please let us know where you saw this article.

Acknowledgements

Reprinted and updated from Library Hi-Tech January 1999, incorporated by reference in Archive Planning, Volume 3, number 10, 1999, Archive Builders' analysis newsletter for document management.

See http://www.ArchiveBuilders.com

Note to Editors

Paper 22014v048

We will continue to update these articles as we get comments. Please contact us for the most current version before you publish. Also, please request permission to publish the article. Permission will be given freely for most non-commercial purposes.

Steve Gilheany
Archive Builders
1209 Manhattan Ave.
Manhattan Beach, CA 90266
Tel: +1 310-937-7000 Fax: +1 310-937-7001
SteveGilheany@ArchiveBuilders.com

Dividing this Article into Parts for Serialization

If you decide to divide this article into parts please print at least the updates, comments, and acknowledgements sections in each of the parts along with:

'by SteveGilheany@ArchiveBuilders.com'

Bio

Steve Gilheany, BA in Computer Science, MBA, MLS Specialization in Information Science, CDIA (Certified Document Imaging System Architect), AIIM Master (MIT), and AIIM Laureate (LIT), of Information Technologies, CRM (Certified Records Manager, ARMA) has eighteen years experience in document imaging and is a Sr. Systems Engineer at Archive Builders.

Author

Steve Gilheany is a Sr. Systems Engineer at Archive Builders. He has worked in digital document management and document imaging for twenty years.

His experience in the application of document management and document imaging in industry includes: aerospace, banking, manufacturing, natural resources, petroleum refining, transportation, energy, federal, state, and local government, civil engineering, utilities, entertainment, commercial records centers, archives, non-profit development, education, and administrative, engineering, production, legal, and medical records management. At the same time, he has worked in product management for hypertext, for windows based user interface systems, for computer displays, for engineering drawing, letter size, microform, and color scanning, and for xerographic, photographic, newspaper, engineering drawing, and color printing.

In addition, he has nine years of experience in data center operations and database and computer communications systems design, programming, testing, and software configuration management. He has an MLS Specialization in Information Science and an MBA with a concentration in Computer and Information Systems from UCLA, a California Adult Education teaching credential, and a BA in Computer Science from the University of Wisconsin at Madison. His industry certifications include: the CDIA (Certified Document Imaging System Architect) and the AIIM Master (MIT), and AIIM Laureate (LIT), of Information Technologies (from AIIM International, the Association of Information and Image Management, www.AIIM.org), and the CRM (Certified Records Manager) (from the ICRM, the Institute of Certified Records Managers, an affiliate of ARMA International, the Association of Records Managers and Administrators, www.ARMA.org).

Contact:

SteveGilheany@ArchiveBuilders.com
Tel: +1 310-937-7000 Fax: +1 310-937-7001

For more information, courses, and papers:

http://www.ArchiveBuilders.com