

Why the way we use the Internet is not Easy or Effective

Gautama Payment, November 2003

“The net has not caught up with the more expansive hopes of unfettered, free flowing information which were once attached to it. But the technical potential it opens up comes close to the enormous system of lateral cross-referencing, which hypertext networks Ted Nelson first named Xanadu in the 1960s, and the system Vannevar Bush called the Memex in the 1940s. Both of these conceptions were far more interactive than the system of the mid 1990s allows.”

- Sadie Plant, from ‘Ones + Zeros’, published by Fourth Estate, UK 1998

WE ALL USE THE INTERNET

The Internet, or the world-wide-web as we also know it, has grown to become the largest openly accessible repository of free information content available to any person, anywhere, who wants to merely ‘find out about’ something. Whether we make the connection from our home, the office or school; by dial-up or broadband; or remotely connecting by WAP or satellite from the North Pole, Sahara Desert, or twenty thousand feet above in a jet plane – it’s all the same.

It is an experience of finding, accessing and using information that is so widespread and integrated into our lives, that for the large part, we take for its use for granted and the way we are required to use it as – the way we use it. Normal.

The Internet has grown from the ideal dreams and visions of early pioneers such as Vannevar Bush, Ted Nelson and Douglas Engelbart; from the cold war military innovation that was ARPANET; and from the fevered almost religious efforts of all the early adopters and developers who tinkered, hacked, rigged, walled and rehacked the system time and time again, to bring to us the world-wide-web: a system for everyone.

Today we can pretty much find any information at all which exists already in the offline ‘real world’ and then some more – from sporting events, news and weather to massively multiplayer computer games, chat networks and intricate online communities. We find real world styled online events surpassing their offline equivalents, such as Timecode (March 2000) the first online world film premiere of a major Hollywood film at the Yahoo! Internet life film festival, or seeing the appearance of new distribution networks, such as Supari (June 2003) the latest Indian Bollywood blockbuster being distributed globally by Sharmon Networks across their peer-to-peer file-sharing network, Kazaa. Today it is just You and the rest of the world, twenty-four hours a day, all from the simple ‘tap tap tap, click’ via the next Google Search box.

The Internet is a hypertext medium – interconnecting a network of dislocated information through a single interface, or, as Vannevar Bush describes, *“a device in which an individual stores all his books, records and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged supplement to his memory.”* Our pursuit and experience of information is that fundamental engagement of two-way communication – a dialogue between human-user and machine.

HOW WE FIND INFORMATION

Speaking of Google, did you know that every time you use Google to search the web, a simple file known as a cookie is updated on your computer, which tracks your search queries, logs the time and location from where the search was made and then relays the information back to Google headquarters in Mountainview California? With over 200 million search queries handled a day, and a cookie set to expire in the year 2038, that is a huge amount of detailed personal information being sent to one private corporation.

There are in fact, many things we do not know about how we ‘find’ information on the Internet and how popular Search Tools work, but what further confuses matters is the fact that each individual Search Tool has its own ways (rules and ethics) of finding and cataloging web sites (also known as indexing). The most popular Search Engines roughly spend a few weeks of each month sending out spiders to crawl the web, going from page to page by following embedded links and taking cache snapshot summaries which are then added to the Search Engine database. If there is no link between a new page and a page which already exists within the database, then the spider has no link to follow, and the page cannot be found. Additionally, Google operates on a policy of popularity, where sites not visited as often as others can be completely passed over and being lower priority, can definitely not expect any of their sub-pages to show up in the Search Engine database.

Google competes along with Teoma/AskJeeves and Yahoo!’s Overture: AltaVista/Alltheweb and Inktomi for the top spot, but even so, can only capture a minority percentage of all the web sites which exist on the Internet. What this leaves us with is a very large part of the Internet completely unreachable through the web’s most popular Search Engines. Other contributors to this ‘invisible web’ are such as: dynamically generated content built on demand without being stored or cached, pages which exist in file formats obscure to Search Engine spiders, and pages excluded from a Search Engine database by policy (such as the exclusion of 113 web sites from the French Google.fr and the German Google.de for their racist content¹ in October 2002), or merely sites which try to manipulate the system so that their site can take the top spot on Google’s search results. The invisible web is extensive to say the least, and in the interest of an uninhibited and free search experience, we can see that users who rely on Search Engines alone are getting a severely limited portion of what is available.

The reality here is that most people are not even aware of any indexing process of web sites. What is more important to the average user is how their search criteria are handled by the Search Tool. When you read *"We've heard users using Google 50 times a day"* published in the New York Times², the [search] tool which accounts for 75% to 85% of all web site referrals³, how queried keywords are matched to web sites and then ranked into a list, sometimes thousands of pages long and so becomes of a much greater concern to everyone. Google and any other Search Engine which employs a method of ranking measured by site popularity, is essentially a self-fulfilling prophecy, allowing sites already popular to grow their rating even more. To Google this method of ranking is their patented and well-publicized PageRankTM technology, and claim that, PageRank *"relies on the uniquely democratic nature of the web,"* but, as Daniel Brandt, creator of Google-watch.org explains, *"In a democracy, every person has one vote. In PageRank rich people get more votes than poor people, or, in web terms, pages with higher PageRank have their votes weighted more than the votes from lower pages."* And this is wholly due to the perpetual recycle of top ten search results retrieved for any given query.

In a way, a search mechanism of this type has aimed to assert a compromise between a one-vote-democracy and a popular representational-democracy; but an ever-widening division between web sites listed on the first page of any given search query and web sites listed everywhere else, is in itself a very strong compromise of integrity. As we are still a long way from systems which can assess the context and intention of our search, I would argue that at least the power of choice be placed into the hands of the user. What we are concerned with here, is authenticity placed in our search results, giving a foundation of truth and validity to the whole process of finding and accessing accurate information content.

Advertising within search results is another area which further confuses the truth. Take for example the Amazon.com sponsored box on Yahoo!'s websearch results, which promise titles generated of your own search criteria – books which may not even exist at the Amazon.com shop itself. Similarly Google itself sells whole keywords as an exclusive commodity to their advertising clients. Some Search Tools fail even to clearly label what is a sponsored advertisement and what is not, something which confuses the authenticity of search results, and so greatly inhibits the use of the internet in general.

This brings us to the simple task of educating the user. Google provides extensive information on how to use all of its features so that one can find exactly what is wanted. But it is when people are enrolling for university courses which teach effective techniques for using the world-wide-web and when Amazon.com lists among hundreds of titles focused on Google alone – a 416 page Dorling Kindersley title – 'DK Google E.Science Encyclopedia' that we begin to question whether this whole process is in fact an easy one at all. The problem here may be merely the lack of freely available information that should help us learn the process? While the link to Google's help guide might be discreetly hidden away somewhere on its home page (try a Google Search for it), most Search Tools fail to recognize the importance of providing this information to make the use of their service easier and much more effective to the user.

WHAT IS NATURAL IS EASY

Perhaps then we should look more at the way we engage ourselves with the experience of information through the Internet. Again so simply identified by Vannevar Bush as a process of access, selection, storage and retrieval of *"the record"* (information content/data). It is also essentially a process of communication, where the man-machine dialogue must work symbiotically together within the framework of the Internet interface, for the mutual goal of data-retrieval and use. For the purpose of this study, it is the human-user experience of which we are most concerned. The easiest and most natural form of communication with which we are well versed is that of verbal language (speech) and body language. But until the day when machines will be capable of accepting this style of ambiguous, non-conforming and most personalized expression, we will have to make do with the 'tap tap tap, click' of today's keyboard and mouse.

In context, the most common forms of data selection we use on the Internet is most definitely the Internet address, also known as url (universal resource locator) and the keyword – from search term to search result. Both are methods of direct selection and forms of textual input severely limited by context⁴ and language – syntax and vocabulary. Think of an Internet keyword as a dynamic link that aims to create a bridge between user intention and the desired target information. If the context of a keyword has been mismatched by either the search tool or perhaps intentionally by a webmaster trying to manipulate the system (or anything else), then the integrity and entire function of that link is compromised.

Additionally, textual input is also very limited by the user's knowledge and application of language. Early users of computer networks were required to keep in memory mathematics, logic operators, assembly level language and sometimes even machine-level binary language, which could at least be sent as input through some form of a command line or terminal interface. Today textual input has been enhanced by such innovations as pre-emptive text, auto-fill, auto-password storage, language translation, syntax and spell checking (see Google) and even some preliminary level of error tolerance in human-user input. This is still a far cry from the ease of use we are accustomed to, with naturally spoken languages like English which we use to communicate with everyday⁵.

Another fundamental method of input takes place through the 'look and point' conceptions of the visually mapped interface which also forms the basis to all modern day WYSIWYG (What You See Is What You Get) styled Internet experience. From Ven Diagrams to the visual-interface innovations of Xerox Park research in Palo Alto California, today we have a model of visual metaphor universally integrated into almost every form of man-machine-data information exchange. There are numerous modern enhancements to

visually accessing information such as – shortcut bookmarks, favorites, spring-loading folders or such as the new exposé window switching innovation of Apple Mac OS X, but even so, usability experts Don Gentner and Jakob Nielsen argue that the “*human interface is stuck. We seem to have settled on the WIMP (windows, icons, menu, pointer) model, and there is very little real innovation in interface design anymore.*”⁶

This is in fact a real concern for the visual metaphor, which aims to bring familiar real world objects online and with them the knowledge of their original use that has been adapted to fulfill a task required by the online experience. Think then of the un-natural metaphors like requiring users to drag an external volume to the desktop trash in an Apple Mac interface, so as to eject the disk, but not erase. Unfortunately visual metaphors can also impede the user with misapplied characteristics or as noted by Gentner and Nielsen, they are “*interactionally cumbersome, overloading users with additional windows and other interface elements necessitated by the metaphor but not by the user’s task.*” Like textual input, an interface that is visual also requires learning; the difficulty of which will determine the success or failure of its use. Using visual metaphors the interface can choose to be invisible, playing no part at all in the learning process, or take the opportunity to be highly visible, which will either aid or make difficult the user experience.

Even so, a visually based information experience promises us a much more natural and intuitive way to navigate, lending itself very well to linking information through visual association, something we do so fluently every day whenever we recall a memory or identify patterns and relationships of real world objects.

EXPERIENCE OF SPACE, IN HYPERSPACE

Online directories such as Google, Yahoo! and About.com all provide data access and selection through a series of well catalogued information content, categorized in just such an associative manner. Unfortunately, like most visually oriented interfaces, online directories also suffer from a linear ‘all stops’ navigation process, where the user must browse through each level of associated pathway points to reach the desired target information. This can be much like browsing the metaphoric files and folders of a generic home-user computer system, though probably less organized, but which can then instead lead to the all too common problem in hyperspace of spatial disorientation. This issue of orientation in hyperspace is also very pertinent to finding and using information. Thus we find the Internet strewn with metaphors and crude adaptations of the real world.

So far I have failed to mention the current interface we most use today, to actually engage ourselves with the Internet and that is the Browser. In truth “*page viewing is truly all they excel at.*” As Bruce Tognazzini and Jakob Nielsen put it, “*Browsers fail to support the actual task of browsing the web ... because the Browser’s capabilities were, for all practical purposes, frozen four years ago.*”⁷ Held back by continued Browser-publisher infighting and an original hypertext markup language (html) limitedly designed for the simple purpose of displaying fixed text and graphic content within a hypertext environment. Today we have grown comfortable with a system which could never see beyond document viewing constructs of the early 1990s – fixed format, unengaging, structure driven, rather than being driven by an immersive content experience worthy of our time.

As the Internet is a hyperspace of dislocated content, any virtual location or geography can only ever be suggested by the user interface. Having one that provides familiar cues from the real world such as ‘forward’, ‘back’, ‘stop’ and ‘go’ are just a few steps in the right direction. In actual fact, today’s Browser lacks so much of the functionality pioneered before there ever was an Internet, that by 2003 we seemed to have most proficiently retrogressed.

INFORMATION STRUCTURE AND ORGANIZATION

In 1945 Vannevar Bush envisioned the ‘Memex’, a system of unfathomable storage capacity using visual display projections and a mechanized method of “*selection by association*” much akin to the workings of the human mind. The system draws dynamic links between information content selected by the user either through association or index. The user’s journey is mapped and stored as individual sequences of cross-referencing hyper ‘trails’ allowing for future re-selection and further exploration, building from stored pathways already traversed. In theory the concept of trails is most outstanding as it accomplishes spatial orientation through a generated virtual geography, mapped uniquely as a personal experience to every individual unique user.

In the 1960s a similar design was also first envisioned by Ted Nelson, a recognized pioneer behind today’s Internet. His system, ‘Xanadu’, had almost limitless inter-referencing between individual files and documents. The key features being the recognition and retainment of the original authorship of information content, while being able to freely navigate and link to any extract version, or sub-component of what was the original.

While Bush’s concept is visionary and Nelson’s possibly revolutionary, I would rather reference a pre-Internet idea that comes a little closer to the system we have in place today which is known as OHS or ‘open hypertext-document system’. It is a system that allows dynamic linking of any individual component of a hypertext document to any other component within the hypertext system, with no additional markup needed in order to identify the component as having any functionality of a link (quite the opposite to html). The possibilities for such a free-form flow of navigation through an ‘open’ style system of information content are immense. It is truly the ultimate ability to freely spider across the information datascape and a step closer to Bush’s own concept of free-form data selection. Such a system could also enhance search possibilities to all hypertext components, independently and automatically

relating content components regardless of format and nature. But this also leaves us with a possibly even greater task of maintaining link relevance to the context of the user intention and a potential failure in an over saturation of information pathways due to the now almost infinite linking possibilities. The structure appears to be much more versatile and advanced, but the problems of assuring link integrity still remain and so for now I would think of OHS as a largely theoretical framework for a truly open hypertext system.

While the exploration into the structural side of data organization is not a key focus of this text, it is important to note that together with the interface, the way in which data is organized and stored, will greatly help to improve the functionality and methods by which the information content can be accessed, selected and used by the user. Additionally with reference to the enormous data organization infrastructure which exists throughout the world (content management, coding platforms, coding standards, database structures, programming languages, etc), innovations and redevelopment of new interfaces which would 'float' above current information structures to provide a more modern, engaging and versatile experience for the future to come, would definitely be a much more practical and less expensive interest to everyone. This topic is decidedly rather large, but in the remaining space I would just like to propose a few suggestions and reference a few key areas that would go a long way to realizing a better experience of the Internet for everyone.

THE PERSONAL EXPERIENCE

Drawing from Bush, the human-user should enjoy a more personalized experience of the Internet. The interface should be one that learns from the success of services such as Amazon.com, where a limited form of selection AI aids the user by putting forth suggestions which cater search results based on previous user trails traversed and also providing a good level of error tolerance through such features as natural language spell checking and associative suggestions of related sites. Such an AI always places the options in the user's hand and will never limit access to information content in any way. By learning a limited level of user intention, we would hope that the accuracy of keyword results and suggested links would be improved.

User input would use a combination of both textual and visual methods (soon oral) making the most of modern enhancements using such a Meta-Search Tool as Kartoo.com creating an immersive and combined visual and textual interface which provides virtual orientation, long term personalized preferences exceeding individual user sessions, stored trails which map out previous searches and the most necessary ability to navigate information directly and associatively. For security and privacy user data must be stored locally on the access device, rather than on the private data servers of Google or Microsoft. Yet such an interface would need to be developed by a vendor capable of widespread standardization (Apple, Microsoft, Redhat) and who can allow transcendence beyond the traditional Browser, web service or desktop application environment; as access to and the engagement of information content now proliferates so totally throughout our entire man-machine information exchange.

Above all else such an interface will actively move forward to educate the user in a suggestive manner that does not impose on the experience itself, as at the end of the day it is the interface itself that should hold the user's experience and needs in the highest regard. (3535 words)

GAUTAMA PAYMENT 2003

1. see searchengineethics.com at searchengineethics.com/02-news/10-google4.htm a situation which took place on the 29th of October 2002 based on a study conducted at the Harvard University's Berkman Center, Cambridge USA. see also a similar article at the SanFrancisco Independent Media Group at indybay.org/news/2003/09/1644738_comment.php which describes a similar incident where Google News was found to have banned all derogatory references to Israelis that could be found through it's service (this included references made by external sites which were found by Google News).

2. see referral made by geeklife.com/Article.aspx?ArticleId=1235&CommentCount=16

3. web site referrals describes the number of visits made to a web site found using a Search Engine. Google is noted to account for 75% of the world-wide-web's web site referrals, while a further 10% are made by AOL and Yahoo!, who both in turn also employ Google as their service's web-search Search Engine. These statistics are originally sourced from google-watch.org/big-bro.html by creator David Brandt but similarly mirrored in articles from Nielson NetRatings Search Engine Ratings at searchenginewatch.com/reports/article.php/2156451 placing Google, Yahoo and AOL together at 86.8% and About.com – websearch.about.com/cs/articles/a/googletraffic_3.htm quoting Google at 70% without including third party search tools like Yahoo! or AOL who also use the Google service, sourced by Mike Banks Valentine September 2003, a Search Engine optimization specialist and creator of WebSite101.com Small Business Ecommerce. It is difficult to get a single figure, but the area defined from 70%-85% is well quoted.

4. context refers to how the application of a textual link matches the desired intention of the human-user.

5. or as my Dad puts it. "Compare asking a search engine for information with asking a human librarian. The internet is "the greatest library in the world, but the librarians are drunk."

6. quote sourced from the outstanding article – 'The Anti-Mac Interface' by Don Gentner and Jakob Nielson, downloaded from acm.org/cacm/AUG96/antimac.htm

7. quote sourced from another innovative article 'Beyond the Browser' by usability experts Bruce Tognazzini and Jakob Nielson; dated March 26th, 2001 and taken from eweek website

STUDY SOURCES

SEARCH TOOLS

University of California Berkeley: Internet Guides, updated September 10th 2002

Teaching Library Internet Workshops by Joe Barker
Prepared for the United States Air Force Office of Scientific Research
<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/>

Main study source for understanding how Internet Search Tools work, what type of search tools are available, with special attention to the term 'invisible web' given to that region unreachable by general purpose popular Search Tools.

Search Tools studied and Analysed include: Google.com, Amazon, Kartoo, Askjeeves, Yahoo!, Teoma, Surfswax, Alltheweb, Altavista, About.com and Vivisimo

BROWSERS AND BEYOND

Beyond the Browser, August 22nd 2003

By Eric Schonfeld, published in Business 2.0
<http://www.business2.com./subscribers/articles/web/print/0,1650,51810,00.html>
document logged: October 24th 2003

Beyond the Browser, March 26th 2001

By Bruce Tognazzini and Jakob Nielson, published in eweek
http://www.eweek.com/print_article/0,3048,a=7834,00.asp
document logged: October 23rd 2003

The Anti-Mac Interface,

By Don Gentner and Jakob Nielson
<http://www.acm.org/cacm/AUG96/antimac.htm>
document logged: October 23rd 2003

HYPERTEXT

The Electronic Labrynth, 1993-2000

By Christopher Keep, Tim McLaughlin, Robin Parmar
Hosted by the University of Virginia: Institute for Advanced Technology in the Humanities
<http://www.iath.virginia.edu/elab/elab.html>
document logged: October 7th 2003

pages sourced include: Ted Nelson and Xanadu, Douglas Engelbert, Visual Representation of Nodes, Hyperbook Authoring Tasks, Writing and Reading Electronic Hypertexts, Structural Paradigms

Useit.com: Jakob Nielson's web site, updated weekly

By Dr Jakob Nielson, Nielson Norman Group of Fremont California USA
<http://www.useit.com>
document logged: October 23rd 2003

includes papers and a weekly column called 'Alertbox' written by Jakob Nielson. papers sourced include: A Short History of Hypertext, The Death of File Systems, Electronic Books – A Bad Idea, the Future of Hypertext

Bootstrap Institute

By Dr Douglas Engelbart, Bootstrap Institute of Fremont California USA
<http://www.bootstrap.org/>
document logged: October 10th 2003

Source for background information on Dr Douglas Engelbart and his work, with special reference to the concept of OHS (open hypertext-document system) and DKR (dynamic knowledge repository).

Systems and Applications

From the World Wide Web Consortium
<http://www.w3.org:80/History/19921103-hypertext/hypertext/Products/Overview.html>
document logged: October 24th 2003

a list of hypertext projects hosted by W3. This was a good potential starting point for investigating different concepts and ideas being developed related to hypertext. Document found via the original proposal made by Tim Berners-Lee, dated March 1989, May 1990 to CERN.

Hypertext Now: Electronic Reading, 2003

By Mark Bernst for EastGate Systems, Watertown Massachusetts USA
http://www.bootstrap.org/chronicle/press/010221/2001_02_21.cover21.html
document logged: October 23rd 2003

understanding the advances in hypertext over traditional printed text. background reading. provides references that lead on to Ted Nelson's 'Literary Machines' and Jakob Nielson's essay 'Electronic Books: A Bad Idea'

Douglas Engelbart: Computer visionary, Wednesday 21st 2001

By Marion Softky, The Almanac
http://www.bootstrap.org/chronicle/press/010221/2001_02_21.cover21.html
document logged: October 7th 2003

biographical look at Dr Douglas Engelbart, his work and vision. background reading. Keyword: co-evolved human systems

As We May Think, July 1945

By Vannevar Bush, originally published inside of The Atlantic Monthly
<http://www.csi.uottawa.ca/~dduchier/misc/vbush/as-we-may-think.txt>
document logged: October 7th 2003

What has really become the most well sourced resource which contributed to the fundamentals of the essay and will go on to contribute to the final year project. Particularly – concept of hypertrails.

Augmenting Human Intellect: A Conceptual Framework, October 1962

Summary Report By Dr Douglas C. Engelbert
Prepared for the United States Air Force Office of Scientific Research

foundation reading, providing a basis for understanding Dr Engelbart's contributions and theories. Interesting concepts of relationship between man and machine. The idea that the human-user must be educated to keep pace in interface innovations. The purpose of the paper studies how we can expand our conceptual capabilities.

A Mathematical Theory of Communication, Jul-Oct 1948 By C.E. Shannon

Reprinted with corrections from the Bell South Technical Journal

foundation reading which provides the basis for looking at the two way dialogue between man and machine from a communications perspective. Vital in taking on mediums, potential information losses, and in defining a conceptual model of communication between man-machine and interface.

Zeros + Ones, 1997

By Sadie Plant, published by Fourth Estate

book served as good background information with particular note to the chapter titled 'cultures' starting at page 45 and quoted at the front of the written text.