The Dark Face of Google

Ippolita Collective

May 2, 2009

Contents

1 Introduction

Google is the best known and most intensively used search engine of the Internet, so much so that it has, over the past few years, managed to become the principal access point to the Web. Surfers world-wide have become completely accustomed to its sober and reassuring interface, and to the unobtrusive yet ubiquitous advertisements adorning its margins. Its users have massively taken to its convenient ancillary services, and its use has become habitual to the point of an automatism: "if you don't know, search it on Google!". People will also 'Google' instead of checking the post-it they put on the fridge, or looking into their diary, search the yellow pages or consult the Encyclopedia Britannica now gathering dust on the shelve in the company of other hard-wired reference books that have become just too awkward to handle.

Google has managed to exploit this craving for simplicity to the tilt. It aspires to be the perfect and ultimate search engine, that is able to understand precisely the query of its users and give them what they require in a fraction of a second. Its elementary interface, which now can be directly personalised and yet remain immediately recognizable in its minimalist style, has become for a truly remarkable - and still increasing - number of users the daily escape route out of the claustrophobia of the electronic screens. It is a whiff of fresh air, a wide open privileged window giving instant access to the fascinating world of the Net at large. See how many people have taken Google as their personal homepage! And yet, beneath such simplicity and user-friendliness hides a colossus, an unbelievably complex and intrusive system out to achieve total control on

the management of information and knowledge of the 'Mare Nostrum' the Web has become. Google is offering scores of free services to satisfy (almost) any desire regarding research and communication: e-mail, chat, news-groups, a file indexation system for your computer, image banks and archiving facility for pictures, videos, books, and many other things. Why? Who is benefiting? Writing a critique of Google looking through its history, and deconstructing the mathematical objects that make it up, present the opportunity to lay bare a specific strategy of cultural domination. This enquiry provides for a more encompassing approach, in order to shed light on the backstage of a large number of applications we have become accustomed to in our everyday use.

This book starts with a brief history of search engines, reviewing the more crucial moments in the ascent of Google. Having survived almost unscathed the collapse of the 'new economy' bubble, Google has managed to establish solid linkages with a number of big multinational players in the IT industry. Its unremitting expansion in all sectors of digital communication has advanced a particularly distinguishable style which has set the trend across the entire cultural universe of the Web.

"Don't Be Evil" is the motto of Google's two founders, Sergei Brin and Larry Page. The two Stanford alumni have, thanks to an arcane management of their own public image, managed to create the notion of a "benign giant", eager to archive our "search intentions" in his humongous data banks. The digital alter ego of million of users appears to be in safe hands at the Mountain View, California central data hub, better known as "the Googleplex". It is there - and in many other data retention centers that have mushroomed all over the planet - that nothing less than the war for total domination of the Web is being hatched. First step is a cheerful embrace of 'abundance capitalism'. Biopolitical control stricto sensu is the name of the game here: working conditions are much better than merely comfortable, the atmosphere is chummy, and bonuses are raining down on the employees, who, happy and thankful, love to be exploited in the Marxian sense and become the best supporters of the company, and proud to be part of its conquering and 'be good' image.

The methods and objectives of Google have a positive outcome for all; the firm's philosophy, based on certified academic excellence, and the commitment to innovation and scientific research is rolled out in ten clear-cut 'truth' on Google's site. These 'Ten Commandments' constitute a Gospel of sorts for the Digital Era, while the 'Google-thought' is propagated by pure and unassuming 'evangelists', all eminent personalities of the world of information and communication technology.

Google's last but not least weapon is its adoption of the collaborative development methods that are the hallmark of the Open Source movement. Here, products are based on Free and Open Source Software (F/OSS), which is not protected by copyright or patents. In doing so, Google reduces the development and improvement costs of its own services, while at the same time obtaining the support of techies, hackers and various kind of other amateurs, and manages to profile itself as champion of free knowledge dissemination, since using its search engine appears to offer a way to access the Web that is both free and meritorious.

However, Brin's and Page's dream of "The Whole Internet into Google", a dream pursued even inside reputed universities is a demagogic concept in the end, an idea that serves to abide by a near-positivist cult of scientific objectivity:

as if in the chaos that is the Web, only a superior technology could vouch for the transparency of the search procedures and of the accuracy of the results - and all this for the best of democracy!

Google auto-proclamation of itself as a 'democratic' instrument is grounded on the allegedly 'democratic' character of the Web itself. Its indexation algorithm, PageRank[TM] is constantly copying digital data in its data-centers, assigning them a value based on the links that are associated with each web-page. In fact, Google interprets a link from page A to page B as a vote in favor of the latter by the former. But it does not stop at counting the number of votes casts or links established by a page, it also weights that page in. If such a page is deemed 'important', its votes will count for more and will thus 'add value' to the pages it links to. PageRank[TM] assigns a higher value to important and high-quality sites by using criteria and filters whose nature is not in the public domain, and which are used every time a Google search is launched. Google's 'democracy' hence shapes the web in accordance with the number and weight of the votes received by each web-page: it is a democracy filtered by technology.

There are a number of secrets hovering around the Colossus of Mountain View, but a fair number of them, as we shall see later, are fake. The mythical aura surrounding Google's technology is due for a large part to a lack of basic information, and of elementary bits of practice that would enable people to engage the technological revolution in cultural terms. To take just one example, the baffling speed at which search results are produced is merely the outcome a drastic pre-selection process whose criteria are anything but transparent. How could otherwise millions of users browse ever so many million pages within Google's databases, all at once and all the time, if there were not some opportune filters in place to limit the scope of the search, for instance to limit it (mostly) to the data couched in the user's own tongue? And if there exist filters which enable a better navigation language-wise, why would one not assume that there exists many others, whose aim is to orient the searchers in specific directions? Google's miracle is in fact grounded in a closed source technology, in copyrighted tradesecrets, and in non-disclosure agreements regarding scientific research discoveries. Search with Google is neither democratic nor transparent as often proclaimed, for the simple reason that it cannot be, both for technical and for economic reasons.

Google's white space where one types the key-words pertaining to one's search is a narrow entrance, an opaque filter controlling and directing the access to information. By virtue of being an informational mediator, a simple search engine reinvent itself as an instrument for the management of knowledge. This enables it to exert a stupendous amount of power, as it attributes itself near-absolute authority in a closed world. Google's cultural model is therefore one of technocratic domination.

The Ippolita Collective's aim with the present volume is to underscore the problem of digital literacy and critical orientation of the larger public on the issue of knowledge management at large. There is a social urgency in achieving this. Internet offers users extraordinary opportunities for self-education, in a way that even exceeds university schooling, especially in such domains as communication and information technology engineering. As the Collective has shown in previous works, the Free Software movement is the most outstanding example of the necessity of continuous self-schooling and of an autonomous management of digital data.

But there is an other side to this coin, a double negative: on one side, the constant debasement of the humanities as model of education, despite the fact that there are many a domain on the web devoted to them, and on the other, the considerable and augmenting cognitive palsy of the average user. Lost in the manifoldness of the data that are available on the Web, we tend to go for the most visible points of reference - Google being merely the most blatant manifestation of this phenomenon - without too much questions asked about what processes are running in the background; we part with our own private data with gay abandon, enraptured as we are by the myriad of decidedly efficient and useful services, which, as still appears to be the rule on the Net, are free to boot.

Ippolita insists on pointing out the lack of scientific popularisation of the technological phenomena our society has been completely overtaken by. Specialised technological manuals can be obtained a plenty, sociologists wax eloquent about the "network society", politicians start imagining an "open society", where the networks would form the substratum of global democracy. But how many dedicated surfers know exactly what an algorithm is? Very few for sure. And most of them will put their trust in the returns provided by PageRank[TM] - an algorithm - that orders in a split second their search results and in doing so directs their experience of the Web. What is needed is the courage to prioritise scientific] popularisation without enclosing oneself in the academic ivory tower. It is absolutely necessary to be able to speak about macro-economics without being an economist, about info-mediation without being a certified specialist in communication studies, about self-education without being a pedagogue, about autonomous management of digital tools without being an elected politician. It is also absolutely necessary to open up the debate about fundamental concepts such as "algorithms", "sensible data", "privacy", "scientific truth", 'communication networks", all of which are far too little discussed by any sort of 'Authority' or 'Garant Institution', which cannot guarantee anything in the end. [NB: This last sentence is really a bit too vague - CHECK AGAINST ITALIAN ORIGINAL!]

The habit to delegate our work to machines unfortunately engenders a generalised lack of interest for the major changes our technological environment is undergoing all the time. These then take place in silence, or are covered-up in a media-induced fog, without their true nature ever being assimilated by the public at large.

The most common attitude encountered amongst those who are faced with the continuous and unfathomable "miracles of modern technology" waver between wonder and frustration. And even mystical adoration sometimes kicks in, as if the digital realm was gilding the world with an esoteric aura, only to be unraveled by the handful of enlightened minds. Yet at the same time, the inability to celebrate in a meaningful way the cult of these new advances makes for deep frustration.

The Ippolita research group was born precisely out of this conviction that change, and the interfacing of various competences and languages might be able to transform what is commonly called the digital revolution into something that enables one to understand our current epoch, with its anomalies, as well as shedding some possible light on our future. Scientific research, humanistic tradition, political engagements and feminist methods are - among many others - the languages we intend to use in these voyages of exploration.

Ippolita's activities show that 'to put 'it' into the commons" is just not good enough, that the level of thinking about IT technology is still limited and the toolbox of most users far too rudimentary. It is therefore necessary to adopt an attitude that is both curious and critical, and to develop competences at the individual level, so as to understand how to have fruitful interaction with the digital world,and to develop tools that are appropriate to the objects themselves. Our hope is to multiply the spaces and places, and the opportunities to be autonomous without succumbing, either to a facile enthusiasm or to a control-induced paranoia. *Just for fun* is our motto! The practice of collaboration is not a panacea that will magically transform every technological novelty in a shared good; collaboration is also not in itself able to thwart technocratic domination in the name of a greater electronic democracy. Going for this represents a superstitious vision of progress, which runs roughshod of individual choices. The synergy between networked subjects, in realms that are perpetually mutating, is not the simple sum of all concerned parties; it requires vision, passion, truthfulness, creativity and an ongoing re-negotiations of the tools, of the methods, and of the objects and objectives involved.

2 Chapter 1. The History of Search (Engines)

2.1 On searches and engines

Search engines today come up as websites enabling one to identify and retrieve information. The mistaken idea according to which the Web and the Internet are one and the same thing is harboured by the majority of users because the web represents for them the simplest , easiest and most immediate access to the Internet. But the Net is in reality far more complex, heterogeneous, and diverse than the Web: it includes chat functions, newsgroups, e-mail, and all possible other features individuals wish 'to put on-line', and this no matter of under what format these informations take shape. To put it differently, the Net is not a static, but a dynamic whole; resources and their mutual interconnections are changing constantly, in a kind of birth transformation and death cycle. The physical connectivity vectors to these resources are also undergoing constant change. Once upon a time there was the modem connected by coper phone wires, and today we live in a world of broadband and optic fiber. And the individuals who shape the Net by projecting their digital alter ego unto it are also perpetually mutating, at least as long as they stay physically alive. Hence, the Net is not the Web; it is a co-evolutionary dynamic built up of the complex interaction between various types of engines ['machines'?]: mechanical machines (personalcomputers, 'pipes', servers, /modems/, etc. aka 'hardware'), biological machines (human individuals aka 'wetware', and signifying machines (shared resources aka 'software').

As we shift through the dark mass of information that is the Net, we need to realise something fundamental, yet uncanny at the same time: the history of search engines is by way much older than the history of the Web.

The Web as we know it is the brainchild of Tim Bernes-Lee, Robert Caillau (*N1), and other European and US scientists. They created the first 'browser' between 1989 and 1991 at the CERN Laboratory in Geneva, together with the 'http' protocol Hyper Text Transfer Protocol, and the 'HTML' language Hyper

Text Mark-up Language for writing and visualising hyper-textual documents, that is documents including 'links' (internal to the document itself, or external, linking to other, separate documents). This new technology represented a marked advance within the Internet, itself a merger between various US academic and military projects.

As the web was still being incubated amongst a number of laboratories and universities across the world, search engines had been already in existence for many years as indexation and retrieval services to the information extant on the Internet.

Obviously, the first search engines could not be looked into on the not yet existing Web: they were rough and straightforward programmes one had to configure and install on one's own computer. These instruments did the indexation of resources by way of protocols such as 'FTP' File Transfer Protocol for file-sharing, and 'Gopher' (an early rival of emergent 'http'), and other systems which have gone out of use today.

1994 saw the 'Webcrawler'(*N2) come into operation as a search engine solely devised for the Web. It was an experimental product developed at the University of Washington. The innovations this search engines was bringing along were truly extraordinary. Besides functioning as a web-site and making it possible to do 'full text' searches (*N3), it also included a tool, the 'spider' that catalogued web-pages automatically. 'Spider' is a software programme fulfilling two functions: it memorises the informations that are on the web-pages it encounters as it navigates through the Web, and it make these accessible to the users of the search engine. (More about this will be discussed details in the next chapters).

As unbelievably innovative as it was in its time, Webcrawler was only able to return simple lists of web addresses as search result together with the mere headline title of the web pages it listed.

In the last months of the year 1994, a new search engine, Lycos, came up that was able to index in a very short time 90% of the pages that were then extant on the World Wide Web (ca. 10 million in all). Lycos' principal innovation was to do without 'full text' systems, and analyse only the first 20 lines of the pages it indexed. It allowed Lycos to give as search result a short synopsis of these pages, abstracted from these first 20 lines.

It was with Excite, coming up in December 1995, that search results for the first time gave a ranking to web pages in accordance with their importance. Introducing an evaluation system that assigned 'weight' to a web page constituted an first, rudimentary, step towards a thematic catalogue: it would at last put an end to interminable lists of disorderly search results. It made a kind of 'initial checking' possible of a 'directory' of web sites, comparable to a classic library system, with an indexation according to subject, language, etc. - but then for web resources.

Apart from that, Excite entered history for another reason: it was the first search engine equipped with tools that were explicitly geared towards a commercial activity. After having acquired Webcrawler, Excite proposed its users personalisation facilities and free mail-boxes, becoming in less than two years one of the Web's most popular portal (1997). Yet, Excite dropped its original business model not long after that, and chose to utilise other firms' search technologies, Google being foremost among them today (*N4).

This bird's eye view of Google's precursors would not be complete without

mentioning what by 1997 had become the best and most popular search engine of all: AltaVista. AltaVista ('the view from above') was based on the findings of a DEC (Digital Equipment Corporation), Palo Alto, California, research group which had succeeded in 1995 to stock all the words in a random Internet HTML page, in a way precise enough to make a very refined search possible. DEC had granted AltaVista the further development of the first data base that could be directly looked up from the World Wide Web. AltaVista's in-house genius was Louis Monier (*N5). Louis Monier clustered rows of computers together, made use of the latest hardware, and worked with the best technologists on the market, to transform his baby into the most common and best loved search engine of its days. AltaVista was also the Net's first multi-lingual search engine, and the first with a technology able to include texts in Chinese, Japanese, and Korean in its searches. It also introduced the 'Babel Fish' automatic translation system, which is still in use today.

By the time of its collapse in 1997 (*N6), AltaVista served 25 million queries a day and received sponsor funding to the tune of US$ 50m a year. It provided a search facility to the users of Yahoo!'s portal, still today the principal competitor of Google in the Web sphere.

2.2 The Birth of Google. Once upon a time there was a garage, then came the University ...

The name Google stems from "googol", a mathematical term to describe a 1 followed by 100 noughts. According to the legend, this was the number of web pages Larry Page and Sergei Brin dreamed of indexing with their new search engine.

Both met in 1995 at Stanford, when Larry Page, then aged 24 and graduate of the University of Michigan, came to Stanford intent on enrolling for a doctorate in computer sciences. Sergei Brin was one of the students assigned to guide newcomers around the campus. Stanford was (and is) renowned as the place to develop highly innovative technological projects. This Californian university is not only a household name for its cutting edge research laboratories, it also enjoys near-organic links with with companies in the information technology (IT) sector, and with keen-eyed venture capitalists ready to sink consequent amounts of cash in the most promising university research. Brin and Page turned out to be both fascinated by the mind boggling fast growth of the Web, and with the concomitant problems related to research and information management. They jointly went for the 'Backrub' project, which got its name from the 'back links' it was meant to detect and map on a given web site. Backrub was re-named Google when it got its own web page in 1997.

The fundamental innovation Google introduced in the search process was to reverse the page indexation procedure: it did no longer show sites according to their degree of 'proximity' with regard to the query, but showed them in a 'correct' order, that is conforming to the users's expectations. The first link provided should then correspond to the 'exact' answer to the question asked, the following ones slowly receding from the core of the search question. (*N7)

It is in this perspective that the famous "I am Feeling Lucky" option came about. Clicking it opens the very first link in the Google results, and is profiled as the indisputably 'right' one.

The algorithm that calculates the importance of a web page, known as PageRank[TM] and allegedly 'invented' by Larry Page, is actually based on statistics of the begin of the 19th Century, and especially the mathematical works of Andrej Andrejevich Markov, who calculated the relative respective weight of nodes within a network (*N8)

At its beginnings Google was only an academic scientific project, where the weight evaluation system was mostly dependant upon the judgments of 'referees' operating within the format of 'peer review'. In theory the method presenting the best guarantees of objectivity s called the 'double blind' reading, as is habitually applied to articles before they are accepted for publication in a scientific review. A contribution is then submitted to two readers who are reputed scholars in their field; they are not not know the identity of the article's author (so as not to influence their judgment). The second 'blind' moment is when the article is being reviewed for publication, and the reviewer is deemed not to know who the two referees have been. To sum up, the more positively a scientific article has been received by fellow scientists (who are supposed to be of an independent mind), the more the article is deemed to be important and worth consideration. Page adopts this approach in his research domain, and applies the theory that states that the number of links to a web page is a way to evaluate the value of this page, and in a certain sense, its quality. We will later go into detail as to how this passage from 'quantity' of returned information correlates with the 'quality' of the results that are expected by the user (*N9).

But this criterion is not sufficient in itself to establish quality, since links are not equal and do not represent the same value; or to be more precise: the static value of a link needs to be correlated with the dynamic value of its trajectory, since the Web is an environment (mathematically speaking, a graph) where not all trajectories have the same value: there are varying 'trajectory values' depending upon the 'weight' of the various nodes. And actually, to pursue further the metaphor relating to the scientific/ academic review process of scientific articles, not all reviews carry the same weight. A positive advice by reviewers less prestigious, or worse, by reviewers not very much liked within the scientific community, can be detrimental to the article being submitted as too many insufficiently authoritative reviews undermine the credibility of a publication. Hence, sites that are linked by sites that are themselves extensively refered to, are according to Page more important than others more weakly referenced ones. In this way, a trajectory (i.e. a link) that originates from a very popular site carries much more weight than one coming from a relatively unknown page. This is how a link from page A to a page B is interpreted in the way of a scientific referral whose weight is directly in proportion to the reputation of the reviewer furnishing that link (it should be noted, however, that Brin & Page explicitly talk in terms of 'vote' and 'democracy' in this regard). The authority of the scientific reviewer becomes the measure of a site's reputation.

Google's web pages evaluation, known as PageRanking[TM], is thus elaborated on the basis of a 'public' referral system which is allegedly the equivalent of the way the 'scientific community' (*N10) is operating, only not limited to scientists but including all the surfers of the World Wide Web.

Today, the organisational workings of the 'scientific community' and the issue of data-referencing in general have become crucial problems: in a context of 'information overflow' (*N11), especially on the Web, it has become

increasingly difficult to estimate not only the importance of information, but also its trustworthiness, the more so since the very principle of peer review has been questioned by scientists themselves in the meanwhile (*N12). Amongst the more interesting alternative options are rankings based on the number of publications [???], networks of publications available under copyleft, and 'open access' projects, which include also research in the domain of the humanities, like Hyperjournal (*N13).

This was the background when Page launched his 'spider' web-exploring programme in March 1996 in order to test the page ranking algorithm he has developed.

The spider-based search engine of the two talented Stanford students became an instant hit amongst their peers and more senior researchers alike, gaining a wider and extraordinary popularity in the process. However, the bandwidth usage generated by the search engine quickly became a headache for Stanford's system administrators. Also, owners of indexed sites had some qualms about the intellectual property rights pertaining to their content, and were besides not best pleased by the fact that the Google's ranking system ran roughshod of more established evaluation systems, such as prizes and honorary mentions in favor of the number and quality of links (i.e. popularity) a page was able to garner around it: Google considers only the relational economy of sites expressed in terms of links, and nothing else. "Spider couldn't care less about the content of a page".

Hence, the value of a search result must be based on the weight of the links between two pages, and not on some arbitrary classification enforced by the terms of the search. This breakthrough turned out to be the key to Google's subsequent success in the years to come: search results would in future no longer be fixed once and for all, but would vary dynamically in accordance with the page's position within the Web as a whole.

2.3 Google.com or how ads (discreetly) entered the pages...

Page and Brin went on developing and testing Google for eigthteen months, making use of free tools provided by the Free and Open Source Software (F/OSS) community (*N14), and of the GNU/Linux operating system. This enabled them to build up a system that is both modulable and scalable to an extremely large extent, which can be augmented and tweaked even while being fully in use. This modular structure constitutes to-day the basis of Google's data center, the 'Googleplex'(*N15), and makes possible the maintenance, upgrade, changes and addition of features and software, without the need to ever interrupt the service.

By the middle of 1998, Google attended to something like 10.000 queries a day, and the in-house array of servers Page and Brin had piled up in their rented room was on the verge of collapse. Finding funds vastly in excess to what usually is allocated to academic research became therefore a matter of some urgency.

The story has it that Google's exit from the university is due to a chance encounter with Andy Bechtolstein, one of the founders of Sun Microsystems and a talented old hand in the realm of IT. He became Google's maiden investor to the tune of one lakh US Dollars (100.000 in Indian English -TR ;-)

The birth of Google as a commercial enterprise went together with its first hires, needed for further developments and maintenance of the data center.

Among them was Greg Silverstein, now the CTO. Right from the beginnings, Google's data center took the shape of a starkly redundant system, where data are copied and stored in several places, so as to minimize any risk of data loss (a move that amounts to print the currency of search). Its most important feature is the possibility to add or to remove modules at any given time so as to boost the efficiency of the system. An other major trump card, as befit university hackers, was Brin's and Page's habit to recycle and tweak second hand hardware and make extensive use of F/OSS. Their limited financial resources enabled them to evolve what would become the core of their business model: nimble modularity at all levels. The Google-ranger[???]'s modularity means that it can scale up and down according to need and availability. No need to reprogram the system when new resources, whether hard- wet- or software, is added: the highly dynamic structure integrates the new modules, even if they are stands alone.

Google's formally opened its offices on September 7, 1998 in Menlo Park, California. As the story goes, Larry Brin opened the doors with a remote, since the offices were located in a garage, a friend of theirs had subleted to the firm. A Spartan office-cum-garage then, but one featuring some not to be spurned comfort components: a washing machine, a dryer, and a spa. Right from start, Google's company philosophy is about making employees' life very cushy indeed.

By January 1999, Google left the Stanford Campus for good. The official statement reads: "Google research project has now become Google Inc. Our aim is to give to the world searches of a far higher quality than what exist today, and going for a private company appears to be the best avenue to achieve this ambition. We have started to hire people and to configure more servers in order to make our system scalable (we are ordering servers 21 pieces at a time!). We have also started to launch our spider more frequently, and our results are now not only as fast, they have also a much better actualisation rate. We employ the most talented people, and through them we obtain the latest and most performing Web technologies". Brin and Page then went on for a few more lines to talk about the ten best reasons to come work for Google, quoting tech features, stock options, free drinks and snacks, and the satisfaction coming from millions of people "going to use and enjoy your software".

The years 1998 and 1999 would see all search engines and other popular sites world-wide in the grip of the 'portal syndrome', a narrow obsession with developing sites that would attract and retain visitors at all costs on the site by providing ever more services, ads, and personalisation gizmo's. Google contrariwise remained the only web instrument without ads and additional features. It was to remain a search engine pure and simple, but for that also the best, the fastest, and the one without commercial ties-up whatsoever.

But the firm could not survive purely on the money given by Bechtolsheim without generating any substantial profit while at the same time pursuing its research on identifying and organising information. Displaying a remarkable aptitude at talking the language of high finance, while constantly emphasising their commitment to research, Brin and Page then managed to reach an agreement with California's two topmost venture capital firms, which astonishingly assented in co-financing together one and the same company. A totally unique occurrence, seeing two giant venture capital institutions agreeing to share risks and profits of a single business proposition. On June 7, 1999, Google was able to announce that Sequoia Capital and Kleiner Perkins Caufield Byers had granted it US$ 2,5 crore in finance capital [N*16] (1 crore = 10 million, in Indian English

-TR ;-).

While one PhD thesis after the other saw the light at Google Inc., its two researchers-CEOs were looking for avenues to commercialise one way or the other the mass of indexed data. As a start they tried to sell their search service to portals by profiling themselves as OEM (Original Equipment Manufacturer [*N17], [I leave the original authors' explanation out, since it's rather confused - in the French version at least] but this was not very successful. But on the other hand, the business model that appeared to be more compatible to the new firm was one of direct advertisement, integrated within the search engine itself, and working by way of doing the count of the number of visitors who access the sites through commercial advertising links. This business model, called CPT, Cost per Thousand Impressions [*N18], has a structure that is as little intrusive as possible for the user. It is not based on flashy advertisement banners, but relies on discreet, yet very carefully selected links that appear above the search results. As these links are in a different font and color than the search results proper, they tend not to be perceived as too disturbing to the user's search activities.

2.4 Self-Service Ads, or Beyond the Dot Com Crash...

A business model based on simple sponsored links appearing alongside search results does not make very much sense in terms of profit generation: at this stage, Google's long term commercial strategy was in need of a qualitative jump. So the presidents came together in search commercially more promising solutions and came across Goto, a company founded by Bill Gross (*N19), now owned by Ouverture/Yahoo.

Goto's business model was based on mixing real search results with sponsored returns, and billing advertisers only if users actually clicked on their web-address, a format known in the trade as CPC, Cost per Click.

Compared to previous methods, this was particularly innovative. Sponsored links would only appear if they were functional to the user's actual search query, thus maximising the likelihood of a transaction to take place in the form of a click-thru to the commercial site. Google tried to reach an agreement with Goto, but its CEO's refusal forced it to seek an alternative, similar solution in-house. At that time, portals (think Excite, Lycos, infosite, AltaVista and Yahoo) were all using the CPM format, and CPC was something of a repressed wish. This tends to show that if you're not able to buy up a superior, mission critical technology from someone else, you'll have to develop it yourself /in an autonomous fashion/.

March 2000 saw the implosion of the NASDAQ bubble, sinking in its wake all pipe-dreams of the 'Dot Com' sphere. With them went also the CPM model, with its illusion of an unlimited cash flow thru its myriads of ad banners "with millions of eyeballs" each. However, these banners were most of the time totally out of context on sites that had nothing to do with the advertiser's line of business. Google faced at that stage the dire need to look very closely at its cost/earning accounts, and urgently find a way to make search technology acquire financial value.

The response came with AdWords, which saw the light in October 1999. Adwords functions as a sort of advertisement self-service, where commercial parties could chose the search key-words most likely to be associated with their

own sites. AdWords was Google's application to put Goto's 'keywords-based advertisement' into effect.

Google hence not only survived the Dot Com bust, it also was able, thanks to being a not - yet - publicly traded private company, to make good use of the opportunity to fish right and left for talent beached by all the other 'dot coms' gone belly up. By mid-2000, Google was answering 18 million queries a day and its document index contained 1 billion unique items. Six months later, queries had reached the 60 million mark.

2.5 Style, Form, and Services Overflow ...

For Google begin 2000 meant most of its competitors had gone South, and the times were ripe for a new round of innovations, starting with proposing users a bevy of new services [*N20]

Every new service constitute a piece of a complex, constantly re(di)fined mosaic, branching out to each and every domain of IT. At the time of writing [end 2006/ begin 2007 -TR], Google offered 17 different types of searches in data-banks of images, blogs, notes, books, maps, videos, financial services, etc. But it can also search and retrieve documents into a user's own computer. And there are many more services, present and to come. Two are specifically geared to application development and new projects under elaboration in Google's labs. There also 6 communication enabling services: Google Mail, VoIP telephony, instant messaging, discussion groups ('Google-groups'), picture-sharing, and translation services. Three services are for mobile devices (GSMs, PDAs, etc.). And finally, there is another service about software suggested by Google. And the number of services keeps adding up...

Even the most dull-witted user can easily grasp the reach and power of these instruments. By now, it is possible to key in a postal address or a phone number, and Google will instantly disclose all you need to know in order to contact a person or localise an object. One can also save one's search preferences, making the repeated use of the search engine a breathtakingly smooth experience. A typing error in the search query is promptly corrected by a highly advanced spell-checker, which is also able to 'learn' incrementally along the search process.

In 2001 Google launched "Google Images", a dedicated search engine which in just a few weeks became the most sought after resource for do-it-yourself graphic production and has become one of the Web's biggest image banks. At the same time, Google also bought up Deja.com, and hence the Usenet archives, which constitute, with over 650 million posts on various newsgroups, a kind of "historic memory" of the Internet in its pre-World Wide Web days, when such discussion groups were the life-blood of the Net. By April 2001, Usenet got re-christened "Google-groups" with a new, pleasant interface making it easy to follow the more elaborate, edgy discussion threads.

¿From 2001 onwards new services followed each other in quick succession, or were upgraded without any apparent economic purpose or immediate financial return, as if Brin and page were thrilled to show that a sheer inexhaustible data retention center could also bring about next to any technological feat one could dream about. The most illustrious instance of this is probably "Google Map" [Google Earth?] a behemoth repository mapping the Earth in detail, with some maps of the Moon, Mars and the Oceanic depths -TR thrown in for good measure. Google Earth is a freely downloadable software suite enabling you to

visualise by way of satellite images parts, or details, or at least a photographic rendering of any surface of the Globe. And the "Google Directory" is home to the contents of Dmoz.com. a collective of human agents organised within a collaborative and decentralised system of 'open publishing', that has been present on the Google home page for ages, but now sports an increasingly sophisticated graphic design.

"Google News" saw the light in 2005, making Google's humongous databases available to journalistic work. GMail started the same year, offering each user 1 gigabyte of personal storage. Beta-launched on invitation-only basis, it immediately created a network of personal linkages completely internal to Planet Google. Privacy nay-sayers were promptly silenced with the somewhat cranky argument that GMail was "an outstanding product", that "its advantages far out-weight the doubts it may raise", and that "it's bound to get ever better with time". Any user of GMail is however liable to be controlled by Google in its use of the service, since the enormous storage space made available is likely to incite her to leave all her mail messages on its servers. And since usage of the service was spreading by way on invites already registered users could freely extent in her circle, Google obtained crucial information on individual networks of friends and acquaintances. With other words, the archetype of an intrusive feature geared towards 'data-mining'.

And then came the "Google Scholar" project, a universities-oriented search engine still in 'beta-testing' mode [2006], which enables to retrieve academic literature, but also articles submitted to reviews, working papers, M.A. and PhD theses, university publications, reprints, table of contents, bibliographies, reports, and reviews published across all sectors of scientific/ academic research. And then we have Google Library, whose ambition is to make available on line *all* books in digital format, by going into agreements with libraries and even interested publishers world-wide, and scan publications into e-books. Only the Google data center could make into reality the dream of a global digital library, accessible from Google pages. But this dream is meeting fierce opposition from the side of a large part of US publishers who are members of the AAP (American Association of Publishers). They fear a meltdown of their profits. In 2005 AAP demanded the digitization of those works still under copyright to be frozen for six month, pending further, and comprehensive, explanation from Google's on its 'Library' project. Yet, despite appearances, and copyright owners opposition, this initiative of Google does not have the free circulation of knowledge as its aim. It is more about a shift in the monopoly to information, which in this scheme would be transferred from a handful of publishers to the one and only Google. Like in all dreams, there is a flaw: a solitary, private entity, named Google, is going to decide what constitutes the stock of collective information, by making the same available through proprietary formats. The Open Content Alliance was started in direct reaction to this project, and it is supported by the Internet Archive, a non-profit, and by Yahoo! Its objective is to make as much material as possible totally accessible, through open formats.

Parallel to its opening new services, Google showed a remarkable ability to milk the relational economy to the max, made possible by a keen utilisation of the commercial data it indexes.

AdSense, launched in 2004, offers site owners the possibility to host certain commercial links on their site, as suggested by Google on basis of the site's subject and particular keywords. Revenues accruing from such links are shared

between Google and the owners of the participating sites. The innovation there lies in monetizing the trust the site's users network put in it. Google is then no longer on the Google site only, but everywhere where the Google 'window' is welcomed, and that unobtrusive little space promises to be always full of accurate and interesting data, such as befits Google, even if these data bits are now commercial suggestions. AdSense is thus factually the materialisation of a "Google network", a specific network meant to cross-link users data with their interrelationships for the benefit of advertisers. According to Google, AdSense is the network of "sites and products partnering with Google to put targeted AdWords advertisements on a site or a product[*N21]." Obviously the AdSense system is also part and parcel of the "Google Network".

And obviously, once you have put such a network in place, revenue must be extracted. We are still in 2005, and Google now experiments with a 'rerun' of the CPM model on the AdSense platform, following a 'site-by-site targeting' model. Advertising will now again 'pay for eyeballs', but this time not according to number of clicks on their banners but as a package deal sold through an auction process. Advertisers are able to choose in detail the profile of their prospective viewers: language, geographical area, issues of interest, etc. But moreover, such views will only happen within the 'Google network'. This appeals mostly to those who want to sell a brand rather than a product, i.e. those vendors favouring indirect marketing strategies. Here, 'brand awareness' is the name of the game, rather than selling specific products to key-word selected potential buyers such as is the case with the CPC advertising model.

This virtuous, or hellish, circle linking up the value management of its own immaterial products with the organisation of the labour force, and the framework of project development, is perfectly attuned to the modular building blocks system upon which the entrepreneurial philosophy of the firm Google is based. An endless growth is the precondition for the system not to flounder. The number of users searching with Google and hence trust their data unto it must increase ceaselessly in order for the advertisers peddling their wares in the "Google network" to keep growing alongside. There must be a continuous launch of new services, of new machines to keep track of it all, of new employees to maintain, improve, and invent them, of new users to make use of them, and of new advertisers to extract a profit from, and, and, and ...] Every new 'piece'of the system is being introduced as a new module, in an endless cycle: ever growing stockpiles of data, brains, users, and of their respective data, increasing quality of the handling of these data, in the dealing with employees, in the interaction with users and the management of their data archived in Google's data centers. And this always under the imperative of speed and further development.

Brin and Page don't hide where their ambitions lie. "Why would we let our employees start their own firms only to buy them up later on when we can pay them to stay with us, and do what they would have done in any case?" The "Googleplex" [*N22], Google's operational Head Quarters in Mountain View, California, is a kind of university campus where people are pampered all the time. Employees are even given one day off a week to work on their own projects, which are then shown to the "Google Duos", who offer both money and the support of the firm to the most promising talents, as reward for their efforts.

2.6 Google, the Good Giant, goes IPO ...

"Don't be evil" or you can do anything you want provided you're not naughty: thus is the motto of Google's "capitalism with a human face"[*N23]. But already, quite a number of cracks are showing up in this 'being Good' PR image: lawsuits galore, suggestions of fraud, sites being blacked-out, etc.[*N24] ...

In 2002 Google had 1000 employees on its payroll and owned in excess of 10.000 computers [servers?]. Its service indexed over 4 billion documents and its net profits (somewhat reluctantly disclosed) amounted to close to US$ 185 million. Given such a size, investors were starting to demand more transparency, more control, and a more credible business profile. It's allright to have two brilliant - if eccentric - engineers at the helm, but please hire also a general manager with a proven development track record!

After a few less than felicitous get togethers and some intemperate public statements, the role of CEO of Google Inc. finally devolved to Eric Schmidt (who was already a top dog at Sun Microsytems and then Novell). The two young prodigies keep taking pot shot decisions but this strategic managerial move soon proved to be a sound economic choice. Schmidt's arrival actually coincided with the first semester that the firm was in the black, demonstrating herewith that it has succeeded in making its products billable.

Page and Brin had postponed as long as they could the moment their company needed to go public, as they feared they would be forced to go on record regarding their business perspectives and profit expectations and that this would make their life less fun. It would also have made Google a much more open book, and presented its competitors on the market with sticks to beat it with.

But after the introduction of AdSense in 2004, and despite Page's pronouncements to the effect that "Google is not your run of the mill company, and has no intention to become one", the colossus became to all intent and purposes precisely that: an all-American publicly traded company.

Just before the IPO, Yahoo! and other competitors lodged scores of complaints against Google, claiming copyrights and patents infringements with the aim to ruin the firm's reputation even before it has sold its first share.

Wall Street was then on the verge of lowering the initial floor price for the bid in view of the encountered difficulties, but Brin and Page managed to bury the biggest lawsuit, the one with Yahoo! by paying Filo and Yang a compensation in Google shares and settling the different regarding patents. Upon which the duo, against the Stock exchange's best advice, proceeded with the IPO, in the midst of August, and with a US$ 20 reduction of the share price.

Yet within a day of trading, Google shares lifted from their US$ 85 launch price to US$ 100, leveraging a cool US$ 1,5 billion of paper profit in the process. One year later Google shares were quoted at US$ 400, or a 300% increase in value. Google Inc. appeared to be surfing the wave in a marvellous world where nobody is bad, everybody wins, and evil simply does not occur. Granted, with such figures even a small downturn in share prices means millions of Dollars going up in smoke, as happened in March 2006 when Google lost seven percentage points. Google is now a giant amongst the giants on the world's stock markets, and if it ever sneezes, many risk catching a cold with it.

2.7 Google Inc. or the Monopoly on Search

In October 2004, Brin and Page were flying their company jet when they learned that AOL (America On Line, the biggest US Internet access provider) had just closed a deal with Yahoo! to incorporate its search engine into their service. The youthful entrepreneurs immediately ordered a change of course, flew to London, and managed to prevail on AOL to shred the contract they just had signed and opt for a sweetheart deal Google, to the tune of a cool 50 million US$. It's not exactly what you would call the gentle and open approach you would expect from the "good giant", but hey, business is business, even for the two nice guys research scientists from Mountain View!

In the meanwhile, Google's profits have grown by a multiplier of 4000 in the course of a mere 5 years, making it the closest direct competitor to Microsoft and Yahoo!, and this not only in terms of stock market capitalisation, but foremost in terms of popularity and hence of cultural domination of the consumer's mind. Millions of users are now using Google as their starting page when they go on the web. And they trust the results they get through the tools developed in Mountain View. Today Google's name is uttered in the same breath as the Web or even the Internet. The Californian search engine scores best when it comes to milk the relational network of its users and extract every cent possible out of millions small advertisers, so much so that for 2005, available data suggest an income in the range of 6 Billion Dollars on advertising products (whereas estimates for Yahoo!'s similar activities amount to US$ 4,6 bn).

The first swamp Google got bogged in had to do with complaints that its searches were conflicting with the (US) legislation on trademarks. Symptomatic were the cases of Geico and American Blind & Wallpaper Factory [& More ?] vs Google[*N25]. In both cases the complainants alleged that Google's AdWords service was illegally selling trademarked name-words. The tricky question was whether complainants could prevent Google from making appear their competitor's links when users would query on terms like 'geico', 'american blind', or 'american wallpaper'. Would a court follow that argument, then Google's and its partners' would face a severe drop in their revenues, since any owner of a trademark could deny its use by AdWords, and sue Google if it ever did. In France, luxury goods firm Louis Vuiton went to court on this and won. Google's answer is that if any tort occurs, it is the responsibility of announcers themselves and not of Google, since its role is merely that of a neutral carrier, and that besides, "attempts to limit the sale of trademarked terms amount to a denial of the freedom of expression". Sounds like making good sense - for Google at least. [? intp. by TR]

however, Mountain View's giant itself falls foul on the freedom of expression issue it argued against complaining firms, where it breaches the trust many users have given it in a matter that constitutes one of the most important sources of revenues. Google has always shielded behind the argument that the actualisation process of of its search algorithms and the objectivity of the workings of its machine were proof that query returns were beyond any kind of manipulation. But then, just before the American Blind case went to court, it had decided to withdraw a number of AdWords that had been purchased by Oceana, an activist group [*N26]. Oceana's 'mistake' had been to publish an environmentally motivated critique of the operations of Royal Caribbean Cruise Lines, itself a major Google investor. This could be retrieved when searching for the

'AdWorded' terms 'cruise vacation' or 'cruise ship', keywords users would normally use to look for information about cruise holidays or associated activities. Google's official statement was that being a neutral medium, it could not condone any propaganda campaign deemed to be detrimental to the good name of other enterprises. Obviously, in such cases, freedom of expression is no longer a paramount concern.

To make things even weirder, on the very day the San Jose District Court was in session on the American Blind case, Google's search results in that very district were mysteriously at variance with results obtained in any other part of the world! For the first attested time, Google was caught manipulating search results with an other aim than to return "the best possible answer to a query". The fact that the court ruled in favor of Google in the Geico case (which was analogous to the American Blind one) does little to detract from this unsavoury episode.

The most embarrassing and best known case till now pertains to Google entrance into the Chinese market. In order to penetrate this fast-growing [potentially immense market, google for the first time, publicly abode by a demand for censorship, making sites deemed illegal by Beijing authorities inaccessible to searches from out the Chinese territory. A Harvard study in 2002 had already shown that Google was blacking out 113 sites in its French and German language versions (Google.DE and Google.FR). Google confirmed the facts, but argued that these pages had only be withdrawn on request of local government agencies and police authorities, and only after a careful analysis of their contents. Many sites were racism-oriented, others were informed by religious fanaticism. Someone then raked up a controversy, stating that Google's much vaunted transparency was crumbling and that users should be made aware of the existence of a 'hidden censorship'. Others countered that Google was not to blame, but rather the law system in particular jurisdictions where you could get sued merely for the providing a link to an incriminated site on your page. In such cases, it is natural that Google chooses to avoid legal consequences by withdrawing links after assessing the risks on individual basis. It should noted, while we are at it, that the issue of the 'right to link' is going to be a major bone of contention within the issue of digital liberties at large: who decides what is legitimate censorship? An umpteenth 'Authority'? Or an international body? Or will it be 'might is right'? In a market economy, that amounts to the right of the party that pays the most, or carries most weight. Or will local, usually religious, fundamentalists have the last word, who black-mail with reprisals every time a 'subversive' site runs foul of their particular world-view? This problem is as far-reaching as the issue of freedom of expression itself, and obviously cannot be resolved in a court room. Emotions ran high in the Chinese case, because the censorship bid came from a government. Yet Brin and Page were too focused on the potential of a market representing a quarter of the world population to backtrack, despite this massive scale-up of the issue at stake.

For Google, the world will fast become a gigantic index in which a perfect correlation will obtain between digital resources and ambient reality. Each and every index will become computable by an algorithm and presented as a search result in the most convenient manner. And Google will be in pole position to be the instrument that shall maintain that index.

But, quite aside from the obvious observation that digital and real worlds do not necessarily coincide, even they are very much intertwined, and that not even

from a technical point of view. The perfect algorithm simply does not exist. It is simply not possible to retrieve *all* informations that exits on-line. Also, nothing that is in the technological domain can be considered really neutral, especially not if it pertains to real-world data of on line individuals.

Stemming from the partnership that are likely to be entered upon, and of the technological convergence coming every day nearer, a new direction appears to emerge, and Goog;e's 'vision' is forcing it upon us as the one and only access point, and management and mediation of digital data. Google's dystopia as Big Brother wannabee becomes more precise, and is bot dangerous and fascinating, as every historic power struggle: the Web is the new stage for a fierce competition to establish the new standard of communication. A standard that, paradoxically, is "personalised", with offers and services that are geared towards the users' individual needs and tastes. for a few years now, the keyword has been "mass personalisation". An oxymoron for sure, but one that comes loaded with the importance of the game, and which represents a paradigm shift, away from mass production consumerism towards a personalised one, sold to us as "freedom of choice". As for us, beyond rhetorical platitudes, we could find a response to this by simply making different choices: the question is not whether or not to use Google and its services, but to choose other ways to put our personal information on the Internet, and to learn how to link them up in a new fashion, making for more innovative and interesting trajectories for each one of us [*N28].

Since a number of years, Google has been learning to its own costs (and those of its users, of course) that innocence does not really belong to this world, and even less to the world of business, that total goodness amounts to stupidity in general, and more particularly so for a firm whose main goal is to make a profit, and that finally, neutrality is a very uphill road when war is raging between competing search engines. And at this juncture it may be recalled that those nations that are traditionally neutral, like Switzerland, are also traditionally militarised to the core. And so we can see which kind of 'good' weapons Google has been using to achieve the status of a world class phenomenon.

3 Chapter 2 BeGoogle!

3.1 Google's Brain-drain or the War for Control of the Web

"I'll ... this Google ... Eric Schmidt is a ... and I'll bury him alive, like I did with other ... like him!". Thus foulmouthed Microsoft's CEO Steve Balmer when he learned in May 2005 that Google had just headhunted Kai-Fu Lee, a high ranking employee of his, and key-man of 'Redmond' for China. Kai-Fu was the one who had developed MSN Search (engine) for the 100 million Chinese Microsoft users. Balmer's expletives were of course targeted at his opposite number at Google, a former Sun Microsystems and Novell honcho, firms also that Microsystem had battled with before, both on the market and in court. Kai-Fu Lee was boss of the MS research lab near Shanghai.

Microsoft immediately started a court case against its former employee and Google specifically, accusing Kai-Fu Lee of violating extremely confidential contractual agreements existing between the Redmond and Mountain View rivals.

Microsoft lawyers argued that Kai-Fu Lee, as executive director, must be in the know of MS industrial and trade secrets, and would not hesitate to put these technologies and the social network and economic know-how he had accrued at MS to use to bolster the profits of the competitor's firm. This contentious personage didn't come cheaply by the way. His entry 'salary' amounted to US$ 2,5 million, with 20.000 Google shares as a side perq. Exorbitant figures which give some idea of the wager at stake - and we were not only talking about the Chinese market.

The lawsuit between the two giants was finally settled out of court in December 2005 - with just one month left before the case was to come up. The particulars of the deal are completely confidential. May be large amounts of money have changed hands, or may be Microsoft managed to force Kai-Fu Lee to keep mum about anything he knew from his previous employment.

This story is merely one of the most curious and emblematic illustrations of a trend that had become noticeable for a few years now: Kai-Fu Lee was actually only the umpteenth senior employee that had switched to Google, "the firm that looks more and more like Microsoft" as Bill Gates had loudly complained. Bill himself was left in a cleft shtick as he faced the nasty choice of either diabolising the two student prodigies - reinforcing thereby their image as his 'kind and generous' opponents in the world of IT - or to pretend that they did not really matter and were not very much worth of attention as competitors .

The truth is that Bill Gates knew all too well how much a switch-over of managers means for a firm's core business, especially in the IT sector: Microsoft had often enough made use of the same trick against its own competitors. The commercial tactic consisting in headhunting key personnel of rival firms in order to tap their industrial secrets and their production and resources management know-how has always been part and parcel of industrial competition. But in the era of the information economy, the practice has become markedly more prevalent, and more diffuse.

So this management choice of Brin's and Page's clearly indicates what Google's ultimate aims are: to become the Web's most comprehensive and customizable platform, by adapting all its services to the singular needs of each of its users, and this together with maintaining an immense reservoir of information. To put it simply, Google is pushing full speed ahead to catalogue every type of digital information, ranging from websites to discussion groups, picture galleries, e-mails, blogs, and whatever you can think of, without any limit in sight. This amounts to open war with Microsoft, whose Internet 'Explorer' browser, MSN portal, and its Hotmail e-mail service, makes it after all, and for the time being, Google's principal foe [competitor].

The overlap between the domains of interest of both firms is growing by the day: both aspire to be the one and only medium to access whichever digital activity. Microsoft has achieved predominance by imposing its Windows operating system, its Office software suite and its Explorer browser as the current computing standard both at work and at home. On its side, Google has been profiling itself as the global number one mediator of web services, especially with regard to search, its core business, offered in all possible formats, but also with particular ancillary services such as e-mail ('GMail'). At the risk of simplification, one could say that Microsoft has been for years in a dominant position thanks to products that pertain to services, whereas Google is now seeking dominance through services running on products.

The outcome of this competition, hence, is dependant on users' choices and on the future standards Google wants to impose. Developing certain web programmes intended to funnel requests for services only through the browser amounts to deny a market to those who have always invested heavily in products and in creating new operating software architecture. [French text unclear here, I guess the gist is: Google's going to literally vaporize all 'static' M$ products by going full tilt for the 'Internet in the clouds' paradigm, cf. next sentence -TR]. The same holds true for /markets in/ the economy at large: there is a shift from a wholesaler/mass market approach (Microsoft), trying to sell licenses of one and the same product or service, to a completely customised one, where products can be downloaded from the web.

3.2 Long tails in the Net. Google vs. Microsoft in the economy of search

Google's second line of argument is based on the key point John Batelle made in his numerous writings [*N2]: the ascent of the 'economy of search'. In his essay "The Second Search", Batelle, who is a journalist and counted amongst the founders of WIRED magazine, argues that the future of on-line commerce lies with personalised searches paid by the users [themselves?]. Google, which is sitting on top of the largest data-bank of 'search intentions' by users, finds itself in the most advantageous position to make this possible, thanks to its very finely ramified network, made up on one side by a famously efficient advertising platform (AdWords) and on the other, by a bank of advertisers (AdSense) now good for several millions of websites. Google's wager is that it will be able to satisfy any wish the users may express through their search query, by providing new services geared towards 'consumerism at the individual level'. Each and every user/customer will hit exactly what she/he wants, the product that is precisely geared to her/his needs. The best known of these 'mass personalised' online services is the one offered by Amazon.com, which is well on its way to make far more money out of selling books or Cd's one at a time to individual customers than to pile up hundreds or even thousand of copies of a best seller. The numerous customers buying not particularly well-selling books online constitute a myriad of 'events' infrequently occurring in themselves, and sometimes even only once. To be able to satisfy nevertheless such 'personalised searches' is the secret of Amazon.com's distribution power. It would be impossible for a traditional book-seller, whose operational model is stacked on shops, stocks, and limited orders, to have the ease of delivery of million of titles at once Amazon.com has: most of its revenues has to come from novelties and best-sellers. Selling one copy of a book to a single customer is not profitable for a traditional bookshop, but is is for Amazon.com, which capitalizes on the 'economy of search' of the 'online marketplace'.

This type of market is called 'long tail' in new economic parlance [*N3]. The theory of 'long tails' goes at least back to 'Pareto's distribution' [*N4], where "there a few events that have a high occurrence, whereas many have a low one". Statistically such distribution is represented by a hyperbole graph where the 'long tail' is made up of a myriad of events that are pretty much insignificant in themselves, but which taken together, represent a considerable sum. Mathematically speaking, a 'long tail' distribution follows the pattern of what is called "power's law" [*N5].

The 'winning strategy' in a long tail market hence is not to lower prices on the most popular products, but to have a wider range of offerings. This makes it possible to sell 'searchlight products' while selling few items at a time, but out of a very large range of different products.

Commercially speaking, it turns out that the highest sales occur in the realm of small transactions. The largest part of sales on the net is a long tail phenomenon. Google makes turnover by selling cheap advertisements to millions of users with text ads, not by selling a lot of advertising space in one go to a few big firms for a hefty fee.

Batelle takes interest in the application of search into not yet explored markets. In the case of Google, the enormous amount of data that is available in order to make searches possible is what has made the milking of the 'long tail' possible. In the domain of e-commerce, long tails have three consequences: first, thanks to the Internet, it becomes possible for non-so-frequently asked products to collectively represent a larger market than the one commanded by the small number of articles that do enjoy large sales; second, the Internet favors the proliferation of sellers - and of markets (such as is illustrated by the auction site eBay); and thirdly, thanks to search, the shift from traditional, mass market to that of niches becomes a realistic scenario.

This last tendency finds its origin in the spontaneous emergence of groups of like-minded people, something occurring on a large scale in networks. On the Internet, even the most important groups by number are not necessarily made up of homogeneous masses of individual people, but rather of colourful communities of users banding together because of a shared passion, or a common interest or goal. The opposition between niche and mass therefore is not very relevant to the identification of the segment of the market that is aimed at. From a commercial point of view, this leads to the creation of e-commerce sites for products attractive only to a very specific type of potential customers, who would never have constituted a profitable market outside online distribution. Take for instance typically geeky tee-shirts, or watches giving 'binary' time, flashy computer boxes or other must-have items targeted at the techie crowd. The amplitude of the supply makes good for the narrowness of the demand, which is spread over a very extensive range of highly personalised products. An interesting article by Charles H. Ferguson [*N6] points out that in such a scenario, it is most likely that Google and Microsoft will confront each other for real for the control [monopoly?] of indexing, searching, and data-mining, and this over the full spectrum of digital services and devices. Microsoft is now massively investing in web services: in November 2004 it launched a beta version of a search engine that would answer queries made in everyday language, and return answers that would be personalised according to geographical location of the user; in February 2005, this MSN Search engine was improved further [*N7]. with MSN Search, it becomes possible to check out Encarta, Microsoft's multimedia cyclopedia. But for the time being, browsing is limited to two hours, with a little watch window telling you how much time remains ... Microsoft has thus decided to develop its own web search system on PCs, without resorting to Google, despite the fact that the latter has been for years now #1 in the search business (with Yahoo! as sole serious competitor).

Taken as a whole, it would appear that the markets that are linked to the economy of search are much larger than the existing markets for search services as such. Microsoft is undoubtedly lagging behind in this area, but the firm from

Redmond might wel unleash its trademark savage strategies, which would be difficult for Google to counter. It could for instance take a loss on investments, integrate its search engine to its Explorer browser and offer the package for free, or start a price war on advertisements and so starve its competitor of liquidity. And in the meanwhile, the new Windows Vista operating system developed in Redmond is supposed to offer innovative search options [looks like fat chance...;-) -TR]. Also take note that Microsoft was lagging very much behind Netscape (the first web browser that was freely down-loadable) in its time , and yet Explorer managed to displace and dispatch it - and not really because it was so much better! But if Microsoft indeed has a long experience of the market and also very deep pockets, Google has not a bad hand either. It is the very incarnation of the young, emergent enterprise, it has built up a reputation as a firm that is committed to research and technical excellence, it preaches the gospel of speed with regard to users' search satisfaction and does so with nifty and sober interfaces, in one word it imposes itself by simply being technically the best search engine around. In the battle for control of the Web Google appears to have a slight advantage. However, one should not forget that Microsoft's range of activity is without par since it covers not only the Web but the whole gamut of information technologies, from tools like the Windows operating system or the MS Office suite, to contents like Encarta, and hi-end research platforms like dotNet, etc. Given the wager at stake - basically the access to any kind of digital piece of information, and the profits deriving from it - peaceful cohabitation between the two giants looks unlikely. Google is still in the race for now - but for how long? [MR/FCG: " but (G) won't be able to stand up very long"]

3.3 The War of Standards

Let's follow up on Ferguson's argument: the story starting now is a war of standards. Three actors are in the game, for now: Google, Yahoo!, and Microsoft. The industry of search, as Batelle pointed out also, is growing at a fast pace. Emerging technologies, or the ones that are currently under consolidation - think broadband enabled audio & video streaming for instance, or VoIP telephony (e.g. Skype, or Google's GTalk), or instant messaging - all are generating Himalayas of data still waiting for proper indexation. And proper 'usabilitization' for the full spectrum of new electronic vectors like pals, gsm's, audio-video devices, satellite navigators, etc - all these being interlinked for the satisfaction of users - but all milked in the end as supports for intrusive advertising. In order for these tools to be compatible with all kinds of different systems and which each other, new standards will be necessary, and their introduction /in the market/ is unlikely to be a painless process. What causes a war of standards is the fact that technology markets demand common languages in order to organise an ever-increasing complexity. The value of information lies in its distribution; but it is easier to spread around real tokens [? analog stuff?] than audio, or worse still, video documents: the heavier the data-load, the more powerful 'pipes' it requires and the more demands in puts on the way in which the information is managed [French text somewhat unclear here]. Traditionally, legal ownership of a crucial technology standard has always been the source of very comfortable revenues [*N8]. It has indeed happened that the adoption of an open, non proprietary standard - such as the 'http' protocol -

created a situation that is beneficial to all parties. But often, the dominant solutions are not qualitatively the best, as "it is often more important to be able to rely on a winning marketing strategy". This being said, there are a number of trends emerging regarding the winners. They usually sell platforms that work everywhere irrespective of the hardware, like for instance Microsoft's operating systems, and this as opposed to the closely integrated hardware and software solutions offered bay Apple or Sun Microsystems. Winning architectures are proprietary and difficult to reduplicate, yet they are at the same time very 'open', that is, they propose publicly accessible interfaces so that can be developed further by independent programmers, and in the end by the users themselves. In doing so, the architecture in question is able both to penetrate all markets and at the same time create a situation of attraction and 'lock-in'. Or to put it differently, it pulls the users towards a specific architecture, and once in, it becomes next to impossible to switch to a competing system without incurring great difficulties and huge expenses [*N9]. The aim: impose a closed standard, and obtain a monopoly.

A very clear illustration of the battle for hegemony in the domain of standards is the challenge Skype and GTalk are throwing at each other. For the time being, Skype enjoys a position of near-monopoly on VoIP for domestic use. Yet it is possible that it has not well estimated the time needed by development communities to assimilate [embrace?] open technologies. Till not so long ago, Skype was the only solution that really worked for anyone wanting to put phone calls through the Internet - even if the person was technically clueless. Skype's proprietary technologies, however, could well be overtaken by GTalk, which is entirely based on F/OSS (and especially on the 'Jabber' communication protocol), all offered with development libraries under copyleft licenses, something that attracts a lot of creative energies in the project as coders vie to increase the punch of Google's VoIP network. In which case the adoption of F/OSS would prove to be the winning strategy and erode Skype's dominance. Of course, Skype could then choose to make its own codes public in order to tilt the ante back in its favor. The choice between adoption of proprietary technologies and platforms, closing access - but keeping public the development interfaces, and going for openness, is therefore of paramount importance in the strategy of control of the Web, and of the economy of search in general.

But access to the economy of search is already closed, or 'locked-in', as economists would say: that any new entrant, or 'start-up', could ever compete with Yahoo! or Google in the indexation of billions of web-pages is clearly unthinkable. Even if such a firm had a better algorithm for its spider, the investments needed would be prohibitive. Yet there are a lot of side-aspects, especially with regard to the interface between various search systems, which lend themselves for a bevy of 'mission critical', yet affordable innovations. This is for instance the case with 'libraries', small pieces of software which make it possible to link up heterogeneous systems together and function as 'translators' between systems, languages, and search engines. Together with integration methodologies between various arrangements and ways to share data and search results, they represent areas that could be developed by individual, independent researchers rather than by large companies [*N10]. We will later into more details into the issue of interfaces and libraries.

For now, it is important to note that none of the players in this game is in a position of absolute dominance, something we can be thankful for. Imagine

what the situation would be if a complete monopoly of search, by whatever private actor, existed by virtue of its factual imposition of one standard. Obviously, the first problem to arise would be the issue of privacy: who would own the indexed data on which searches would take place, reaping humongous profits in the process ? Moreover, since it is already possible to tap into quite an unbelievable amount of information just by typing the name of an individual in the Google search window, and since in a near future the quality and quantity of such information will not only be greatly increased, but even further augmented by the possibility to cross-search among heterogeneous data, one can assume that the control exercised on individuals will become ever more suffocating and totalitarian: it will cross-aggregate confidential data with medical records, phone conversations, e-mails, pictures, videos, blogs and opinion pieces, and even ADN info. Google would then become the premiere access point to the digital panopticon [*N11]. So let's us now have a look at the weapons that are deployed in this very real war for the control of the networks.

So let's us now have a look at the weapons that are deployed in this very real war for the control of the networks.)

3.4 Exhibit # 1: the Googleplex, or nimble Capitalism at work

The customary panegyric of Google tells with glee the saga of the firm's impressive growth, which saw Brin and Page move from their dorm in Stanford to the Menlo Park garage sublet by a friend to the newly founded Google Inc., and then on to the offices on University Avenue, Palo Alto, to culminate with them taking possession of the Mountain View, California Googleplex, where the firms is now headquartered. Between 1998 and 2000 the pair fleshed up their formula through a company philosophy based on innovation, creativity, and sacrifice. The sort of commitment you see in science, but then applied to commerce, is their key to success. Right from its beginnings, the Googleplex attracted droves of eager collaborators: here they find back the environment typical of an American campus, where study, commitment, sport and games mesh in a whole. The idea being that if a comfortable and relaxing environment stimulates the students' creativity, it obviously will also boost the productivity of workers. The spirit of fraternity like at the university, the academic elite mentality of working with total commitment for the very best results, all seem to form the bread and butter of stories concerning the Googleplex. The rumor goes that large swatches of the car parks are earmarked twice a week for roller-skates hockey. masses of gadgets and gizmo's cramp the offices, with multi-coloured lava-lamps being favourite. A chummy easy-going atmosphere has been made the norm, with 'Larry and 'Sergei' chairing the weekly 'TGIF' (Thanks God It's Friday) meetings with dozens of employees assembling in an open space created by pushing the office furniture aside.

Right from the beginning, Such an informal atmosphere was intended to build-up a community spirit and encourage the sharing of ideas. Indeed, the Googleplex looks like a place to celebrate one's passion for research rather than an everyday workplace - what it of course is. But not an ordinary workplace, despite its by now gigantic dimensions. Granted, the 'campus style' organisation of work had been widespread in the USA for the past thirty years at least: Microsoft and Apple, to take but two examples, have always worked that way.

Silicon Valley's mythology is replete with stories illustrating the paramount role assigned to creativity, and stress the importance of collaboration between co-workers. No better boost to productivity than happy employees happy to work for a company whose objectives they hold equal to their own, as opposed to workers oppressed by a rigid hierarchy, enslaved by rules and inflexible schedules in a dreadful environment. Perhaps that the novelty of the Googleplex then resides in having promoted, deliberately and right from the beginning, the idea of a 'different' 'new-fashioned', 'made for the best /brains/' place of work. You can't come in the Googleplex unless you know someone working there. And once in, photography is forbidden - in theory at least. As if to shield it from the mean world outside, full of finance sharks and other malevolent IT predators out to pry on the talents of the 'Googleboys'.

Everybody wants to work in the Googleplex. An unofficial survey of all the fantasies out there would for sure list: company work-out room, swimming pool, free food in the four staff restaurants (one of which vegetarian) free drinks and snacks everywhere (who needs vending machines? Google picks up the tab!), volleyball and basket ball fields and other outdoor sport facilities, buggies to dart from one building in the campus to the next , and so on. But that's all nothing compared to the kiddies day- care, kindergarten, and primary school run by the company - free of costs of course, and [don't forget the dental surgery, actually a mobile dental lab in a van, also completely free of costs. In a country like the USA, where education and health care come with a huge price tag, and leave so many people out, these are truly unbelievable perks.

The work spaces also are spectacular, the dreams of a IT ueber-geek come true. 21' LCD plasma monitors are standard all over the place. Toys and games galore (life-size Star Wars figures, a riot of hi-tech gadgets, etc.). Fluo-coloured lava lamps as the omnipresent accessoires du jour. Googleplex is a dreamland, a green workspace, with flexible hours, and where everything appears possible. In one word, the Googleplex radiates the Google philosophy, and unfolds the Google life-style - of course there is a collection of all imaginable must-have enterprise gadgets, one can shop on a dedicated merchandising site. Most are, as befits gadgets, totally useless and/ or superfluous, but all contribute to impart a sense of pride boosting the feeling of being part of the firm. Gone are the dull sweaters and jackets with the firm's logo embossed: Googleplex's conditioning is much more nimble than that! Google is anyway not the only firm taking that road but it has gone furthest along it. Sure, Apple and Yahoo! have been providing a catalogue of firm-related goodies, ranging form a complete line of attires to all kinds of hi-tech accessories, MP3 readers and USB keys, all in the colours or with the logos and motto's of the firm. But Google's trade shop is much more versatile: from foibles for new-borns to 'Google Minium' the system that enables you to index your data 'just like Google'.

The Googleplex is abundance capitalism in the informational era made 'flesh' [*N12]: all the world's information made available to all, for free. The era of scarcity is over. The plenty and availability of goods (in this case, of information) is simply limitless. But let us not forget that, in the end, next to all this plenty comes from advertising, itself mostly text links-based. All the remainder is free - as in free lunch. But B/by the way: not everything works to perfection [*N13]. Mikie Moritz, a Welshman who also has a stake in Yahoo!, and John Doerr, who's also a major investor with Sun Microsystems and Netscape, are amongst the most influential 'apostles' of this abundance capitalism [*N13, *N14].

3.5 Exhibit #2: perfecting the strategy of accumulation

The reason for the 'flight of brains' towards the Googleplex, as was hinted at at the beginning of this chapter, now becomes clearer. For an average employee in the IT industry, and even more for an 'independent' (read: precarious) IT worker, a job at Google's is a dream come true. In this branch of industry, exploited precarious workers are more and more numerous. An exemplary figure is that of the independent coder who labours on personal 'projects', maybe by publishing them on sourceforge.net or Slashdot.org [?], and who offers her/his competences on the market without any kind of status or union protection, nor any of these other guarantees that look like prehistoric remnants in our times of total flexibility. But at the Googleplex not only she/he will get all these, but, she/he will be even able to devote these famous 20projects while being paid for it, and invited to do ever more his best.

To find life boring would be rather difficult amidst games of volley or basket ball, dogs running all over the company campus and its corridors, and casual meetings around a table tennis table. Since it is difficult to find new recruits that would be able to improve further the prevailing atmosphere, Google is resorting to rather novel hiring techniques. The most curious being probably the riddle they splashed in July 2004 on gigantic white billboards along the highway and in a few mass transit stations in Cambridge, Massachusetts. They bore the following text:

First two digit prime in consecutive digits of e.com

The natural logarithm meant here is the number 7427466391. Going to the address http://www.7427466391.com one found a Google IP address asking one to complete a sequence of numbers. Upon finding the number 5966290435 one had to follow instructions, using that number as password to enter into a section of the http://www.linux.org (!) site from where one was redirected again to Google's site http://www.google.com/labjobs where one could downloads one's CV. If you'd managed to resolve all these riddles, chance was you'd make good Google material! But Google does not only attract the best techies, hackers and assorted ueber-geeks. Quickly enough, the highest rewarded It managers got wind of the Google's career potential and vie with each other to enter into the company.

Google's accumulation strategy of both data to conduct searches and of networked computers to stock all these data and back-ups follows closely its brain accumulation equivalent. Semantic [?] machines, electronic machines, biologic [?? bionic? ;-)] machines, all accumulate at the Googleplex, in order to nurture a life-style, or maybe even a kind of cult of excellence, incarnated in an 'evangelist'.

The person representing best the company's style, the one who is called Google's 'Chief Evangelist', is not one of the many youngsters around, but an true sea dog of the Web: Vinton G. Cerf, who invented the TCP/IP protocol together with Robert Kahn. The particulars of his arrival at Google are worth a little diversion: In February 2005 Google let know that ICANN, the supervisory body of the Internet's domain names and numbers, had allowed it to set shop in the domain registry trade. By next September, Google announced that Vinton Cerf had become "senior vice-president and Internet Chief Evangelist for Google, with the mission to identify new technologies and strategic applications for the firm, on the Internet and on any other platform" [*N16]. Till then, Vinton Cerf

was, among many other occupations, ICANN's board senior adviser. But unlike CEO Eric Schmidt's and other top-level management who were headhunted at Google's competitors, Vint Cerf's hire looks more like a PR stunt. Amusing as it may sound, he is unlikely to be a regular at the Googleplex...

3.6 Exhibit #3 Image is all, but a little bit of 'philosophy' doesn't harm either ...

Google's public image cannot be reduced to its site and sleek interface, whose simplicity and speed has earned the firm so much success [*N17]. In cannot be reduced either to the Googleplex, the Valhalla of technology's ueber-gifted. And its not only about 'Being Good', and yet make a lot of money, by combining brazen commercial strategies with the suggestion of Google being part of academic culture and F/OSS communities through its incentives and funding.

But where its image also, and mostly, resides, is in the enterprise's 'philosophy', which is expressed, in clear and easy to understand language, as the 'Google thought' [Google-think?]. The word 'philosophy', however, might be slightly misplaced, since this 'thought' is not really informed by the love of knowledge and transparency. But anyway, one can find online the Ten Commandments that guide the actions of the 'Good Giant' Google. The first sentence of this gospel already sets the tune: "Never settle for the best", as indeed, according to Larry page, Google's ultimate goal is the "perfect search engine", which would understand exactly "what you mean and give back exactly what you want". Thus Google does not strive to reach the greatest number of people possible, it wants to reach all people, satisfy everybody's desires, in one word, bring happiness to Earth. In order to achieve this, it works relentlessly on research and innovation, as is laid down in "The ten things Google has found to be true".

1. "Focus on user and all else will follow". Google's growth was fuelled by word of mouth, and attracted users who were enthusiastic about its performance. This is the exact opposite of aggressive advertisement campaigns. Ads should not jump on users, but present something useful.

2. "Its best to do one thing really, really well". "Google does search. With one of the world's largest research groups focused exclusively on solving search problems, we know what we do well, and how we could do it better. Through continued iteration on difficult problems, we've been able to solve complex issues and provide continuous improvements".

3. "Fast is better than slow". "Google believes in instant gratification. You want answers and you want them right now. Who are we to argue? Google may be the only company in the world whose stated goal is to have users leave its website as quickly as possible". Two major intuitions, and realisations, have enabled Google to arrive at this 'speed': the development and constant amelioration of the PageRank[TM] algorithm, continuously indexing through the networks, en the use of modulable platforms that are interlinked and extremely flexible ('clusters'). Now speaking of speed as the Holy Grail, it might be a timely idea to think a little deeper. sometimes, even in the realm of IT, slow maybe a virtue [*N18] [Makes sense, comes from the country that invented Slow Food ;-) -TR]

4. "Democracy on the web works". "Google works because it relies on the millions of individuals posting websites to determine which other sites offer

content of value". We already know that Google uses PageRank[TM] to evaluate the sites linked to another web page and assign them a value partially based on that of the sites to which they are linked. the representation of this electronic democracy is rather idyllic: Google's index results are allegedly an 'people-based ranking index' based on an algorithm doubling as electoral law which supposedly would enable the users-citizens of the Net to express their confidence/ cast their vote by way of providing links to other pages, and to find the result of this vote regularly expressed through the respective position of favoured web sites. The equation 'link = vote' is rather simplistic and forced, as 'refinements' are constantly being introduced to calculate rankings, by selectively tweaking the value of these votes/links. One could speculate that a link provided by a porn site might weight less than one coming from an university ... In which case one might ask whether academic cultures ranks higher in popularity than porn... what is certain however, is that with the continuous growth of the mass of information, this 'democracy' is bound to expand exponentially.

5. "You dont need to be at your desk to need an answer". "The world is increasingly mobile and unwilling to be constrained to a fixed location. Whether it's through their PDAs, their wireless phones or even their automobiles, people want information to come to them". Flexibility of time and space is an important objective. The convergence of electronic media (TV, radio, phones, Internet ...) towards miniaturised mobile platforms is an unheard of boon for the world's largest supplier of search solutions. As we saw earlier with the 'war of standards', early penetration of future markets is /strategically/ vital, especially for Google, which produces search interfaces, but not the electronic supports on which it could impose its own software (like Microsoft and Apple). Each new device out on the market is therefore a new territory to be conquered.

6. "You can make money without doing evil". "Google is a business. The revenue the company generates is derived from offering its search technology to companies and from the sale of advertising displayed on Google and on other sites across the web". But advertisements are text only, hence not very intrusive. The proposed links are relevant to the search query (AdWords). And users can very easily become advertisers themselves: it's a DIY formula. If you are maintainer of websites , you can even make money on the Google network through AdSense, by putting ads that are relevant to the content of these sites. 'Don't be evil' and 'Don't harm anyone' apparently also means 'Don't advertise those who don't advertise you', and of course guarantee that Pagerank[TM] is not for sale. The trust users put in the correctness of the search returns is Google's major asset and shall not be squandered for the sake of short-term benefits. Its function is to generate indirect, 'second line' incomes, based on advertisements.

7. Theres always more information out there". "Once Google had indexed more of the HTML pages on the Internet than any other search service, our engineers turned their attention to information that was not as readily accessible. Google indeed accumulates a bevy of heterogeneous databases: images, newsgroups posts (Usenet), telephone numbers, postal addresses, financial information, etc. If your aim is to be the world's largest info-mediator, accumulation of data should know no limits!

8. "The need for information crosses all borders". "Though Google is headquartered in California, our mission is to facilitate access to information for the entire world, so we have offices around the globe". An Academic, American

Culture for All. You need to have a grand vision of things: whatever happens, index more and more information, and make it accessible to everyone. 'Localisation' is an essential part of Google's universalism: speakers of be it Korean or hackers' jargon, Hindi, Xhosa, Star Trek's Klingon or even 'Pig Latin', Zulu, Esperanto, Muppet Show's'Bork Bork Bork' - all should have access to a dedicated Google search site. The interfaces languages run into 100+. Google is #1 search engine in over one hundred countries. A very impressive performance, but verging a trickle towards the totalitarian... The whole operation smacks of political correctness and appears respectful of minorities, but reality is that we have to do with a 'super-layer', the surface sheet of the one and only interface, which flattens and homogenises differences, spreading the Mountain View style all over the planet.

9. "You can be serious without a suit". "Google's founders have often stated that the company is not serious about anything but search. They built a company around the idea that work should be challenging and the challenge should be fun.". This injunction aptly resumes the Googleplex, which is organised like a campus in order to maximise profitability. Hence we are told that "There is an emphasis on team achievements and pride in individual accomplishments that contribute to the company's overall success", and that "this highly communicative environment fosters a productivity and camaraderie fueled by the realization that millions of people rely on Google results. Give the proper tools to a group of people who like to make a difference, and they will". Maybe this is the ultimate method to exploit 'creatives', transforming them into enthusiastic supporters of the 'Google experience' at the same time.

10. "Great just isnt good enough". "Always deliver more than expected. Google does not accept being the best as an endpoint, but a starting point. Through innovation and iteration, Google takes something that works well and improves upon it in unexpected ways". Of course, in order to satisfy all the desires of all the world's users, and that ever faster and ever better, one needs to ever push back the point where one's desires are satisfied. One must desire to desire to be the best. Seen in this context, being second is worst than not to exist at all. But as far as we are concerned we'd rather go for the following motto: "Making money, within a firm devoted to excellence, is moral obligation!".

[NB. All the quotes are from Google's 'Corporate information' website at: http://www.google.com/corporate/tenthings.html]

3.7 Exhibit #4 Google and Open Source

Probably Google's most complex weapon is its strategy of co-operation-cum-exploitation of the world of F/OSS. The Google Code initiative (started March 2005) is a token of honour towards the F/OSS community: "we are friends of theirs" say the Google's founders, "because we owe them a lot". The site of Web's most used search engine explains that Google Code is not about promoting the development of applications working on its own APIs (Application Programming Interfaces), since there is already a site devoted to them, but to make F/OSS development tools that are of public interest available to everybody. The first four projects on Google Code were actually programmes created by Google's own engineers to optimise creation, optimisation and debugging of code. The projects affiliated to Google code are also hosted at sourceforge.net and distributed under a BSD 2.0 license (meaning the code may be used both

on F/OSS and on proprietary applications). Moreover, Google has recently promised to make all kind of software available to the F/OSS community, and these are mostly the outcome of the famous 20% of their working time employees are free and encouraged to devote to personal projects.

So it's not a total coincidence that shortly after launching this initiative, Google embarked on a robust recruitment drive of F/OSS developers: the "Summer of Code", a contest of talents with a US$ 4500 prize money to be won. And then came 'Google Earth', and finally, like every power that has achieved to create a distinct life-style of its own, Google materialises a long-cherished dream: http://www.google.com/moon/ ! /Yes, Google's on the Moon!/ /To honour the first landing of Man on the Moon, on July 20, 1969, we have added a few NASA images to the Google Maps interface so that all can pay a visit to our celestial neighbour. Have a nice trip!/ : "After over three decades, we're finally getting ready to go back to the Moon. To help you prepare, and to whet your appetite for exploration, we teamed up with scientists at the NASA Ames Research Center to bring you this collection of lunar maps and charts. This tool is an exciting new way to explore the story of the Apollo missions, still the only time mankind has set foot on another world."

Google's moves, which are those of a typical 'quasi-monopolist' in both its methods and its aims, already have had a direct effect on its competitors. Today, Google is fast on its way to become a giant occupying all spaces of the market; its constant stream of new services choke smaller companies to death, as they are desperately battling to recruit engineers and developers, and live in the constant fear to see their products poached and duplicated.

The continuous launch of new services, coupled to the in-house funding of potential spin-offs by its own work-force, make that Google today factually has closed the market in terms of technological innovation. Indeed, who would risk today financing a Web-based project, knowing full well the risk that in a matter of days, it would be Google that launches it?

Google has managed to represent itself, both to observers and to the average users, as a stalwart of progress. Starting with its search engine, designed in a way to be rapidly and easily understood by its users, it has multiplied ideas and proposals for new services.

With its choice for F/OSS, the relational economy that Google engineered has become a 'world view' that can immediately be adopted as a desireable evolution, towards a 'benign capitalism' as a dispenser of abundance, the kind of 'ethical' economic dispensation that individuals are looking for.

4 Chapter 3 Google Open Source

4.1 Theory and Practice: 'open' is not 'free'

'Free Software' and 'Open Source' are terms, often used as synonyms, referring to code or portions of code. Though both terms often are used to describe the same objects, their perspective is radically different. 'Free Software' is a term coined in the beginning of the eighties of the previous century by Richard Stallman, and is about the absolute freedom it allows the user to use, modify, and improve the software. This liberty has been precisely set out in the famous Four Fundamental Freedoms:

(...) [Follows here, for a few pages, a general description of Free Software / Open Source and its history, Free Software Foundation, etc. I'll translate it later, it's fairly common knowledge stuff, and I need the original English language texts, not always at hand due to failing connectivity... I resume where Google comes into play - p84/5] (...)

The interaction between 'free' methods of development and the net.economy at large would lead, in the years following 2YK to an explosion of the number of 'Open Source' products as well as to heated political debates around software patenting, digital property rights and generally about ethically and politically acceptable norms of 'intellectual property' management.

Google, despite not being a producer of software as such [? means probably Google doesn't produce software for sale, it develops software for its own products and services], was very much involved in the rocky history of F/OSS, was it only because it adopted, like many other dynamic and innovative firms, the F/OSS methods in order to pursue its 'mission'. The contiguity between F/OSS and Google is one of place and of time. A number of important free software projects were seeing the light at Stanford University in 1998, just as Brin and Page were putting the last hand on the first version of their search engine. Think for instance of SND and 'Protege', which were to be extremely successful in their respective digital domains (audio and semantic web).

The Stanford hacker culture (from which in last analysis F/OSS stems from), lending to all students the feel to belong to the same family, it comes as no surprise that the pair, whose formation straddled those very years, was always to have a preference for the GNU/Linux development platform.

Even though there are non-trivial differences between Free Software and Open Source, there are also many common elements and shared viewpoints. For the sake of clarity, we will use here the term /'Open Source'/ 'F/OSS', for Free and/or Open Source Software to design the phenomenon variously embracing Free Software, Open Source software and its manifestations as competitive element in the IT market [French text : competition —?—]

The first characteristic of a F/OSS community consist in adopting working methods that are open to the collaboration of all comers, meaning that it will potentially accept spontaneous input and interaction from any party that is involved in the creation of digital artifacts, be it a coder, a programmer, or even an ordinary user. In the hacker jargon, this approach has been described as the 'bazaar' model and its widespread acceptance stems from the way the Linux kernel was developed in the beginning of the nineties. This project, initiated by Linus Torvalds forms the basis of all GNU/Linux distros (distributions: software suites, often a whole operating system ('OS')).

The new co-operation techniques developed by the digital underground dispatched Brook's infamous law [*N7], which up to now had been the bane of IT projects' development teams. Following Brook's law, which predicates that the number of errors grows exponentially as complexity and lines of codes increase, a project in which thousands of developers participate must end up in a chaos of unstable code and innumerable bugs. Quite on the contrary, the publication of the source code, and the free circulation on the Internet of the documentation, together with the co-operation and spontaneous feedback of an ever growing number of participants, have all enabled F/OSS communities to demonstrate that it was possible to considerably improve the development of digital artifacts, both process- and results-wise. Software developed this way is usually

shipped under a General Public License ('GPL'), leading to 'viral' distribution of products under copyleft.

Despite the fact that the GPL license does not restrict commercial use, it has been often superseded by 'diluted' variants, just as has happened with 'Free Software', where emphasis on freedom was perceived as too pronounced. This is the case with the BSD (Berkeley Software Distribution) license, which does not restrict closure of the codes, and hence impairs viral transmission, as the 'free' code could be augmented with 'non free' portions, resulting in an originally free creation becoming proprietary in the end. Other forms of 'free' licenses also exist these days: MPL (Mozilla Public License) for instance. And more are custom-made for various new F/OSS products as they appear on the market.

This way, the market economy also hosts a sustainable development model, and the developers community is becoming the kernel of a truly 'Open Society' [*N8], often thought as a chimerical Shangri-La. This imaginary posture is not only determined by the moral allegiance inspired by the practice of collaborative development, but also, and actually foremost, by the fact that F/OSS applications are usually superior to proprietary ones, despite (or thanks to ...) the fact that they are often a labour of love.

4.2 The era of the 'Open Source economy': be good and compete...

The arrival of 'Open Source' on the markets has been, according to some observers, the vindication of 'technological convergence', a by now somewhat paradigmatic slogan in IT circles. This convergence means the coming together and synergy build-up between various technologies, which up to now had been separate, and were developed in separate R&D environments.

Within these often extremely rapid transformations, the creation of open standards has merely marked a new phase in the 'war of all against all', also known as 'free trade', with "co-operate on the standards, compete on the solutions!" as motto. This is also IBM's catchword, one of the largest players in this field. When even 'Big Blue' is willing to co-operate, then the cake must be worth it...

Indeed, for many firms, F/OSS solutions have become one of the few ways left to compete successfully against monopolies (and consolidated oligopolies), and to escape classic style competition, which due to ever increasing investment costs is no longer a viable proposition for many smaller companies. But with F/OSS in hand, firms can lower their development costs, and hence the 'price' of their services. Firms have been now familiar for long time with the dynamic advantages of networked development and network partnerships: it is a well-known fact that a network's worth goes up in the square proportion of its nodes [*N9]. The larger the network, even larger the profits /,exponentially so/.

F/OSS would seem to offer a number of interesting guarantees for the development of high added value networks: on one hand it allows the software to remain, in a certain sense, a 'public good' (since it follows an open path of development and benefits from community support); on the other hand it helps reduce the migration, or 'switching'. costs, from one system to the other, especially in the case of switching from proprietary models to 'free', but even more so, in the case of 'legacy' issues (abandoning obsolete platforms). When adopting new technologies, the major expenses reside in the formation of the users,

not so much in the costs of acquiring the technology itself, and certainly not in the case of outstanding software carrying next to no price-tag. But the greatest boon, even though it is difficult to quantify, resides in the /creation of an/ entirely new , attractive image for the firm adopting F/OSS and its products.

The performances and success of F/OSS has led to various attempts to put its format in practice in various other sectors. Such attempts inevitably went along with the use of exalted formulas such as "Open Law', 'Open Science', or even 'Open Society' though the term had been coined by Karl Popper much longer ago. Today, the idea of an 'Open Source Society' has almost become the paradigm for a new epoch, dedicated to the collaborative search of common means to achieve a 'politics of what is feasible'. 'Open Source Society' indeed is meant to consist in an 'open code' dispensation where the possibility to provide input for improvement is freely available to all. When expressed in such terms, one can only agree. Yet one might be surprised by the facility with which such a concept, whose origins are in a very precise, technical, IT context, has been metaphorically 'translated' to philosophical, economic, and societal domains, without very much thoughts being given to modifying or adapt it to the demands of its new usage.

In the branch of the IT industry were it was born and is put into practice, F/OSS, and more specifically, "Open Source" has also meant market competition, battle for the best brains, race for the lowest costs, venture capitalism, and mergers and acquisitions to the tune of billions of US Dollars. We have to do here with large markets where capitalism organises itself in a nimbler, more 'democratic' way. A business dynamism that is no longer bent on submitting the labour force, but to intimately associate workers to the 'mission' of the enterprise, itself increasingly equated with the realisation of individual desires [*N10].

Amidst ever so many firms surfing this wave in the pursuit of various benefits, Google stands again out as the one which sets the tune: "don't be evil", avail yourself of F/OSS, it's free, it's better than proprietary software, and its developers are proud to be part of it. A look at the Googleplex has shown how this strategy of deep penetration in people's everyday life has been refined into a fine art: happy, rewarded, and incentised creative employees, producing far more and much better than oppressed workers.

4.3 Seducing the hackers: autonomy, easy money, free tools.

Google's F/OSS exploitation peaks in 2005, just as the firm's reputation hits low tide due to its competitors' moves and a some murky judicial affairs [*N11].

Even though Google's business model was firmly rooted in IT culture and the practice of scientific excellence, the mere usage of the GNU/Linux operating system to run Google's humongous data center(s) was not enough: a stronger initiative was needed to strengthen further the faith in F/OSS, and focus the attention again amidst a by now disparate mass of free production networks.

Developers could no longer be seduced just by providing a 'authentic h4x0r' version of the site - or a Klingon one. And the intellectually elitist attitude of the in-house academic brains started to wear thin on investors. They expect substantial returns on their investments and are less interested in the cult of excellence, meritocracy, and the attendant academic arrogance, even though

their outcome is an invariably outstanding quality of products. It was therefore unavoidable that the period would come to a close where the two founders friend could jokingly quote shares and do a wager on the stock exchange for US$ 2.718.281.828, being the mathematical constant "e", or to make completely 'crazy' moves, as in August 2005, when declared to have sold 14.159.265 Google shares in order to rake up US$ 4bn in liquidity, without telling the investors nor explaining what they intended to do with that money.

A bold strategic move was called for in order to materialise further Google's aim to invest in research, and to demonstrate that it is possible with such a strategy, to be not only outstanding quality-wise, but also competitive on the markets. This move was to be targeted not so much at the 'average user', but at the 'young brains', with the future, and innovation as goals. Operationally speaking, this meant to create and nurture a community, by giving it tools and means, and by signing agreements with other firms in the same sector. The F/OSS world was to be brought under Google's spell.

Google got seriously into sponsoring new F/OSS communities in October 2005: Oregon State University and Portland State University were [each?] granted US$ 3,5 lakhs (see Chapter 1 ;-) to improve the quality of their F/OSS development projects, spawning new software. Shortly afterwards the "Summer of Code" programme was inaugurated with a splash, PR being made directly on Google's home page (and is still accessible today at http://code.google.com/summerofcode05.html)[¡¡ recommended click! -TR]. The message was loud and clear: the best were to be concretely rewarded. Every coder coming up with a new F/OSS projects or with substantial improvement of an existing one, was to receive US$ 4500. The whole operation was of course meant to be perceived as one big love bum to F/OSS, stressing the fact that there was the strategic ground where innovation was happening. Also, the sympathy of young developers was to be courted by offering them a cash incentive. And finally, Google was seeking to create a real, 'open' style community, which it would sponsor.

More than 400 young developers ended up with a reword. Most were students, and most had made improvements or introduced new features in already existing projects, rather than having developed entirely new software packages. They had added all kinds of features to software suites like Apache, Fedora, Gaim, Inkscape, jabber, KDE, Mozilla, OpenOffice, Python, Samba, Gnome, Mono, Ubuntu - and even Google. Quite some success, especially for the firms that were going to benefit as owners of these projects: IBM, RedHat, LinSpire, Novell, Mozilla.com, sun Microsystems and Hewlett Packard.

A number of these projects, together with those that were developed within the famous 20contribute towards achieving the second goal of the firm's plan to pony up with the F/OSS world: to provide for development tools and means. By 2002 Google was already offering freely downloadable tools on its site. Today, the dedicated page hosts proprietary projects developed by Google teams as well as the winning projects of the Summer of Code which are not linked to Google's own products or services.

The "Code" section of the site presents a number of projects by software developers that are devoted to the most diverse programing languages (Java, C++, Python, etc.). Making development tools available is absolutely essential if you want to foster the creation of software and communities, because the investment is directly linked to the instruments that are necessary for that purpose. Those projects that are developed by Google's own coders are called Google APIs,

and are proprietary libraries to ensure the interface and run Mountain View's /colossus'/ principal services.

A library is a collection of shared subroutines and compiled portions of code that provide services to other, independent programmes needing simplified functions [?]see en.Wikipedia: 'library (computing)'. An eloquent example are the graphic libraries GTK, QT, and FLTK, which make use of the standard visual applications like buttons, menus, icons, making the work of coders easier. Coders will then go to their favourite libraries and will need to write only those lines that are unique to the programme. The libraries will take care of the buttons, the mouse's moves, the inking of shadows, in short of everything we, as users, are accustomed to. Given the fact that the average coder will be less than enthusiastic about doing all this dreary work her or himself, graphic libraries are an essential link between various projects. One one hand, they lend a certain graphic homogeneity to the different applications, and on the other hand they enable coders to concentrate on the real work without losing time creating interfaces.

There are development communities which take care of libraries in order to provide for generic and transversal [cross-over?] tools needed for solving complex tasks (network connexions, communication between applications, word processing, image compression, etc.). Just like a software suite is made in order to reach out to as many users as is possible, a library is there to be used by the maximum number of developers.

Libraries hence allow coders to create software starting from an assemblage of shared resources, which function as 'de facto' standards. Making use of existing libraries while programming means to benefit from a basis that is already very large and complex, uses existing code in the most effective way, and allows for a layering of competences. Libraries therefore represent a strategic asset both in the dynamics of spontaneous F/OSS co-operation as in the relational economy oriented world of 'Open Source'.

Google libraries, the Google APIs run under a proprietary license, hiding to programmers their actual mode of functioning. But that's not all: they also include a special control device, as the developer who downloads libraries for free needs to authenticate her/ himself by way of an identifying code. This enables Google to trace in an invasive manner all moves and all changes that are made subsequently to the use of its APIs. Coders making use of these libraries are allowed to integrate Google search in their site and to know its PageRank[TM] ranking in real time. They can also make use of software that manage advertisements through AdWords, generate dynamic maps of their data with the Google Maps interface, or open a VoIP account for online telephony with GTalk. In one word, they can deploy Google services as they like, making use of the programming language of their choice, and all this under the watchful eye of Mountain View.

The vast diffusion of Google services goes together with the possibility to personalise them down to the minutest detail. It is possible, by writing appropriate XML documents [*N13], to establish bridges between the various Google service. For instance, all elements of Google's home page may be tweaked to one's own requirements, as if it were an application. Same possibilities exist with Google Earth: one can install 3-dimensional surfing on satellite images, or one can highlight geographical areas, or buildings, or weather data. [?]

All these tools, which are intended for those who know how to write code

- at least in one language - are essential to create new combinations of programmes, or simply to use whatever Google makes (at least partially) public in its applications [*N14]. There is even a portal, called googlearthhack.com , where one can find numerous tricks and 'hacks', so that one can do the most unexpected things with the site, for instance merging satellite maps with any other database.

All the facilities offered /to us/ by the Google libraries carry with them two strict rules to be respected: registering and licensing. In order to activate the functions of the Google API, one need to request for a key first, that is an access code, and to mention exactly to which purpose one wishes to employ it. Only then are the APIs activated. The second requirement is the license. These APIs are not under copyleft: they can only be used up to a certain extent: it is mandatory for instance, to have a Google account, as the hunger for gathering more information never stops; moreover, the maps are the exclusive property of Google (or of an third party), an may under no circumstances be altered. And of course, in case of commercial use, an agreement must first be entered into.

The activation code enables Google to retain total control on the /new/ programs that come about by making use of the APIs. Google can block these applications without any reason being given, or it can simply control either the way they access its services, or the usage that is being made of them. All this comes about because the source code is not public and not free, making it impossible to understand the internal working of the libraries.

Besides getting development done free of costs while still keeping it under control, Google has another good reason to foster the creation of communities along this somewhat bizarre formula, which we may call 'quasi-open'. It can be also put to use to compile even more data, do research, and sell statistics.

To welcome and host for free individual developers' projects means also obtaining their trust. Allowing people without restrictions to search the database of ongoing projects amounts to trigger a solid chain of users into existence. Moreover, such a costless incubator of young talent secures the availability of a pool of highly motivated human material whose formation, one of the major cost items in the IT sector, has already been taken of in an autonomous fashion, and in way that is in complete alignment with the style of the firm.

The offer of development tools as a form 'talent scouting' mechanism has been known for long time: it is, for instance, the battle horse of a few robust IT market players such as the Va Software Corporation, which puts extremely powerful computers at the disposal of the F/OSS community for free, together with unlimited bandwidth, memory space, and even technical assistance of a kind that is beyond reach to the most. There are two digital Valhalla's which may claim world-wide fame a a number of project hosted far above that of any competitor: sourceforge.net and freshmeat.net - and both are property of Va Software. So big is the appeal of such portals that even very small projects appearing on their front pages will attract hundreds of unique views. All projects hosted on code.google.com also have a sister page on freshmeat or sourceforge.

Thus, all the ensuing applications will have Google's visibility together with all the services offered by /the/ Va Software /colossus/: discussion forums, mailing lists, debugging tools and machines, control devices such as CVS (Concurrent Versioning System), controlling the versions, editions and changes made to /of/ the code.

It is not difficult to imagine how, with data bases used /for free/ by thou-

sands of coders at its disposal, Va Software can offer an outstanding 'business to business' service to companies that are active in the domain of F/OSS - and not only to those. Its data mining represents a virtual heap of gold in the feverish world of billion Dollars deals. RedHat, Microsoft and many other corporate heavies are among the advertisers and sponsors of sourceforge and freshmeat.

There are many ways to bring F/OSS developers and firms going for F/OSS together. In Italy for example, Sun Microsystem allows you to publish your CV on a Google Map API through its javaopenbusiness.it portal. It is up to the developers to create their own profile, helping thereby Sun Microsystem and Google to sketch a map of Italy's F/OSS resources with the tools they provide.

And so Google can bank on advancement of its products being done by hundreds of users, and this at next to no costs. To which one can add the organisation of talent competitions such as the 'Summer of Code', serving both the development and the advertising of its services. And finally we see extremely dynamic methods of recruitment: Google now even practices video-hiring at http://video.google.com where enthusiastic employees and Sergey Brin himself will tell you all the benefits of working for Mountain View [*N15]

4.4 Hybrid worlds of university and enterprise

With the benefit of hindsight, the coming together of Google and the world of F/OSS would appear to be very much a strategic and calculated move, despite a commonness of origin and purposes regarding the dynamics of collaboration among F/OSS communities which came out of the academic/ scientific scene. The accumulation strategy we discussed earlier is at work even here: Google operates a bit like a black hole, using, even fostering, open codes in order to subsequently suck them in and integrate them into its business. A number of changes Google engineers made to open tools have never been made public for instance. This applies to their server 'GWS (Google Web Server) which is a modified version of Apache, the most widespread F/OSS server of the Web. This amounts purely and simply to availing oneself of the potentials and realisations of the open development formula without sharing developments and improvements afterwards.

An important factor in the relations between Google and the F/OSS world is the fact that it had its origins in Stanford, a university well-known for its capacity to spawn aggressive and competitive, high quality research-backed start-ups. Despite the fact that Stanford did constitute - and still does so - an environment very favourable to F/OSS development projects, the narrow links that exist with venture capitalism make it rather difficult to pursue purely academic excellence once one has left the campus behind.

A small digression on academic research, US style, is needed here to shed light on the intertwined origins of Google, the F/OSS world, and commercial profit-oriented research. On a more general plane, universities in the USA are remarkably intent on capitalising on intellectual creation: the custom is that a university will retain the copyright on the results of all research projects that were developed within its walls. Universities in the United States are historically connected to business, and are often real businesses themselves. University originating patents on invention made by its researchers bring benefits in all senses of the term, besides enhancing the prestige of research centers, their staff and students.

These universities constitute hybrid environments, public and private at the same time. Up to 2002, public universities were not allowed, in theory at least, to patent their inventions, and the same applied to publicly funded private research labs (often at - private - universities). Rights payments impede the free circulation of knowledge in scientific research and makes reproduction, verification and/ or invalidation of experimental results difficult. This was based on the "Experimental Use Defense", a legal principle dating from 1813 that allowed for the free usage of patented technology in experimental research. This jurisprudence was quashed in 2002, in Madey vs. Duke University. John Madey had sued his own university because it made use of a device he had patented to conduct research on free electrons. The Federal Circuit) Court (of Appeals ruled that the "Experimental use Defense" was intended to protect a scientist who is engaged in research in a free and financially uninterested way, but that within universities such activity was obviously no longer to be considered so innocent, since, even in case there was not a direct commercial connection at stake, it still could be considered akin to a 'legitimate business', because it generated funding and benefited the research personnel and the students being educated. And so, any distinction between research for public and research for private goals was made to disappear. [NB interesting article & comments on the case at: http://tinyurl.com/dyd9mc -TR]

Naturally, all projects conducted at Stanford are patented by the university, and this mixture of incentives bestowed on F/OSS projects on one side with a mad run on patents on the other, does not sit well with the ideal, let alone the practice, of 'research for its own sake', such is being trumpeted by Google as its strength and its pride.

The issue of patents becomes even more interesting in the light of the fact that Google's success primarily rests on an algorithm invented by Larry Page in collaboration with Sergei Brin, at a time when they both were still researchers at the Computer Sciences Faculty of Stanford. The algorithm that revolutionised the indexing of the Web is hence the property of Stanford, subjected to a regular patent. In the next chapter(s) we will look into how this prodigy functions, how it manages to return results in less time than any competitor, as if it was able to give each and every user "exactlywhat she/he wants".

5 Chapter 4 Algorithms or Bust!

Google's mind-boggling rate of growth has not at all diminished its reputation as a fast, efficient, exhaustive, and accurate search engine: haven't we all heard the phrase "if it's not on Google, it doesn't exist!", together with "it's faster with Google!". At the core of this success lies, besides elements we have discussed before, the PageRank[TM] algorithm /we mentioned in the introduction/ which steers Google's spider's forays through the Net. Let's now look more closely at what it is, and how it works.

5.1 Algorithms and real life

An algorithm [*N1] is a method to resolve a problem, it is a procedure built up of sequences of simple steps leading to a certain desired result. An algorithm that actually does solve a problems is said to be accurate, and if it does so speedily,

it is also efficient. There are many different types of algorithms, and they are used in the most diverse scientific domains. Yet, algorithms aren't some kind of arcane procedures concerning and known only to a handful of specialists, they are devices that profoundly influence our daily lives, much more so than would appear at first sight.

Take for instance the technique used to tape a television programme: based on algorithms; but so also methods to put a pile of papers in order, or to sequence the stop-overs of a long journey. Within a given time, by going through a number of simple, (re)replicable steps, we make a more or less implicit choice of algorithms that apply to the problem solving issue at hand. 'Simple' in this regard, means foremost unequivocal, readily understandable for who will put the algorithm to work. Seen in this light, a kitchen recipe is an algorithm: "bring three liters water to the boil in a pan, add salt, throw in one pound of rice, cook for twelve minutes and sieve, serve with a sauce to taste", all this is a simple step-by-simple step description of a cooking process, provided the reader is able to interpret correctly elements such as "add salt", and "serve with a sauce to taste".

Algorithms are not necessarily a method to obtain completely detailed results. Some are intended to arrive at acceptable results within a given period of time [French text: 'without concern for the time factor' - which doesn't sound very logical to me -TR]; others arrive at results through as few steps as possible; yet others focus on using as few resources as feasible [*N2].

It should also be stressed /before going deeper into the matter/ that nature itself is full of algorithms. Algorithms really concern us all because they constitute concrete practices meant to achieve a given objective. In the IT domain they are used to solve recurrent problems in software programming, in designing networks, and in building hardware. Since a number of years, due to the increasing importance of network-based reality analysis and interpretation models, many researchers have focused their studies on the construction methods and network trajectories of the data which are the 'viva materia' of algorithms. The 'economy of search' John Batelle writes about [*N3] has become possible thanks to the steady improvement of the algorithms used for information retrieval, developed in order to augment the potential of data discovery and sharing, this with an ever increasing degree of efficiency, speed, accuracy, and security. The instance the general public is the most familiar with is the 'peer-to-peer' ('P2P') phenomenon: instead of setting up humongous databases for accessing videos, sound, texts, software, or any other kind of information in digital format, ever more optimised algorithms are being developed all the time, facilitating the creation of extremely decentralised networks, through which any user can make contact with any other user in order to engage in mutually beneficial exchanges.

5.2 The Strategy of objectivity

The tremendous increase of the quantity and quality of bandwidth, and of memory in our computers, together with rapidly diminishing costs, has enabled us to surf the Internet longer, better, and faster. Just twenty years ago, modems, with just a few hundred bauds (number of 'symbols' transmitted per second) of connectivity, were the preserve of an elite. Today, optic fiber criss-crosses Europe, carrying millions of bytes per second, and is a technology accessible to all. Ten years ago, a fair amount of technical knowledge was required to create

digital content. Today, the easiness of publishing on the World Wide Web, the omnipresence of e-mail, the improvement of all kinds of online collective writing systems, such as blogs, wikis, portals, mailing lists, etc. together with the dwindling costs of registering Internet domains and addresses, have profoundly changed the nature of users: from simple users of information made available to them by IT specialists, they have increasingly become creators of information themselves.

The increase in the quality of connectivity goes together with an exponential augmentation of the quantity of data send over the networks, which, as we have pointed out earlier, entails the introduction of steadily better performing search instruments. The phenomenon that represents this pressing necessity exerts a deep attraction on social scientists, computer science people, ergonomists, designers, specialist in communication, and a host of other experts. On the other hand, the 'informational tsunami' that hits the global networks cannot be interpreted as mere 'networkisation' of societies as we know them, but must be seen as a complex phenomenon needing a completely fresh approach. We therefore believe that such a theoretical endeavour cannot be left to specialists alone, but demand a collective form of elaboration.

If indeed the production of DIY network constitutes an opportunity to link autonomous realms together, we must also realise that the tools of social control embedded in IT technologies represent a formidable apparatus of repression.

The materialisation of this second scenario, most spectacularly exemplified by the Echelon eavesdropping system [*N5], looks unfortunately the most probable, given the steadily growing number of individuals who are giving information away, as opposed to an ever diminishing number of providers of search tools. The access to the information that is produced by this steadily growing number of individuals is managed with an iron hand by people who are both retaining the monopoly of it while at the same time reduce what is a tricky social issue into a mere marketing free-for-all contest where the best algorithm wins.

A search algorithm is a technical tool activating an extremely subtle marketing mechanism, as the user trust that the search returns are not filtered and correspond to choices made by the 'community' of surfers. To sum up, a trust mechanism is triggered into the objectivity of the technology itself, recognised as 'good' because it is free from human individuals' usual idiosyncratic influences and preferences. The 'good' machines, themselves issued from 'objective' science, and 'unbiased' research, will not tell lies since they cannot lie, and in any case don't have any interest in doing so. Reality, however, is very much at variance with this belief, which proves to be a demagogic presumption - the cover for fabulous profits from marketing and control.

Google's case is the most blatant example of this technology-based 'strategy of objectivity'. Its 'good by definition' search engine keeps continuous track of what its users are doing in order to 'profile' their habits and exploits this information by inserting personally targeted and contextualised ads into all their activities (surfing, e-mailing, file handling, etc.). 'Lite' ads for sure, but all pervasive, and even able to generate feedback, so that users can, in the simplest way possible, provide information to vendors, and thus improve the 'commercial suggestions' themselves by expressing choices. This continuous soliciting of users, besides flattering them into thinking that are participants in some vast 'electronic democracy', is in fact the simplest and most cost-effective way to obtain commercially valuable information about the tastes of consumers.

The users' preferences and their ignorance about the mechanism unleashed on them) is what constitutes and reinforces the hegemony of a search engine, since a much visited site can alter its content as consequence of the outcome of its 'commercial suggestions': a smart economic strategy indeed.

Seen from a purely computer science point of view, search engines perform four tasks: retrieving data from the Web (spider); stocking information in appropriate archives (databases); applying the correct algorithm to order data in accordance with the query, and finally, presenting results on an interface in a manner that satisfies the user. The first three tasks each requires a particular type of algorithm: search & retrieval; memorisation & archiving; and query. Google's power, just as Yahoo!'s and other search giants /on the network/ is therefore based on: 1. A 'spider', that is a piece of software that captures content /on the net/ 2. An enormous capacity to stock data on secure carriers, and a lot of backup facilities, to avoid any accidental loss of data. 3. An extremely fast system able to retrieve and order the returns of a query, according to the ranking of the pages. 4. An interface at the user's side to present the returns of the queries requested (Google Desktop and Google Earth, however, are programmes the user must install on her/his machine beforehand).

5.3 Spiders, databases and searches

The spider is an application that is usually developed in the labs of the search engine companies. Its task is to surf web pages from one link to the next while collecting information, such as document format, keywords, page authors, next links, etc. When done with its /exploratory/rounds, the spider software sends all this to the database for archiving /this information/. Additionally, the spider must monitor any changes on the sites visited so as to be able to programme its next visit and stock fresh data. The Google spider, for instance, manages two types of site-scans, one monthly and elaborate, the so-called 'deep crawl', the other daily, 'fresh crawl', for updating purposes. This way, Google's databases are continuously updated /by the spider through its network surfing/. After every 'deep crawl', Google needed a few days to actualise the various indexes and to communicate the new results to all its data-centers. This lag time is known as the "Google Dance": the search returns used to be variable, since they stemmed from different indexes. But Google has altered its cataloguing and updating methods from 2003 onwards, and has also spread them much more in time, resulting in a much less pronounced 'dance': now the search results vary in a dynamic and continuous fashion, and there are no longer periodic 'shake-ups'. In fact, the search returns will even change according to users' surfing behaviour, which is archived and used to 'improve', that is to 'simplify' the identification of the information requested [*N6].

The list of choices the application is working through in order to index a site is what constitutes the true force of the Google algorithm. And while the PageRank[TM] algorithm is patented by Stanford, and is therefore public, later alterations have not been /publicly/ revealed by Google, nor,by the way, by any other search engine company existing at the moment. And the back-up and recovery methods used in the data centers are not being made public either.

Again, from a computer science point of view, a database is merely an archive in digital format: in its simplest, and till now also its most common form, it can be represented as one of more tables which are linked together and which have

enter and exit values: these are called relational databases. A database, just like a classic archive, is organised according to precise rules regarding stocking, extraction and continuous enhancement of the quality of the data /themselves/ (think recovery of damaged data, redundancy avoidance, continuous updating of data acquisition procedures, etc). IT specialists have been studying for decades now the processes of introduction, quality improvement, and search and retrieval within databases. To this end, they have experimented with various approaches and computer languages (hierarchies rankings, network and relational approaches, object oriented programming, etc.). The building up of a database is crucial component of the development of a complex information system such as Google's, as its functionality is entirely dependent on it. In order to obtain a swift retrieval of data, and more generally, an efficient management of the same, it is essential to identify correctly what the exact purpose of the database is (and, in the case of relational databases, the purpose of the tables) which must be defined according to the domains and the relations that link them together. Naturally, it becomes also necessary to allow for approximations, something that is unavoidable when one switches from natural, analog languages to digital data. The itch resides in the secrecy of the methods: as is the case with all proprietary development projects, as opposed to those which are open and free, it is very difficult to find out which algorithms and programmes have been used.

Documents from research centers and universities allow a few glimpses of information on proprietary projects, as far as it has been made public. They contain are some useful tidbits to understand the structure of the computers used and the way search engines are managing data. Just to give an idea of the computing power available today, one finds descriptions of computers which are able to resolve in 0,5 microsecond Internet addresses into the unique bits sequences that serve to index in databases, while executing 9000 spiders 'crawls') at the same time. These systems are able to memorize and analyze 50 million web pages a day [*N7].

The last algorithmic element hiding behind Google's 'simple' facade is the search system, which, starting from a query by the user, is able to find, order, rank and finally return the most pertinent results to the interface.

A number of labs and universities have by now decided to make public their research in this domain, especially regarding answers to problems that have been found, and the various methods used to optimise access speed to the data, questions about the complexity of the systems, and the most interesting instances of parameter selection.

Search engines must indeed be able to provide almost instantaneously the best possible results while at the same time offering the widest range of choice. Google would without doubt appear as the most advanced search engine of the moment: as we will see /in details/ in the next chapter, these extraordinary results cannot but be the outcome of a very 'propitious' form of filtering...

For the time being, suffice to say that the best solution resides in a proper balance between computing power and the quality of the of the search algorithm. You need truly extraordinary archival supports and indexation systems to find the information you are looking for when the mass of data is written in terabytes (1 TB = 1000 gigabytes = 1000 raised to 3 bytes), or even in petabytes (1 PB = 1000 TB [or 1024 TB , Wikipedia's funny...-TR]), and also a remarkable ability to both determine where the information is in the gigantic archive and to calculate the fastest time needed to retrieve it.

And as far as Google's computing capacities are concerned, the Web is full of - not always verifiable nor credible - myths and legends, especially since the firm is not particularly talkative about its technological infrastructure. Certain sources are buzzing about lakhs [See Chapter 1 ;-)] of computers interconnected through thousands of gigantic 'clusters' [sitting on appropriate GNU/Linux distros - French text unclear]; others talk about mega-computers, whose design comes straight out SciFi scenarios: humongous freeze-cooled silo's where a forest of mechanical arms move thousands of hard disks at lightning speed. Both speculations are just as plausible or fanciful, and do not necessarily exclude each other. In any case, it is obvious that the extraordinary flexibility of Google's machines allows for exceptional performances, as long as the system remains 'open' - to continuous in-house improvements, that is.

5.4 From 'Brand Identity' to 'Participative Interface'

Search, archiving and retrieval of data are procedures so complex that understanding them fully requires an amount of knowledge and explanation that are beyond the scope of this book. We will however look in detail a bit further down at some aspect of their functioning. And we should in any case have a closer look at the interface because that is the element which is fronted and managed by Google as representing its core image, whereas algorithm performances and the database architecture are components that remain invisible to the user.

In the case of Google, [T]he interface is mostly the 'blank box'[*N8], the empty window where the user puts down his/her query, or 'search intention' on Google's universal homepage, which is designed in such a way as to exude welcome, reassurance, closeness.

Google's homepage universal functionality stems from it being iterated across 104 languages and dialects, customizable in 113 different countries as per today [2007, -TR]. In some [all?] of those, the interaction model remains the same and unifies all search behaviours into one single, homogeneous format.

Going to Google's homepage, one first notices a linear interface with key elements, each with a very specific and universally recognizable function. This frame will accept search queries of various nature and complexity, from simple key words (e.g. 'Ippolita') to more complex assemblage of words in brackets (e.g. "authors collective"), or it will enable to narrow the search more precisely: to a particular site, or a specific language, or a particular domain, or only to documents in a specified format, and so forth, depending on the level of specificity one is aiming at. We can take it as an example of a successful interface, in so far that it manages to fulfil the ambitious goal of assigning a positive value to an otherwise white space in a page. The interface presents itself without any adornment, almost empty, or rather, filled with one empty element, the 'blank box', which reassures the user, and induces her/ him into activity, warding of loss of attention and her/ him leaving the site due either to an absence of handles [i.e. something to hold on], or, conversely, because there are too many visual stimuli. This way, the confusion is avoided which often go together with pages filled to the brim (suffering apparently from the 'horror vacui' syndrome), trying to be attractive with a flurry of banners, graphics, animations, etc, only to communicate anxiety to the user in the process.

Actually, surfing is not really possible on a Google page: all different components there have a purely functional purpose. Their goal is to have the user

access a service, not to lead her/ him on a journey; their usage engenders behaviours which subsequently turn into routines of search, and become a default mode within a very short time. The interface is designed in such a way as to make usage, behaviour dynamics, and expectation of the average user iterative. Thus, even after allowing for the 'personalisation' of an user, search comportments remain basically identical, so much so that one can speak of a 'universal tool'.

The organisation of texts and images is linear. It uses recurrent graphic elements, notably primary colors, and the images used are qualitatively homogeneous. The interface's display style is sober, to the point of austerity, despite the 'brand (and corporate) identity' the design reverberates [*N9]. It is informed by a specific aesthetic, which appeals to very elementary, yet in their simplicity, very effective, properties of perception.

¿From this almost instantaneous visual identification stem a facility of use far above that of Google's competitors' search engines. The level of ergonomy achieved by Google is mind boggling: it doesn't even needs to present itself as a basket of services in its interface, its visual architecture screams that message already. The different services' interfaces are all autonomous, separate and largely independent from each other, and they all carry the 'blank box' as hallmark, /and they are not directly linked to each other/. It is for instance not necessary to go through various complicated steps to reach the code.google.com service dedicated to technicians of all levels, when you come from the site images.google.com, which addresses a much larger public. You need only to 'go deeper' in the google.com site, and know how to search. Despite this fragmentation, we are all able to recognize the network of services Google offers; moreover, visitors are able to make use of the information sources in a combined and complementary manner. And this equally holds true for the 'browse-only' types, as for those who have developed a mild - or stark - addiction to Google's services (a.k.a. 'Google-totally-addicted', joyfully jumping on the bandwagon of each and every Google novelty)[*N10].

This decentralisation of services results in a particular relational mechanism, as Google users do not discover these new sectors not so much through Google itself, but rather by way of the informal network of other users, on other sites, where Google visitors tell of their habits and discuss their tastes and preferences. Users then automatically 'localise' themselves within the extensive gamut of Google services, something that happens as soon as they log in for a new service: for instance the appropriate language will immediately be offered according to geographic area of the user's IP address. It also becomes easy for Google to approximate the sort of users to which a particular service is directed, to evaluate what level of technical knowledge it will require, or to what extent there exist an affinity with other users of the same service. Thus, the ear-say mechanism becomes akin to a 'relationships-based informal PageRank' system.

A first approximation would be to say that there exists a local relational dimension, where the ear-say, 'by word of mouth' communication, concerns friends and acquaintances, together with a typological dimension of relationship, which is about particular classes of users, which can be identified by means of statistical parameters (age, sex, occupation, etc.), and who use a particular service and thereby kick a particular type of relational economy into being.

It would appear that Google too [also?] does [not] escape falling victim to the ten problems relating to the use of websites discussed by Jakob Nielsen, one

of the most prominent specialist of user interfaces [*N11]. Although written in [standard?] HTML language, Google's site is completely outside standards, and yet manages to be fully readable [i.e. compatible] with all browsers, whether graphic or linear, in use today[*N12]. [The second sentence seems to indicate a logical flaw in the first, so I'd tend to say that JN's critique does not apply to Google, something the content of *N11 appears to confirm - French text blues ...]

The neat graphic design of the Google home-page is further enhanced by an excellent visual organisation of its commercial aspects. No advertisement link whatsoever on the home page or in the documentation/ information pages. At Google's, ads are only on display together with the query returns, but clearly separated from these, although they are related to the matters the query was about. One can therefore say that Google is able to arrive, in the agencement of its interfaces, at an acceptable compromise between the respect due to its users and the necessity of economic returns. Advertisement, Google's main source of income, are displayed in such a way as not to be invasive and not distract users in their usage of Google's services. Advertisement links are sponsored in a dynamic fashion, adjusting to the user's trajectory within the search site first, and on the Internet [in general?] second.

Commercial links are thus not static, they move along with the users' searches. This is made possible by the RSS-feed (for RDF Site Summary, or Really Simple Syndication), one of the most used formats for the distribution of web contents, and also thanks to the many different digital information (re)sources (dailies, weeklies, press bureaus etc.) Google is using to dynamically modify its home page, when it has been personalised by a user. This as Google lets registered users completely configure their Google start page thought the addition of a RSS-feed, making it possible to have the weather forecast for cities of one's choice, or to go through the history of previous searches, and all this at one's fingertips. Bookmark management, keeping track of the last incoming e-mails all become possible, but also checks on one's not web-related /computer/ files thanks to the 'Google Desktop' application.

The commercial promotion mechanism [i.e. ads], the services and sophisticated profiling of users appear to constitute a coherent whole, both at the aesthetic and at the content level. And as they clamour themselves, sponsored links are nothing more than suggestions, though they are graphically compatible and conceptually cogent with the search operation in progress. Google's economy is so well integrated with its interface that it can vanish without harm being done from the vantage point of users who are not interested, while generating handsome profit from users who do show interest in the suggested commercial link-ups.

Yahoo! and many other search engines and portals offer the same sort of facilities to personalise their home pages. Yet the quality and quantity of what Google has to offer remains unchallenged till the day of today. The configurations are rather simple, yet they do require some familiarity with Web interfaces and need some time to be put to work [by whom? users? Google itself? unclear in French text]. The treshold of attention on the Web is notoriously low, pages are visualised and then left within a very short time-span, often just a few seconds, and thus a user who 'invest'a couple of or even several minutes in a website, reveals /through her/ his choices/ a lot about her/ himself and her/ his habits as a consumer. These informations are then carefully memorised by

the company owning the search engine (whether that is Google!, Yahoo! or another firm) and represent a true source of wealth produced cost-free by the user her/ himself. They are essential to the sponsoring companies offering targeted products and services.

Home page personalisation makes a site more attractive and intimate: the site itself becomes some sort of private tool in which the user goes on investing time by choosing colors, tweaking its outlook, and selecting her/ his favourite content. A recurrent [habitual] visitor who is able to configure her/ his start page participates in the construction of the web interface. Giving the user the freedom of choice and control over a few pages means transforming her/ him from a simple target of advertisement into an 'intelligent' consumer that is one you can extract 'intelligence' from. To foster interaction is surely the best and yet subtlest way to achieve 'fidelity'. This is why one sees the multiplication of participative interface environments, where ads are increasingly personalised, in order to let us all enter together into the golden world of Google.

5.5 PageRank[TM] or the absolute authority within a closed world

The algorithm that enables Google to assign value to the pages its 'spider' indexes is known as 'PageRank'[TM].

We have already seen that PageRank[TM]'s mode of functioning is based on the 'popularity' of a web page, computed on basis of the number of sites that link to it. Given an equal number of links, two different web pages will have a different PageRank[TM], according to the 'weight' of the linking pages: this constitutes the 'qualitative' aspect of sites.

To take a concrete example: quite often, when one checks out the access stats of a site, one encounters an enormous number of link-ups coming from pornographic sites. This is due to the fact that Google assigns ranking according to accessing links which appear in public statistics. There are therefore programs exploiting the invasive aspect of this connexion and node evaluation logic in order to push up the ranking. And pornographic sites are well-known to be pioneers in this kind of smart experiments (they were the first on the Web with image galleries, and with on-line payment models).

As a number of [spider?] programmes are looking for sites with the help of public access statistics, a very large number of links are actually established through bogus visits. These come from a fake link on another site, which is most often pornographic. This devious mechanism literally explodes the number of access to that site, causing its statistics to swell, and its (Google) ranking to lift up, which in last instance benefits the pornographic site issuing the fake link in the first place. It looks like a win-win situation, at least where visibility is concerned. And it is not an 'illegal operation' [;-)] either: nothing forbids linking up to an Internet site. This practice causes sites with public statistics to have a higher ranking than non-public stats sites.

This mechanism illustrates how Google's ranking's 'technological magic' and 'objectivity' are actually connected to the 'underground' of the Net, and is partially grounded on less savoury practices.

Other perfectly legit practices have been documented that exploit Google's approach to indexation, such as Search Engine Optimization (SOE), a suite of operations pushing up the ranking of a website in search returns. Getting to

the #1 position, for instance, is often achievable through spamming /from out improbable addresses by automatic programmes, with stupendous effects/.

"We register your site with 910 search engines, registries and web-catalogues! We bring your site in pole position on Google and Yahoo! Try Us! No risk, just US$299 instead of US$349! - one shot, no obligations!". Of course , confronted to this, Google still plays the transparency card: "nothing can guarantee that your site will appear as #1 on Google" [*N14].

Mathematically speaking, a feature of PageRank[TM], which is based on the analysis of links, is that the data base must be fully integrated, or with other words, that the search operations can only take place within a circumscribed, albeit extremely vast, space. That means that there is always a road that leads from one indexed web page to another indexed web page - in Google's universe, that is.

Searches therefore, will tend to be functional, by avoiding 'broken links' as much as possible, and also returns that are substantially different from what had been archived in the 'cache memory'. But the ensuing problem is that users will be falsely made to believe that Internet is a closed world, entirely made up of transparent links, without secret paths and preferential trajectories, because it would seem that, starting from any given query, a 'correct' response will always be returned.

This is the consequence of the 'Googolian' view of the Internet as exclusively made up of the journeys its own spider accomplishes jumping from one web link to the next. If a web page is not linked anywhere, it will never appear to a user since Google's spider had no opportunity find, weight and index it. But this does in no way mean that such things as 'data islands on the Net'do not exist!

Dynamic sites are a good example of this as their functionalities are entirely based on the choices the user has made. A typical instance is the [abysmally faktap - TR] site http://voyages.sncf.com owned by the French Railways. Filling in the appropriate form gives you train times, onward connections, fastest itineraries etc. for any given destination in real time [or so they say - Good Luck! -TR (uses bahn.de or sbb.ch instead, esp. for domestic France trips)]. Google's system is unable to grasp these forms' queries and hence does not index the results that have been dynamically created by the site voyages.sncf.com . Only a human person can overcome this hurdle. The only solution Google is able to provide is to redirect the user to the rail companies' or airlines' own sites when an itinerary, time table or destination is asked for.

This is the reason why the idea of the exhaustiveness of Google's data bases must be challenged and discounted, as these falsely conjure up the notion of one unique universe for all of us, which it is complete and closed and is called 'Google'. Quite the contrary is the case, as the act of mapping a trajectory in a complex network always means an exploration with only relative and partial results.

The dream of a Google "which has a total knowledge of the Internet" is demagogic nonsense whose sole aim is to perpetuate the idea that the information provided is accurate and exhaustive, elevating Google, as it were, into a unique, truth dispensing service - the Internet equivalent of the One Party System. Such an absolute fencing-off works admittedly well in everyday searches, because it leads speedily to results. But in reality, within a complex networked system, there is no such thing as an absolute truth, but only a trajectory-induced evaluation, or even a time-induced one, depending on how long one wishes to spend

on a (re)search. The quality of a search is also entirely dependent on the subjective perception we have of the returns, considered as acceptable, or less so. The networks we are able to explore, analyse and experience are complex objects whose nodes and linkages are constantly shifting. And if the decision as to find the results of a search acceptable or not depend on the user in last instance, then the exercise of critical faculties is essential, together with a sharp realisation of the subjectivity of her/ his own viewpoint. In order to generate a trajectory that is truly worth analysing, it is necessary to presuppose the existence of a limited and closed network, of a world made up only of our own personal exigencies, yet at the same time knowing full well that this is a subjective localisation, neither absolute, nor remaining the same in time. To explore the Net means to be able to carve the Net up in smaller sub-nets for the sake of analysis; it amounts to creating small, localised and temporary worlds [*N15].

It turns out that in everyday practice, chance linkages are of utmost importance: the emergence of new and unexpected relationships can by no means be predicted by the analysis of the web's separate elements, such as Google's ranking system suggests. These linkages fulfill the function of 'dimensional gateways' and allow for the lessening, or even the rank abolition, of distances between two nodes /in the network/.

5.6 PageRank[TM]: science's currency?

Contrary to common belief, the PageRank[TM] algorithm is not an original discovery by Google, but is based on the works of the Russian statistical mathematician Andrej Andreievich Markov, who analysed statistical phenomena in closed systems at the beginning of the 20th Century. Closed systems are understood as ones where each and every element is by necessity either the cause or the outcome of (an) other element(s) in that system [*N16]. Sergei Brin's and Larry Page's work must have been based on this /theory/, although the further advances they made therein have not entirely been publicly disclosed, aside from the Stanford patent [assuming that that is public - the French text, convoluted and unclear, would almost say the opposite].

Maybe the best way to understand the nature of this algorithm is to look at what happens between friends. In a community of friends the more one talks about one shared event or experience, the more it grows in importance, to the point of becoming something of common lore [here again, the French text lets me down, talks of ... "password between friends" ...] If the knowledge about this given event is confined to a narrow circle, it will not become very famous. The same logic applies to celebrities /in the show business/. The more they manage to be talked about, the more their ranking rises, the more famous they are, the more they become celebrities (this is the reason why there are so-many self-referential shows on television, like "Celebrity Farm" and others.) Google puts exactly the same mechanism to work in handling data.

But Google is much more convincing in its image management by spreading the idea that Internet should be seen as a vast democracy, since the algorithm functions as if links were votes in favor of sites. And it doesn't matter whether the link speaks good or bad about a site, the important thing is to be spoken about i.e. linked. The deception inherent to this 'global democracy' arrived at by an algorithm is immediately obvious: as if democracy was something coming out of technology and not of the practices of human individuals! We have already

stressed [*N17] that the cultural origins of such a worldview stem from the extremely elitist peer review system as practiced by scientific publications, where each researcher's contribution fits into a network of relationships, of evaluations and verifications enabling the communication and control of scientific research results. Google's 'global democracy' hence amounts to transfering the 'scientific method' /of publishing/ on the Web by way of the PageRank[TM] algorithm, functioning as 'technological referee' which is able to objectively weight the informations on the web and to order them according to the choices expressed through their links by the 'People of the Net'.

The likeliness is striking: on one hand we have scientific publication which acquire influence and authority in accordance to their ranking within their particular discipline, and this ranking is obtained by way of citations ('quotes'), that is by being cross referenced in the specialised literature. This is the way scientific research guarantees coherence: by ensuring that no new publication exits in a void, but function as the 'current art' within the long history of scientific endeavour. And then on the other hand, we have web pages whose links are taken by Google's spider as if they were as many 'citations' which increase the status, and hence the ranking of these pages.

Scientific elitism, the prime mover of the awe which 'science' inspires, is curiously based on publication. Publishing by the way, i.e. making public, does by no means mean making 'accessible' or 'understandable' [*N18]. Indeed, it was the contention of sociologist Robert Merton in the seventies of the previous century that 'scientific discoveries', whether theoretical or experimental, cannot, will not, and should not be considered truly scientific unless they have been 'permanently integrated into the body of scientific knowledge'[*N19]. This statement might appear somewhat apodictic (after all, science in Antiquity was not at all 'publicly' transmitted - think of the Pythagorean school in ancient Greece, or of the distinction made between 'esoteric' and 'exoteric' writings, etc), but it does clearly evidence the eminently public character of modern day science. Communication hence is not a derived product of research, but the integral part of a form of knowledge based on accumulation and co-operation. Science, at least since the 16th Century, on one hand strives for new results which would constitute an augmentation of the cognitive capital, but recognises on the other previous research as the necessary and unavoidable departure point of those. One can therefore initiate a history of scientific communication which would develop in parallel with that of its media supports: from the voluminous correspondence scientists used to maintain with each others, through the periodical publications in scientific reviews up to the electronic communication carriers of today. And it is not fortuitous that the first Internet nodes were academic research centers /which had the need to communicate and share information/.

Nonetheless, the evolution of carriers did not influence the basic tenets of the scientific method's mode of communication, which remains based on citations. Dubbed 'science's currency' in some quarters, citations function as tokens of honour given by scientists to the people who taught and/or inspired them. More concretely it links present to past research, whether from the same, or from different authors. And it makes indeed sense to consider that the number of citations ('quotes') a certain piece of research has attracted reflects its importance, or at least its impact, on the scientific community. With time, this system has become itself the object of specific research: bibliometrical analysis is a discipline which uses mathematical and statistical models to analyse the way

information is disseminated, especially in the field of publications. In fact bibliometry, and then especially its best-known indicator, the 'impact factor'[*N20] is being commonly used as an 'objective' criterion to measure an individual researcher's scientific output or that of an academic institution. A vast archive of bibliometric data has been put on line in 1993 - at Stanford precisely, the cradle of Google. The SPIRES Project (for Stanford Public Information Retrieval System) was born in 1974 out of the series of bibliographical notes about articles on high energy physics established by the library of Stanford University. Because its domain of analysis is limited and well-defined, SPIRES is an exhaustive, publicly accessible, and free database, making complex searches possible on the body of citations. It is likely that Brin and Page were able to study and emulate this methodology when developing their own PageRank[TM] system.

But besides this algorithm itself, there are more adaptive features which have contributed to make Google a true 'global mediator' of the World Wide Web.

6 Chapter 5 As bonus: other funky functionalities

6.1 Filtered algorithms: ready-made data banks and control of the users

Graph Theory [*N1] is the mathematical basis of all network algorithms, PageRank[TM] among them. This branch of mathematics studies methods to create, manage, and navigate different classes of networks, and to describe them with graphs, and rank them according to their size. The introduction of electronic calculators saw Graph Theory take a huge flight from the mid 50s of the previous century. In terms of geometry, one can figure a graph as a collection of points in space and continuous curves connecting pairs of points without crossing. In Graph Theory, a graph (not to be confused with a graphic) is a figure made up of points, called vertices or nodes, and of of the lines connecting them, called arcs, edges or arrows. [cf. Wikipedia, 'graph' & associated entries][*N2].

A network is a particular type of graph, in which it is possible to assign a different value, or weight, to separate arcs. This enable to establish different values for different routes between nodes. The Internet is a graph, and the same can be said of all web pages taken together. Google's search system is based on this principle.

One of the most fundamental aspect of network algorithms is the relationship between the time factor and the number of examined nodes. The 'travel time' of a search, for instance, that is the time it takes to connect one node to another, is dependent on the number of elements in the network, and is always set between a minimum and a maximum value. This value of which can vary widely according to the type of route algorithm used.

In the network of web pages, every page is a node in the graph, and every link is an arc. Taking the time factor as starting point, it clearly appears that search returns generated by Google as answer to a question (technically the returns of a query on its data bases) can impossibly be based on an examination of the 'entire' Internet.

Google's spider is constantly busy copying the Internet into its data base: not an easy task. However, is is not believable that the search engine browses through its complete database every time in order to retrieve the most important results. The key factor enabling Google to return almost immediate results is dependent on hidden sequences narrowing the general selection of data, meaning concretely, it is dependent on the application of specific filtering devices. Starting from the query itself, the filter makes sure the final result is promptly arrived at by way of a successive side-steps and choices which have been developed with the explicit aim to limit the range of the blocks of data that are likely to be analysed for that particular query.

This is how Google can return results for queries in an astonishingly short time. However, this makes the search process just as opaque as it is fast, or with other words, the search shows no coherence with the body of data extent on that indexed part of network. Results for a search will be returned very quickly not only thanks to the massive computing power available, but also, and foremost, because filters are there to limit the extent of the data pool that will be searched.

The filter's difficult task is to make a drastic selection of the network nodes to be looked at in order to exclude some and valorise others and their associated linkages. This method aims to exclude or include whole blocks of data amidst those that would generate results [French text not really clear].

All this is made possible by the existence of pre-set, ready to use search databases, returning standard answers to standard questions, but also tweaking the returns through individual user's profiling. The user's profile is made up from her/ his search history, language, geographic position (IP address), etc. If a user habitually conduct searches only in French for instance, not the whole of Google's database will be queried, but only the French language part, obviously saving a lot of time in the process.

Given the humongous amount of data, it is simply unthinkable to use transparent algorithms, meaning ones that will hit all the network's nodes. It is therefore unavoidable that some manipulations, simplifications, and deliberate limitations in the number of possible analyses are taking place, and this both for technical reasons of computability in the strict sense, as well as for evident economic reasons. And one can, without falling into unjustified vilification, easily conceive that within a system already biased by approximations caused by filtering, further filters could be added to add, or maneuver into a better position of visibility, those returns that go with paying advertisements, or which carry some doctrinal message [?].

However, seen from Google's point of view, it must be noted that filters are not directly linked to an economic motive, since they are not meant to sell something. They are linked to the habits of the user, and her/ his personal interests. Google sells ads, not products (or if so, in a very limited way only, like Google Minium hardware, or indexation systems for companies and organisations). Google's prime consideration is therefore to obtain data generating parameters which can be used to target advertisement campaigns accurately. The personalisation of results according to their recipients is made possible by the information Google /furnishes and/ gathers in the most discreet way possible. E-mail, blogs, 'cloud computing' (or 'virtual hard disks') and other services function as as many databases in a way that is much more suitable to profile users than these could or would ever fathom.

Hence, the additional services Google offers over and above search are very useful to the firm for experimenting new avenues of business, but also and foremost because they play a key role as 'aggregators' of personal information' about users.

A prime example is the electronic mail service GMail, a virtual hard disk of sorts (2GB for the moment, and counting...), which [in its beta phase, when the book was written -TR] is made available through a distribution system based on PageRank[TM]. Put simply, each (user) node of the Google network has a certain weight (allowed number of invitations to join) and can use it to offer the service (via a link) to her/ his acquaintances. This method enable control over the usage made of the service, and at the same time the user discloses to Google key intelligence about her/ his own network of friends and acquaintances.

In a second stage, this mechanism spreads out among invited individuals, who may extend new invitations: this way, a graph of /human/ relationships between the users will be created, representing an enormous strategic value with respect to 'personalised' ad targeting.

If one considers all the information that can be gathered from e-mail traffic (to whom, why, in which language, which formats, which key words, which attachments, etc.) one can surmise the existence, in Google data vaults, not only of a partial - but significant - double of the Internet, but also, of a copy, equally partial, equally significant, of the relationships, personal, professional, and affective, of the service's users.

In theory, filters merely serve to make the query process faster and more conform to individual requests. They are even necessary, technically speaking. Their usage, however, shows to which extent it is easy, for a party actually in position of dominance as regards to search services, to profit in a commercial sense of the data at its disposal, without much consideration to the privacy of its users.

To resume, Google's database today is able, based on what it knows about this or that user, to marshall with the help of a few key words a query in a manner that varies according to the type of user. Far from being 'objective', search returns are actually pre-set and fine-tuned, and using the search service enables Google to 'recognise' an individual better and better, and to present her/ him with 'appropriate' results.

Use of each Google service goes with acceptance of a whole set of rules and liability disclaimers by the users. Google, from its side, promises it will not reveal personal information to third parties. Yet, it is easy to presume that Google is able to exploit and commercialise users' data to its own ends [French text: different ends]. And then we need not even to consider the possibility (or rather: the probability) that all sorts of intelligence and police services can access these informations for any reason of 'national security' they may like to invoke. [The addition of more search filters in order to further personalise results is the most likely outcome. - unclear sentence in French text]

6.2 Google's cookies: stuff that leave traces

Users profile are always based on a system of search and selection *N3]. Two types of profiling are prevalent on the Internet, one is straightforward, the other is by implication. Explicit profiling necessitates registration whereby the user fills in a form, disclosing personal details. The information send are archived in

a database, to be analysed with the help of a string of parameters partitioning registered users into homogeneous groups (according to age, sex, occupation, interests, etc.). Conversely, implicit profiling is arrived at by tracking anonymous users during their visit to a site, through their IP address, or through cookies. Cookies are little text files used by web sites to leave some data behind in the user's computer. Every time the user comes back to that site, the browser resend the data stocked in the cookie. The aim is to automatise login authentication, to refresh /eventual/ running operations, but mostly to 'reunite' the user with data from her/ his previous visits.

Most Internet sites offering online services use cookies, and Google is surely no exception [*N4]. The combination of cookies and filters on an algorithm enable to track an individual's navigation, and hence to accumulate information on her or his 'fingerprint'.

Let's take an example: Individual 'X' has a mobile phone number on her name, and uses her mobile to call his family, friends and a few work colleagues. After some time, she decides to do away with this phone and take another one, not in her name, for the sake of her privacy. Now with her new phone, she reconstruct her circle of acquaintances by contacting family, friends, and colleagues at work. This sequence of 'social links' /(family, friends, colleagues)/ is, within the totality of all the world's phone calls, a unique one, which cannot be dissociated from the individual in question. So it is not impossible to formalise such a sequence as graph representing the nodes and the arcs of a network. The values of which (the respective 'weight' of the links between different nodes) could be determined by assigning 'affinity value' to 'proximity', starting from the departure point of the analysis, in our case individual 'X'.

Getting rid of cookies is an excellent privacy protection practice, but [as?] one can easily extrapolate to search engines from the preceding example. With cookies, just by looking at some specific search themes, it becomes possible to identity groups, or even single individuals, according to the 'fingerprints' they leave behind on the Web.

The unique trace which identifies our movements, our social contacts, our telephone calls, is just as unique as our preferences, our tastes, our idiosyncrasies, and our passions, which make each one of us different. Passions would be in this case, the sites we visit, and for Google more specifically the searches we are launching during our navigation. This mass of information we are giving to a search engine makes the compiling of our 'fingerprint' possible [*N5]

Like all cookies, the ones on the Internet have a 'sell-by' date. Internet sites sending cookies to our browsers must give a date by which the browser is allowed to delete the information contained in the cookie. A smart use of cookies is not something often encountered: the fact that Google was able to exploit to its own advantage a technical trick to POSIX developers is interesting in this regard. (POSIX is the international standard that permits interoperability between Unix and Unix-like OSs, such as GNU/Linux). The expiration date for Google cookies is somewhere in 2038 - more or less the maximum possible. This means that for all practical purposes, the browser in our respective OSs will 'never' delete these cookies and the information contained therein [*N5].

6.3 Techno-masturbation: create! search! consume! ... your own contents!

It is next to impossible to follow the rapid evolutions Google is going through on permanent basis. New services are launched in a quasi-convulsive way, and it is difficult to understand which ones are actually meant to have an impact on our lives, and which ones are likely to be discarded in the next few months or even weeks. And anyway, it does not make very much sense, in view of the fast rate of innovation and information 'burn' on the Internet, to lose oneself in complicated descriptions and exhaustive classifications which would inevitably contain errors and omissions. The natural dynamics and fluidity of the networks should dissuade from any attempt at attaining complete knowledge - in case someone would be attracted to do so. One would get lost even before having started on such a ill-advised endeavour.

This being said, one can, albeit from a subjective and fragmentary viewpoint, try to formulate a general critique of Google, without going into technical details and even less engaging in uncertain forecasts. As far as personalisation is concerned, the increasing prevalence of the concept of 'prosumer'[*N7] is probably the most worth considering.

Google is well-known for its propensity to launch 'beta' versions of its services, when these are still provisional and under testing mode. This is a dynamic, as we have seen in the previous chapter, which is directly inspired from the modus operandi of the Free Software development communities. Users contribute significantly to the development of the new service through their feedback, impressions, and suggestions regarding its usability; they are at the same time users and producers of the service - 'prosumers' being the name given to this hybrid breed.

In its aim to become the position of global mediator of web contents, Google sells technology and search results (through advertisments) to users, who on the other hand, tend more and more to be the creators of net content, and the consumers of the same through the services of Google, which are more and more personalised.

Two examples, which would seem, de prime abord, to have very little to do with each others, should make the point regarding this closed content production and consumption cycle: Google Web Toolkit (GWT)[*N8] and the alliance between Google's Gtalk and Nokia [*N9].

In May 2006 Google launched 'Google Web Toolkit' , a 'framework' that enables to develop Ajax applications written in Java script. Ajax (or Asynchronous JavaScript and XML) is a technique for the development of dynamic, interactive web applications using standard HTML (or XHTML) together with CSS for the visual part, and with JavaScript for the dynamic display of and interaction between data. What you then get are extremely fast-moving sites since it is no longer necessary to download all information of the page afresh every time. GMail uses Ajax for instance. This is an important innovation which does transform the approach to creating web applications, as it is written in a language with high level objects (Java), which are then paired to GWT and networks compatible with all browsers [French text probably alludes to the "write once, run everywhere" mantra]. But on the other hand, there is no justification for a high-pitched announcement to the effect that an imaginary 'Web 2.0' has now come out, revolutionising the Internet by making it 'machine readable'. After

all, multiple platform software creation for bookmark sharing, social networking, automatic data aggregation techniques, etc. have been there for years. And the hypocrisy of large corporations like Sun, hyping up an alleged entry into the 'Participation Age', occults the fact that the aptitude to co-operation had been for decades the hallmark of hacker culture. And so the most elementary parts of the innovations that had been advanced by organisations such as the W3C for the Semantic Web (like XML/RDF standardisation) [*N10] are being peddled as revolutionary developments! Of course, Ajax and affiliated technologies do solve the very recurrent problem of portability of web sites, which for the time being are difficult to visualise on all browsers. The framework code is available under a Apache license, and is thus Open Source, but - as often happens with initiatives coming out of code.google.com - a number of essential elements (in this case the Java-to-JavaScript compiler and the 'hosted web browser') are distributed in binaries only, and one has to subscribe to an ad hoc license, which to all practical purposes prevents redistribution, further development of derivates, and its inclusion in commercial products. Furthermore, every time one uses the 'hosted web browser' permitting one to try applications out on one's machine before launching them on the Internet, a connection is established with a Google server, officially in order to verify that the latest version of the programme is being used. It is however obvious that this constitutes a very efficient way of controlling developers rather than serving the interests of users. Sure, the code they develop may, on the other hand, be freely distributed, even in commercial products. [unclear to me!]

GWT in effect, is a very simple way to create sites that are perfectly compatible with Google's indexation systems. For the time being, knowledge of a specific programming language like Java is mandatory, but one can easily fathom the moment that new developments will enable a first time user to put web objects like taskbars or image galleries, menus of all kinds and whatever what else on her page, without having to write or to know how to write a single line of code. And indeed, there are already extremely simple programmes for websites creation(e.g. WYSYWIG - What You See Is What You Get), but GWT operates directly on the Web. Such contents are thus immediately usable on static or mobile platforms of whichever type, provided these are able to access the Web.

Let's now imagine Google signing up commercial agreements for the manufacture of bespoke devices, by proposing to those who make use of its services simple, PC, pal- , etc visualisable web page making instruments, which at the same time make it very simple to be indexed by Google's spider. This because contrary to Microsoft, Google does not offer programmes for money. I needs, as we have noted earlier, to spread its standards around in order to manage profitably its economy of search.

And now for the Nokia-GTalk agreement. Gtalk is Google's VoIP service [*N11], and has recently been meshed into GMail, Google's e-mail service, so that the members of the 'Google communities' now can not only mail, but also chat and talk in real time with each others. At the end of May 2006, Gtalk became available on Nokia's new mobile platforms called 'Internet tablets', a kind of mobile phones specifically designed for web browsing. With this alliance, Google entered the world of mobile telephony through the main gate, with the prospect of being rapidly integrated in the public wireless networks (or wimax) that are being deployed in various cities, airports, motorways rest

areas, etc. And there is also a definite outlook for an agreement on video distribution: video.google.com is a warehouse of 'tapes', and television on the mobile is definitely the next stage in product evolution coming up.

To put it differently: Google provides the instruments to create contents according to its own standards. In the domain of content creation for the Web we see extreme personalisation, corresponding to the 'long tail' mechanism (which means providing each individual consumer with precisely the product demanded): the user creates 'exactly' what she wants - in Google's standard format. The complete decentralisation at the level of content creation parallels complete decentralisation of the advertisements, and hence the delivery of 'personalised' products.

What we are seeing is an invasive system imposing its own standards under the formal appearance of being democratic because it is allegedly put into the hands of the users, one click on the browser away. What is being peddled as electronic democracy morphs into far- reaching standardisation which makes it possible to absorb the contents created by myriads of users and to target the most appropriate advertisement at them in return.

6.4 Browsers as development environments

The outburst of ever more powerful new web services since 2005 has transformed browsers from simple navigation instruments into full-fledged development tools. There is a whole gamut of technologies which trump the current web programming standards by putting in the hands of developers a cool, easy, complete, secure and multi-platform tool: the browser itself.

Since a few years there has been a new trend in Internet sites creation, especially because of more importance being given to portability and accessibility of contents: this is what is clearly marked out in style sheets (Cascading Style Sheets, standard CSS and CSS2) on formatting, replacing the pure HTML of the validators, even of the XML standard itself [?]. Graphic and web designers find their browsers to be excellent auxiliaries, as these are ever more sophisticated and ever more compliant with various standards. This enables then to realise websites that can be visualised on various devices and platforms, yet while retaining, or even increasing, their range of expressive possibilities.

The birth and rapid diffusion of the Mozilla browser demonstrated the reality of a massive interaction between site developers and browser developers, which enabled them to do away with nearly all bugs and bridge almost all incompatibilities on web standards in a relatively short span of time. The incompatibility between the browsers Internet Explorer, Opera, and many others, whether proprietary or not, is a well-known problem among all webpage developers. The synergy Mozilla achieved to develop, which may look simple or even trivial, is an absolute novelty in Web history. Another interesting characteristic of Mozilla products is the modular structure which has been built around the Gecko layout engine, through which any functionality can be added. Real time stock market quotes, local weather forecasts, and programmes eliminating ads from websites are amongst the most widespread tools used.

Browsers have thus become ever more fiable instruments, enabling the creation of complex websites and have now all the characteristics of full-fledged programmes, so much so that they tend to replace more common applications. One of the more tangible example is the Office suite of tools Google offers as

an alternative to Microsoft's, or even to the F/OSS OpenOffice variant [*N12]. It is now possible to use 'Writely' (a product developed by a company Google bought up) as text processor. Other 'Internet in the clouds' options: Google spreadsheet, and Google Page Creator - the names say it all. All these services are [were in 2007] in beta-testing, on invitation only phase: needless to say, strictly for Google account holders - Mountain View control rulez!

Developers, from their side, show increasing interest for the Web side of things Internet, thanks especially to instruments like GWT. Naturally, Microsoft is not taking all this lying down. Taking its cue from its competitor's beta-testing strategy,/coming as we know from the practice of F/OSS,/ it has already launched the beta version of its own Office System (aka Office 2007), which integrates a lot of web-oriented tools, but remains nonetheless an application that has to be installed beforehand.

Browser are hence in the process of becoming full-fledged development environments for the sake of creation of standard content, also known as SDK, Standard Development Kit. But what is exactly the innovation that made browsers morph into SDKs? One can speak of a truly new paradigm in programming, so much is clear: it has now become possible to create fully multi-platform, client-side distributed programmes, which are hosted on a server, and therefore need not installation of complex frameworks on the users' boxes. Contents, including personal data, are stored in ever more remote sites (on Google's own servers, for instance) accessed through the Web /, i.e. bwo the browser/.

The choice for an 'Open Source' browser like Mozilla Firefox is often driven by the simplicity of configuration and the fact that so many powerful extensions go with it - for free. Developers use this particular browser to engage in ever more complex and structured programming. The emergence of programmes that live only on the web has two far-reaching consequences on both the market and users: programmes with binaries that need to be installed on the users' machines become obsolete and browsers themselves become more complex pieces of programme[s], they are very modular, and they gain increasing favor on the IT market. Thus, one can therefore expect to see less '.exe' (MS Windows), ".dmg' (Apple Macintosh), or Unix packs in future, and more browser-integrated tools, and more extensions to read RSS feeds, from GMail up to complete office software suites.

The detailed control on use Web service providers obtain through these instruments make these dynamics potentially fraught with dangers for us all, since all parties offering this type of)[these] services know the precise digital ID of their users, the length of time spent with the software and the contents under elaboration, because they control every step and know every detail about access and usage.

Seen from a technical point of view, this mechanism is based upon the fact that there is a permanent connection between the 'client' (the browser) and the server (literally, the service provider), which enables the latter to constantly monitor the requests, the time spans, and the intentions at stake. Moreover, allegedly in order to 'protect' the service against all kinds of hackers and crackers attacks, the authentication process is no longer taken care of by a compiled, source-less [?] programme, but is directly hosted on the providers' servers. Now, malevolent hackers bent on 'penetrating' a software programme, must 'crack' the remote server first.

6.5 Privacy, Paranoia, Power

The accumulation strategy pursued by Google has now enabled it to put the giant Microsoft itself in a difficult position, foreboding a merciless war of standardisation and of control of the access to the Web and to all other [the] networks we use everyday. From the moment that the Google phenomenon addresses the global mediation of information, it concerns all the users of digital data, that is: us all. To go through Google's history means thus to look at our own past as explorers of the Internet and of the Web. Too often have we outsourced the management of our information, of our sites, of our picture galleries, of our blogs, of our SMSs, of our phone conversations, etc. etc., to companies that are everything but free from ulterior motives.

The 'strategy of objectivity' pursued by Google emphasises scientific research, academic excellence, technological superiority, and sophisticated interfaces. But) This is merely a veil occulting the frightening prospect of a single access point to all data generated by naive users.

The F/OSS strategy then, allows Google to adopt the collaborative methods of development typical of digital communities - adapting it its own 'mission' in the process. But even in this case, as we have seen earlier, Google makes preposterous claims by proposing so-called new methods to exploit well-known dynamics, the 'Summer of Code' being a prime example.

Google's activities, therefore, constitute a clear and present danger to all who think that privacy, or, at a higher level, the whole issue of due diligence in the matter of 'being digital', is of primary importance. We are witnessing the emergence of a power conglomerate which is gaining, even as we speak today, far too much influence in the life of far too many people. Google is the holder of confidential information which it analyses all the time in order to achieve a steadily more personalised distribution of the plague that are advertisements. And since the accumulation of powers usually leads to the syndrome of domination, it becomes urgent to look in depth at this phenomenon.

There is no such thing as a global answer to resolve once and for all the issue of privacy. Big Brother does not exist, and like all paranoia, the fear his image induces blots out possible escape routes: it only serves the dominant power that thrives by it.

Secrecy, cryptography, steganography are examples of useful practices, but they are not definitive solutions. Communication and sharing remain the object of a desire, that only can be brought about by 'publication', i.e. by making public. Conversely, obsession with secrecy rapidly leads to paranoia and complot theory. Seen this way, what is the purpose of constructing complicated alternatives to arrive at absolutely secure and sheer impenetrable networks? Technology offers the opportunity for openness and sharing. To make use of machines means to make use of hybrid creatures, of material artifacts (which in this sense belong to the world of 'nature') that have been invested with cultural meanings and values (something that pertains to the domain of 'culture'). Networks are the outcome of a co-evolutive dynamic of mechanical, biological, and signifying machines: technology is essentially a mongrel. To create networks mean to connect machines of different types. It means creating methods of sharing, of exchange, of translation: one cannot withdraw in one's shell. It becomes necessary to engage in self-questioning and change.

We need informed research and analysis; it has become more urgent than

ever to denounce the mechanisms of technological domination. Conversely, to renounce critical thinking and attitude amounts to the same as giving in to the syndrome of control, which is becoming increasingly invasive. Google's history can be used in an exemplary manner to sketch out and promote the ideas of openness and to imagine practices towards the autonomous self-management of technologies. This because Google represents the meeting point between the meritocratic habitus of the University, the desire for boundless and unfettered innovation, and the edgiest form of financial capitalism. Here then rises the occasion for the development of autonomous and decentralised networks, and the opportunity to confront the desire to 'explore' and to 'navigate' the Internet, to the necessity to 'access' the data, this in order to focus attention on the trajectory, rather than on the result.

7 Chapter 6. Quality, Quantity, Relation (part 1)

7.1 The Rise of Information

The information society is heterogeneous in the extreme: it uses network communication systems like telephony, digitalised versions of broadcast [*N1], pre-Web traditional media, like dailies, radio or television, and Internet-born ones like e-mail or P2P exchange platforms, all this with gay abandon, and even without an afterthought. But a closer look reveals that all these systems are based on one single resource: information. Now within the specific domain of search engines, and thus of information retrieval, one can state that what consists information is represented by the sum total of all extant web pages [*N2].

The quantitative and qualitative growth of all these pages and of their content has been inordinate and continue to be so. That comes from the fact that it has become so unbelievably easy today to put up content on the Web. But contents are not isolated islands, they take shape within a multiplicity of relationships and links that bind together web pages, websites, issues, documents, and finally the contents themselves.

Direct and unmediated access to this mass of information is well-nigh impossible, even as a simple play of thought: it would be like 'to browse through the web manually'. This is the reason why there are search engines, to filters the Web's complexity and to serve as interface between the information and ourselves, by giving us search results we are happy with.

In the preceding chapters, we have reviewed the principal working tools of a search engine, that is the instruments Google, and other search companies, have put in place to scan through web pages, to analyse and order them with the help of ranking algorithm, to archive them on appropriate hardware supports, and finally to return a result to the users according to their search queries.

The quantity of stored web pages in memory is thus crucial to estimate the technical and economic potency of a search engine. The larger its 'capital' of checkable web pages, the higher a search engine will score on fiability and completeness of its returns, but this obviously within the limits of the specified context.

Yet, however enormous a search engine's 'pages capital' may be, it will, and could, never be entirely complete and exhaustive, and no amount of time, money or technology invested in it could change that. It is absurd to think that it would be possible to know, or, at a more down-to-earth level, simply to copy and catalogue all the Internet. It would be like the pretense to know the totality of the living world, including its constant mutations.

The information storage devices used by search engines like Google are like vessels: let's imagine we'd have to fill an enormous vessel with diminutive droplets (think all the pages who constitute the Web's information). Assuming that our vessel is able to contain them all, then our task would be to capture and identify them all, one by one, in a systematic and repetitive manner.

But if on the other hand, we'd think there are more droplets then our vessel can contain, or that we cannot fathom an algorithm to capture them all, or that the capture may be possible but will be slow, or even that the whole task may be hopelessly ... endless, then we would need to switch our tactics. Especially as our data-droplets change with time, pages get modified, and resources are jumping from one address to another...

At this stage, we might decide to go only for the larger droplets, or to concentrate our efforts on those place where most droplets fall, or we could chose to collect only those droplets that interest us most, and then try to link them together in the way we think is the most relevant.

As search engines companies continue to go after the holy grail of cataloguing 'everything' on the Net, it might be better to take a more localised approach to the Web, or to accept that for any given 'search intention', there may well be many answers possible, and that among all these answers some may be 'better', because they conform to specific demands regarding [either?] speed [or?] and completeness. One should always keep in mind that the quality of results is dependent upon our subjective perception when it comes to being satisfied with a search return. And in order to accept or to reject a search return, it is essential to apply our critical faculties and to be conscious of the subjectivity of one's viewpoint. In order to establish the trajectory one is really interested in, it is necessary to assume the existence of a closed and delimited network, a kind of world that is bounded only by our own personal requirements, yet always realising that this concerns a subjective localisation, which is neither absolute nor constant in time. [I am not completely happy with this, but then the French text... etc]

¿From an analytical point of view, charting a network means being able to partition the network for examination into sub-networks, which amounts to creating little localised and temporary worlds (LCWs Localised Closed Worlds) each containing at least one answer to the search that has been launched. Without that many searches would go on with no end in sight, especially since the amount of data to be analysed go well beyond the ability of a human person to capture them all: hence this would be a non-starter. Conversely, altering and specifying the query, and refining one's vantage point, will generate a trajectory that is more concordant with the departure point [of the search?]. By looking at the Web as a closed and localised world we also accept that the very dynamic of birth, growth and networked distribution of information (even happening while this information may already have become invalid) is an 'emergence' phenomenon, which is neither fortuitous, nor with[out?] a cause.

Emergence [*N3] is a occurrence which can be described in mathematical

terms as an unexpected and imprevisible outburst of complexity. But it is foremost an event that generates situations which cannot be exhaustively described. To analyse and navigate an 'emerging universe' like the Web demands a permanent repositioning of oneself. This not only determines a 'closed and localised world' of abilities and expectations, but also the opening up towards new avenues of exploration (other worlds are always possible, outside one's own closed one), and thus the appreciation that results can only and always be fragmented and incomplete.

7.2 Quantity and quality

Indexation by way of pages accumulation is a quantitative phenomenon, but does not in itself determine the quality of information on the Web; there the prime objective is to capture all pages, not to make a selection. The relationships between the pages give rise to emergence because they are generated on basis of a simple criterion, links existing between them. The quality of information springs hence forth from their typology, and is determined by their ability to trace trajectories, without bothering about a need to capture 'all' information available [?]. Quality therefore depends mostly on making a vantage point explicit through a particular search trajectory: basically, it are the surfers, the pirates, the users of the web who determine, but also increase the quality of information by establishing links between pages. The power of accumulation of Google's algorithms is useful to achieve this, but is insufficient in itself.

The evaluation of the pages' content has been outsourced to algorithms, or rather to the companies controlling them. The whole Google phenomenon is caused by our habit to trust an entity with apparently unlimited power that is able to offer us the opportunity to find 'something' interesting and useful within its own resource 'capital', which itself is being peddled as 'the whole Web'. However, the limits of this allegedly miraculous offer are occulted: no word about was not in that 'capital', or only in part, and especially not about what has been excised from it.

The thorny ethical and political problem attenant to the management and control of information still refuses to go away: who is there to guarantee the trustworthiness of an enterprise whose prime motive is profit, however 'good' it may be?

Even though considerable economic resources and an outstanding technological infrastructure are put to the task of constantly improving the storage and retrieval of data, the political question that constitutes the accumulation of data by one single actor cannot and should not be sidestepped. Google represents an unheard of concentration of private data, a source of immense power, which is yet devoid of any transparency. It is obvious that no privacy law can address and remedy this situation, and that it would be even less the case through the creation of ad hoc national or international instances /towards the control of personal and sensitive data/. The answer /to the issue of confidentiality of data/ can only reside with a larger awareness and taking responsibility by the individuals who create the Web as it is, and this through a process of self-information. Even if this is no easy road, it is the only one likely to be worth pursuing in the end.

7.3 The Myth of instantaneous search

Since it is clear that Google's data 'capital' , gigantic as it is, will never correspond to the totality of the information present on the Web, presenting oneself as an 'instantaneous' interface, bridging the gap between the users search intentions and the so-called 'exact' result smacks of naivety - or of deceit.

Since the Web consists of nodes (pages) and arcs (links), every time one browses it by visiting pages, one is follows up links constituting a trajectory analysable through the mathematical models of the graph theory.

The pre-set orientations search engines will propose us will always lead us to the 'right' object, indifferent of the dimensions the Web might have or get in future. By applying efficiency and efficaciousness criteria, a search engine will chart out of query the 'optimised' trajectory, meaning that the number of nodes hit will be low, and hence the time taken by the search will look nearly instantaneous. Google actually pushes in the direction of one single trajectory, something illustrated by the "I'm feeling lucky" button on its main page.

This 'optimisation' squeezes search into a three pronged sequential scheme: user-algorithm-goal. On the long term, this dynamic leads to 'digital passivity', a stage where we simply wait till results are brought to us, for us to choose among them.

Moreover, this efficiency/ efficaciousness is paradoxically grounded not on an increase in the size of the data pool where searches are conducted, but on its opposite, on a limitation of the access to the information 'capital' , since no trajectory proposed by the search engine will ever take place in real time [French 'the moment t'] on the network, but will be calculated first according to what has actually been archived, and the user personalisation obtained through filters and cookies.

The access to the information offered by Google is fast, very fast, and looks even immediate, to the point of suggesting the annihilation of time, and to imply the existence of an immensity of data that have been perused for the purpose. The mediation of technology (through interfaces, algorithms, pre-set searches, etc.) makes this temporal 'annihilation' possible as well as the feeling of practically immediate access [*N4]. The rapidity of results return, however, has a detrimental [Indians would say:'deliterious ;-)] effect on the quality of the search. As everyone is aware who has conducted (re)search herself, the time one spend on (re)searching is a determinant element of the experience: to map out one's own research path, to make choices according to the moment, all this generate a feeling of being into it and is deeply satisfying. Google allows us to 'localise' in space (that is its own multidimensional space) what we want, but, however brief the time spent waiting for the result, we always adopt a passive attitude in front of the technological oracle.

In an active (re)search drive, the aim is no longer about 'access' to the data, but to accomplish a rich and variegated journey, and to use the (re)search endeavour for mapping out complex trajectories. Efficiency as a concept vanishes. The larger the number of visited nodes, the greater the complexity of the interlinkages we conceive, the more numerous occasions will be to trigger significant choices, and to refine our (re)search. This approach allows for a cognitive enrichment going well beyond the immediate performance. For instance, when we visit links offered to us by a site we are visiting, and then continue our navigation on sites that have been marked as congenial, we create every time a unique

trajectory; maybe we'll even resort to bookmarking them. Such a procedure is starkly at variance with a coherent user-algorithm-result sequence, but it does create a rich path full of sidelines, of branches, of cognitive jumps and winding detours, all catering to a non-linear cognitive desire [*N5].

To conclude, search engines are perfect tools for fulfilling the quantitative aspects of a (re)search taking place within an already fully structured resource pool, such as are lexicons, cyclopedias, etc. Here,The quantity is directly in proportion to the accumulation and computing potential: Google's reach obviously dwarfs that of all its competitors, but in order to retain its position, Google needs to constantly expand in terms of algorithms, machines, users, etc.

Conversely, quality needs not necessarily to reside with technological prowess or economic might. Nobody in her right mind believes that the results returned correspond to the full spectrum of available information: the emergence of the best possible path cannot be foreseen, cannot be computed, but can only be arrived at step by step.

7.4 Under the veil of the myth

The positioning values of Google's ranking do not correspond to any clear evaluation criterion: yet, in the majority of cases results returned are [look?] exhaustive, that is, we can in no way tell whether something has escaped the spider, unless one is an expert in the issue at stake and knows a resource that has not been indexed by Google.

The capillary distribution of its search tools has made Google a 'de facto' standard. The white space ('blank box') where we type the keywords of our (re)search functions for the user as 'Weltanschaung' of sorts, promoting a very particular world-view, that of the idea of 'total service': the search engine will answer any question, and will satisfy all requests made in the realm of the Internet.

Epistemologically speaking, the 'blank box' represents a cognitive model of the organisation of knowledge: We request through the white space an answer to all the search intentions we have put forward: indifferent whether we wanted documents, or further information, or data, or that we simply wanted to 'navigate'. The (re)search activity becomes completely merged with the entity that provides the service, Google, [of which we have an invading perception (?)].

The habit of using this tool becomes ingrained behaviour, a repetitive activity: it becomes very difficult for users to imagine a different way to satisfy their need for 'input'. They have become tied up to the reassuring efficiency/ efficaciousness of the 'blank box'.

To be active on the Web, and hence to need access interfaces and tools for unearthing information and setting out paths is a is a profoundly contextual and diversified occupation. (Re)search is everything but homogenous and cannot be reduced to the use of the 'blank box'. What we request and what we desire does not solely stem from a desire that can be expressed in the analytical terms of quantitative information, but is something that also hinges upon the way we approach (re)search, the context in which we undertake that (re)search, our own cultural background and last but not least on our aptitude to confront novelty, explore new territories, and face diversity in general. It is impossible to satisfy the quest for information through a one size fits all solution.

Since the indexation of webpages is by definition only incomplete, in the sense that it is a selection obtained through the ranking system, what Google does offer us is the prosaic possibility to 'encounter 'something' we might find interesting and/ or useful in its overflowing amount of data in its collection of subjects [issues]. A (re)search intention, however, implies a desire to find, or even to discover, "everything what one doesn't know but that is possible to learn about". The 'good' giant then appears for what he is: enormous, extended, branching out, but not necessarily adapted to our (re)search purposes.

7.5 (Re)search Models.

The ambiguity entertained by search engines, wanting us to 'search in their) infinite environment' rather that in a closed, localised world that conforms to our (re)search intentions, comes from the formal superimposition of two distinct levels, that of the interface [*N6] and that of the organisation. The interface, in this particular context is the technological element through one accesses the information and the search gets executed; the organisation, on the other hand, is the architecture, the technological model through which information is archived and disposed. The two levels obviously influence each other: organisation-realated choices prescribe the use of specific interfaces, while the information that are visualised through these interfaces betray in their form[at?] the way they are archived. [?]

The problem with this superimposition is that such information is presented in the form of identifiable and unambiguous, single data. The user of Google moves in a linear fashion through the results list of the ranking; in order to move from one result to the next she needs to go back to the start list, with no cross-over linkages possible at the level of the interface [?]

With search engines, one retrieves information, but without any consideration being given to path that have been followed to obtain it. The interface which directs our interactions is the 'blank box' where our queries are inserted: at this first level of access, all information are on the same plane [have the same rank(ing) ?] They are homogenous, yet at the same time separate and fragmented in order to allow the listing of the results as they have been arranged in order of pertinence by the algorithm.

However, as far as the (re)searches one does on daily basis are concerned, the same results can be linked together in all sorts of ways, and it is not necessary to arrive at the same ordered arrangement every time, and neither does only a single 'correct' result obtain; on the contrary, a (re)search which is not about data structured like in a cyclopedia or a dictionary or any other object of that kind (and that may also change in nature over time), could well remain without an immediate answer, but would on the contrary require an effort of creativity, of 'mixage', and of recombination.

When a formal identity is being imposed between the level of the interface and that of the organisation, the outcome is by necessity a constraining model. In Google's case, as we have to do with what is perceived as an infinite power of search, the means to arrive at a result are being substituted for the (re)search activity itself.

Let's take an example: ... [the example taken is a French word, 'plume', whose English equivalent ('feather') would not yield the same illustrative power. Briefly, the authors argue that if you 'Google' for that word, the first returns

(out of 6.700.000 !) will be about everything (various IT companies, a circus, etc.) but birds-feathers or ink pens (also 'plume' in French) - I'll need to sort out a nice equivalent with the collective (or 'invent' one myself) - but maybe you have an idea?] ... A more extended perspective of what it means to 'discover' information, and that would take the cognitive potential underlying every information resource pool into account in a critical manner, would tend to see the access-search function as a process of exploration and creation rather than as one of localisation. The emphasis would then shift from epistemology towards ontology: it is non longer sufficient to know the information, but to become aware of our true role as creators of information [N*7]. Search engines that operate at the access level are therefore of no use for exploration, as they merely intervene on the first and basic level of the presentation of information.

Browsing is the moment of true dynamism in the linking together of digital objects, which are then able to express to the highest degree their heuristic and communicative potential. This is something that is learnt through experience, and it mutates as we are learning it, during the very activity of exploring.

There is a major difference between searching and finding. Google makes us 'find' things, causing the satisfaction that goes with the feeling of accumulation. But far more interesting that 'finding' is the search itself. And maybe it would be even more rewarding to find, but not completely, because that would mean that we are sill engaged in the act of (re)searching.

A search engine is an instrumental model that arranges information into a certain order. It would be more useful and also more commendable to imagine models that (re)combine information, and so generate knowledge.

8 Chapter 7 Technocracy

Analysis of the Google phenomenon reveals a colorful landscape, in which the economy of search is but one element within a far larger and more complex mosaic. Eric Schmidt himself states that Mountain View is setting up the foundations for a global information technology enterprise, a "One Hundred Billion Dollars Business", obviously something that is more than a mere search engine firm.

What it actually is, is an invasive knowledge management system, whose most significant development and methods we have sketched in the previous chapters: strategies that pair aggressive marketing with smart image building, propagation of highly configurable and personalisable, and yet always recognizable interfaces, creation of standard contents 'outsourced' to users and developers, adoption of development methods straight out the co-operative Free and Open Software handbook, use of state-of-the-art data capture and archival systems, information retrieval systems associated with advanced profiling techniques, both implicit and explicit, and last but not least, sophisticated personalisation of advertisements.

8.1 Technocracy or the experts of science

Experts have found in the control and manipulation of technology the ideal tool to maintain their power, impose their personal interests upon society, or acquire more privileges. The mechanism is absurdly simple: technology is being

(re)presented not only as the guarantor of the objectivity in scientific research, it is also used to validate the decisions of politicians in power, or more generally, those of any 'authority' that has access to the technological oracle.

The application of scientific research in its technological form is excessive, yet reality is constantly being interpreted according to that paradigm. The curiosity and desire for knowledge that inspire scientific research are being hampered by ill-informed profitability criterions which are the hallmark of contemporary private and public funding. If research does not come up with immediate profits generating technological artifacts, it is deemed uninteresting. The power's discourse then becomes technocratic, completely at the opposite end of community-oriented sharing, of self-management, of dialogue and mediation between individuals. To sell Google's technocracy as if it were a tool for direct democracy is a charade, meant to make us believe that we participate in some sort of grand electronic democracy game, when it is completely devoid of substance. Sure, we may publish what we want on the Internet, and Google shall index us. But us, who are 'dilettantes' and 'heretics', are not allowed to mention that Google's accumulation strategy resonates remarkably well with the market economy /system/, which is based on endless growth. This makes sense, since we are neither alumni of the London School of Economics, nor successful entrepreneurs, and us are not certified experts either. Hence we have no 'authority' whatsoever. Yet, sound common sense and Orwellian memories are more than enough to realise that such a growth, without end or aim, is the manifestation of the will to technological power that only consider human individuals as potential consumers and nothing else.

That is why PageRank[TM], which, as we have seen, is not merely an algorithm, becomes the cultural prism through which Google intends us to analyse everything. In a certain sense, what we witness, is an enforced extension of the peer review system - which works all right within the academic system - to the whole gamut of human knowledge.

Traditional authorities, like religious or political institutions, have hit rock bottom as far as their credibility is concerned. Nonetheless, their loss of grip on reality, far from having favoured the blossoming up of autonomous spaces has led to an unreal situation where no assertion can be hold for true unless validated by some sort of technological authority. The authority of machines [computers?], in most cases, is only a query return from a data base being dished out by the high priest of technology and assorted experts to the wealthy class of 'prosumers'. An extreme form of relativism is the hallmark of methods which pretend to extract 'the truth' out of the available and allegedly boundlessly numerous data, as one can surmise from the number of algorithms and filters that have been used [?]. The true meaning of 'any search an appropriate answer' is actually: 'a personalised product to every consumer'.

Confronted to this closure of creation, and of the management and application of knowledge at our expense, there appear to remain only two options: refusal of scientific culture considered as the root of all evil; or on the contrary, blind and enthusiastic acceptance of every 'innovation' brought forth by technology. However, between these— two extremes, techno-hate and technocraze, it should be possible to advance the curiosity which is associated with the hacker ethic, viz. the sharing of knowledge, the critical attitude towards 'truths', the rigorous verification of sources, all that while going for the way of open knowledge and free circulation of information.

Education is then a fundamental issue in this context, but the ways to disseminate scientific knowledge on a large scale are simply not there. Educational structures in Europe as well as in North America are only geared towards the production of specialists. As of today, no pedagogic model exits that would correspond to a demand for a 'dilettante' kind of scientific approach to knowledge, not even in countries with a non-western tradition like Brazil or India, which are nonetheless producing high level scientific research and state-of-the-art technology at low costs 'thanks' to unremitting international competition. A scientific activity that would be neither academic nor entrepreneurial, but decentralised and of a DIY kind is nowhere on the agenda, despite the fact that it is indispensable to foster basic competences and the ability to evaluate the technological innovation which concern all of us. More specifically, the whole notion of 'scientific culture' would need to be appraised afresh to cater for the all-round need to have an elementary command of what is needed to confront the technological tsunami that engulfs us.

The rise of Information technology to the status of main mover of technological innovation makes new scenarios possible: IT is not merely a technique to automatise the management of information, it also has a logic of its own, meaning that it constantly strives to alter its own underpinnings. IT is all at once material, theoretical and experimental. IT applies to the formalisation of language (and hence contributes to the formalisation of knowledge), and puts that to work with the physical components of electronics, developing from there languages which in their turn influence theories of knowledge. IT functions as a loop of sorts, following a very particular cyclical process.

In classic sciences one observes stable phenomenons: the science of physics, for instance, constructs natural data and create relevant theories. But with IT, and its derivate, computer science the phenomena theory helps identify are wholly artificial, they continuously change, both in nature and conceptually, in the same time and measure as theoretical and experimental advances make them more refined: the software that was developed on a computer ten years ago will be structurally different from one that has been developed the last month. What we held for true yesterday, we know today that we won't hold it for true tomorrow, when we will have more powerful machines that will do novel things: this is a living world as it where, and hence in a constant state of becoming.

8.2 Miracles of technology: from subjective opinions to objective truth.

It is amidst such a gigantic data base that the 'good giant' Google appears in the landscape with a message for us: we are part of a yet unheard of "global electronic democracy"; the results of PageRank[TM] are correct since they spring forth from a direct democracy, as expressed by the links validated by Google's algorithms, which reinstitute us, in a certain sense, in our rights to 'open it up'.

Epistemologically speaking, however, popularity can never been acknowledged as a test for 'objective quality'. Because if that is the case, then the concept itself of objectivity would be based on an unstated assumption, viz. that a mass of subjective ideas (the 'opinions' expressed by way of links) would somehow, as if by magic, be transformed in their exact opposite (in this case, a 'revealed' objective truth) by the sheer virtue of its number passing the majority

threshold. This is exactly how ranking becomes a token of quality, since it is the explicit outcome of a technology based on the manipulation of information.

But how can quantity ever become quality? One assumes, but without admitting so much explicitly, that the technical mediation of the algorithm is in itself a guarantee of 'objectivity', and one associates to this objectivity the qualitative characteristic of 'good', then of 'best', and finally, that of 'true'. And all this has to be rendered fast, nay, immediate, and transparent, thanks to the annihilation of the time factor and the ergonomic sophistication of the interface.

The consensus creation mechanism Google considers as the manifestation of 'direct democracy' by users voters does not convince, for two main reasons: first, it presumes that the majority is always right, and then, it implies that the majority opinions must necessarily go through a technological mediation to really benefit users. Yet how that precisely works is never explained properly.

The dichotomy between what is objective and what is subjective, superimposed on the concepts of truth vs. opinions, is totally inappropriate in the world of networks. Science, has always created nature-culture hybrids for the sake of exactitude, meaning that it has invented techniques and fostered technologies. On the one hand, observation and experiments take 'nature' as their field of endeavour, and it is in that sense that they can be considered 'objective'; on the other hand, the results obtained are highly subjective because science operates under influence of individual will and political and social perception; this as science is mediated through language (even though it is the language of scientific communication) and because science is also a source of power (up to and including the atomic bomb).

The technology that drives networks is the current application of the scientific method creating "nature-culture hybrids', which amounts to umpteenth scientific objects presenting themselves as ever more fiable tokens of reality, en lieu and place of human beings [*N1]. The technological hybrid that is PageRank[TM]'s verdict is henceforth more valid than anyone else's opinion, and Google's result carry more weight than the view on an expert in the matter. Was it only because PageRank[TM]'s advice is always one click away - unlike the expert's.

As we stated in the introduction, the Internet is a 'natural' phenomenon after all. It is a material object, made up of mechanical and electronic machines; and at the same time, it is a cultural phenomenon, because it would not exist without the meaning that culture has assigned to it, which is constituted by meaningful interaction between human actors, or, to be more precise between biological underling and electronic machines underling, and between both of them. The hybrid character of networks is the necessary consequence of of the hybrid character of the technology itself.

Another possible viewpoint on the issue of subjectivity vs. objectivity is about the decision-making model: how to decide what is relevant? It is easy to assume, in a relativistic context, that an information is 'objective' when it comes from a site (or a blog, or from Google, or from an official source) if this value judgement itself is the outcome of clear assumptions, of a transparent process, and of a limited, localised viewpoint. A network based on trust, that is a group of people who share information, opinions and knowledge in a general sense, readily discloses its working methods, its hierarchy if there is one, and the conditions it imposes to become member of its network or project. Every time one checks out the answers given by such a trust network, one can read these

answers as 'objective' in the sense of 'true' for and in that network, relevant as regard to its experience, and that precisely because they are the outcome of as many different subjectivities and of a strong interaction /between the members of that network/. If one feels to be in agreement with that community, one could then consider the information relevant, or one could also dismiss it in favor of other trusted networks.

Following this approach, and if Google was prepared to disclose publicly its decision-making mechanism, and if Internet users were able to understand that much, then the objective vs. subjective issue could easily be set aside. One would empathise step by step, search after search, with the network that would please us most, and be able to directly influence it, in keeping with our tastes and preferences, our ideas, and our idiosyncrasies - in one word in accordance to who we are.

8.3 Public sphere and private sphere

PageRank[TM] illustrates another dichotomy: the one between the public and the private sphere. In effect, everything that passes through Google is made public: who didn't find private e-mails amidst the ranking's returns, maybe because they were send by error to a public mailing list? Now that an ever increasing mass of personal information transits through Google - and IT carriers in general - the possibility of one's phone calls (channeled through VoIP for instance) being archived and made retrievable by way of a search engines no longer appears so distant. One could say that technology also manages to 'hybridise' public and private sphere: and anyway, to connect /to the Internet/ means to open up to the world, and once we have opened up, it is the world that opens itself up in our lives.

There are already networks which maintain practices that defeat any illusion of objective information. They make up for themselves, in a deliberate, precise, and totally subjective way, what they want to make public and what they wish to keep private. This phenomenon takes its full signification when a search engine turns out to be unable to honour a query whose specified quality is greater than the qualitative availability and proposed technical [infra?]structure. [excuse me? I think this simply means 'when the search engine gets stuck' (because the info looked for is not in its search data-base)].

The best-known example is peer-to-peer network search (P2P)[*N2] like 'eMule' - among others. The mass of data that can be searched in these networks corresponds to the data shared by the users, and changes in an irregular fashion with time; this is why such networks are described as 'transitory' [*N3]. This because a user is free to classify any material she puts in the system as either public or restricted to the private sphere. It is also a ground rule of P2P exchange that contents are to be offered and shared freely (at least those that are searchable within the network) in order to be able to receive other contents in return. In it self, P2P is a legitimate practice; but data (audio, video, text files etc.) may be shared abusively that are protected by copyrights, meaning that at least one fourth of the population should go to jail for illegally downloading a MP3 file or such. Moreover, an individual choice between public and private sphere appears to become more urgent by the day. A broad spectrum of possibilities is available, ranging form the moderate option of fostering information exchange up to the radical option of simply ducking any legal consideration,

and going straight to the source of piracy in order to enjoy anything the Web has to offer.

This spread of piracy by no means should be taken as proof that we are on the verge of a popular revolution, especially since the sort of piracy at stake here is mostly a half (in)voluntary one, and certainly not the outcome of a reasoned choice by committed individual to oppose the current knowledge protection system and face the consequences [N*4] It is more that digital technology has done away with the material constraints of reproducibility and that the consumer culture pushes us to desire without end. At least as far as information is concerned, we seem to find it perfectly natural to desire anything we care to imagine, even though it should be obvious that we will never be able to listen to but a fraction of the music we download from the Net, or to see if only a meaningful portion of the films we have stored on our hard drives. Yet, the existence of desire, which is limitless 'sui generis', when combined with technological opportunity, raises very serious questions regarding the distribution and equality of access to knowledge. The free and costless aspect which is characteristic of these exchanges defies the primacy of productivist economics. The diffusion of opinions, for instance by blogs, throws the very format of traditional mass media into crisis.

Basically P2P is nothing more than the surface level and the most widespread manifestation of a form of exchange which is entirely independent from any authority that is by statute above the community level. There are many more instances of qualitative (re)search which are all taken care of by trust networks which could redefine our orientation perspectives on the Internet. In certain cases , these subjective trajectories are centered around professional or cultural affinities, such as online forums, newsgroups, and specialised blogs; in others the binding element is their opposition to the official sources; one often encounters sufficiently structured examples to constitute an alternative model in terms of knowledge management.

It is therefore possible to imagine a - may be slow - evolution of knowledge circuits from blogs towards P2P networks. An exchange of programmes with no links to web publications, supported by dynamic networks, where it is possible to share information flows and data files between users, who would then really constitute circles of friends. This would enable to make the distinction between what is private and what is public dynamic and fluid, instead of static and frozen: every individual would be free to share her information, and this with a flexible level of accessibility /from public to private/.

8.4 Escape routes: independent media, cryptography, blogs, FoaF

The blog phenomenon [*N5] - websites exposing the personal views of their authors, the links they chose, and the comments of their readers [cf however Geert Lovink's 'Zero Comment' ;-) -TR] - has given rise to what is now commonly called the blogosphere. Estimates today [i.e. 2007 -TR] run to around 60 million blogs, with 4 billion links and 1.3 million postings a day. The blogosphere expands at the rate of 1 lakh [see chapter 1] per day, and doubles in size every 6-7 months [*N6].

Mathematically speaking, blogs behave according to Darwin's law and present the distribution characteristics of a 'long tail' market: a few hundred blogs amass

a considerable amount of links (with 4000 of them making out the blog's 'Who's Who'), but most of them - millions actually - have very few of them. In this sense, as we saw already in chapter 2, blogs are part of the same economic set-up as (re)search. A blog enables one to share her declared, individual, and subjective viewpoint without any filter; the number of links to a blog is witness to its popularity and hence measure of its authority, which can equate or even surpass that of dailies and other traditional media as far as influence on public opinion is concerned. Credibility, trust and reputation are only related to the blog's importance: as a particular blog's popularity grows, it becomes difficult for it to go astray without being immediately strafed - which means not be linked anymore and thus to vanish quickly from memory [*N7].

The authority commanded by Beppe Grillo's blog, the only Italian amongst the world's 100 most popular blogs as expressed in the number of links, is greater than that of the blogs of La Repubblica or Corriere della Sera [the 2 major Italian daily papers -TR]. The personage Beppe Grillo writes in an idiosyncratic vein, and does not claim to sell the truth: he simply tells from his point of view. In a certain sense, blogs create self-managed sharing spaces; sometimes they even become the sole sources of independent information amidst the 'normalised' mass media. This was for instance the case of Iraqi blogger Salam Pax (aka Salam al-Janabi) during the second Gulf War [*N8].

The greatest novelty blogs bought to the spread of information is the automatic bundling together of different sources through RSS feed, which has become the de facto standard of exporting content on the Web. In short, RSS is an automated method to rapidly switch from one web site to another, and to find and export contents that are of interest /to us/. The popularity of blogs is probably the main reason why RSS is so successful: thousands of weblogs are producing RSS contents, so that one sees a profusion of sites, called blogs aggregators, which offer a selection from the most-read blogs, and also of programmes that enable one to access any blog directly/on one's machine/. This makes it possible to search for contents in the blogosphere without going through Google.

The fact that it has become possible to automatically receive on an individual's computer the latest stuff that has been written on subjects of greatest interest to this user is an innovations that is clearly bound to have enormous consequences for the way the Web is being used. Or put more generally, we see here the first step of a development whereby it becomes possible to have any data at hand in a format that is easy to share, transform and expand. RSS hence makes an information accessible on all digital supports indifferent whether it is a site, a programme installed on a machine, a mobile phone, or what ever kind of technological device.

There are however many other ways to go looking after information. Whereas Google presents itself as a public and objective tool, Indymedia.org [*N9], for instance, profiles itself as a "collectively run media outlets for the creation of radical, accurate, and passionate tellings of the truth". Hence the group of people who make out Indymedia act in accordance with a very specific kind of public [publication?] policy: in the right hand column, the so-called newswire, everyone is free to publish. Nothing gets censored, even though "the posts that are clearly 'racist, sexist of fascist' are being hidden, "but not deleted". Indymedia therefore, is a service that spawns a kind of users who are active in creating information and which contributes to the emergence of shared knowledge and truths. The authority that 'creates' such truths is decentralised and

participative, it is also an ethical sort of authority (because it is a collective of human individuals), and not a 'mathematical authority' (i.e. an algorithm).

To question official sources by showing that it is possible to produce credible information in a decentralised and independent fashion is one of the aims of Indymedia.org, and of scores of other networks which have emerged around specific issues of interest [*N10]

Now, should informations be made public, or is it better to keep them private? An interesting answer to this question comes from the Freenet.org project [*N11], a decentralised P2P network, initiated in order to thwart censorship, and which makes use of the resources of its users (bandwidth, hard disk storage) in order to make available any type of information. Freenet was built with anonymity and security in mind, rather than transmission speed. The government or other bodies that censor information all have one characteristic in common: they employ censors who decide which information to let pass or to suppress, and what is offensive and what is not. Freenet is a network that allows nobody the right to impose its own scale of values: nobody is able to effectively decide what is acceptable or not, and technology serves the purpose of unlimited freedom of expression.

When using non anonymous network, if we want our information to reach only pre-established recipients, strong crypto[graphy] is the way to go - actually using crypto) is a basic rule if a minimum of confidentiality is to obtain in digital communication. Because they are not traceable by search tools, encrypted mails will not show up in query returns [French text: are not readable] [*N12]. Cryptography enables to use networks in a subjective and filtered manner, creating private exchange spaces on the users side, but also public spaces that are indexable and open for use by anybody. What is important is not so much to be able to hide something, but to preserve private spaces and to be empowered to decide in an autonomous and independent manner what we want to make public, and how and when.

An other option is to make everything public, or even better, /to decide/ to make public a completely subjective take on oneself, which also makes it 'liable to be searched and read'. A virtual identity (or a digital clone if you like) can be defined with a great economy of details without it being left to the mercy of profiling techniques French unclear - maybe the idea is that a minimally defined virtual persona will not yield very much in terms of profiling?]. Probably the most interesting idea at the moment is the 'FoaF' community, or 'friend-of-a-friend' [*N13]. This project aims at creating a set of machine readable web pages describing individuals, their field of interest, what they are doing in life, and what their mutual relationships are [*N14]. This way, the concept of 'trust' takes precedence over that of 'truth': I tend to trust friends of my friends by engaging in 'networks of trust' which are based on shared affinities, tastes and passions. FoaF is a way to create social networks, by promoting an active involvement: to join a FoaF network one must 'come out', describe oneself, and make oneself 'public' [*N15].

Seen from out 'trusting networks' [support of the Web Trust (?)], FoaF may be a guarantee of transparency with regard to the itineraries of our virtual avatars, by fostering more credibility and trust in interpersonal relationship that found their origin and were developed on the Web. The web then is being used to create spaces in which people are connected together in a big relational map, which one can criss-cross starting from whichever node, and without the need

to go through a centralised searching tool.

It is even possible to use a formula like FoaF for explicitly political aims: the Indyvoter network, for instance, presents itself as "a social networking website for political activists", and has as stated objective the bringing about of an alternative form of government [*N16].

Oppositions like nature vs. culture, objective vs. subjective, true vs. false turn out to be poor tools for real understanding. Information technology is by definition dual and hybrid in character: it is theoretical and practical at the same time, it creates objects which alter our way of thinking, and the use of these objects in its turn modifies the (information) technology itself. The digital domain is very real, and in reality, our black-and-white opinions make room for grey tints, and for an endless variety of colors, for all shades of opinion, for the differences which turn out to be an enrichment, for the heterogeneity which cannot be reduced to one single format. The means to alter, extend and multiply these spaces of freedom are there at our disposal: their potentialities can only be curtailed by our lack of depth and of imagination.

9 Conclusion

We have now arrived to the term of our exploration, having unearthed and hopefully shed light on a number of the more or less substantial secrets of the Mountain View giant...

We have seen that Google profiles itself, with some pride, as the device that is able to integrally manage the complexity of what knowledge is available on the Internet. It 'sells' answers as objective truths which are nothing else than the outcome of subjective trajectories filtered by search technologies. We should be careful not to fall in the trap of an 'esoteric' perception of this mechanism, and let ourselves be fascinated by the speedy returns to our queries. These almost mythical portals are in fact no more than smart strategies combining the use of advanced systems collecting, stocking, retrieving, and ordering data, together with direct and indirect profiling and personalisation of advertisement targeting.

And on top of that, state-of-the-art marketing and sophisticated communication management are the hallmark of Googolian 'evangelisation': see for instance the use of primal colors in the logo /'s visual identity/. Add to this the spread of highly configurable interfaces, which all the samekeep the firm's distinctive outlook under all circumstances - and the trick always works: the firm can enjoy the milking of the relational economy at all the levels of the interface thanks to correlation between users. And finally, Google has adopted the co-operative forms of development that are typical of F/OSS communities, cutting on the costs of its services while at the same time appearing to champion the cause of open access and distribution of knowledge.

Yet to refuse the hypocrisy of 'the perfect search engine' does not mean calling for a boycott. Was it only because members of the Ippolita Collective themselves have often used of Google during their research for this book!

In the same way, recourse to large common and freely accessible resources such as Wikipedia has proven to be very useful from an (en)cyclopedical viewpoint. This is because, if one knows something about the issue at stake, one can verify the correctness of information in an independent manner, by rationally

blending together bits and pieces from the Web [and doing so without fascination, letting the Web speak for itself [?] - French text unclear]. The critical use of sources hinges on people being able evaluate the trustworthiness of information by themselves, not upon the inherent 'goodness' of digital technology.

Information technology is not merely a device for the automatic management of information. It has its own logic, meaning that it constantly adapts and transform its own basic structures. IT is at the same time material, theoretical and experimental: it works on the formalisation of language (and hence of knowledge) , applies the results to the material components of computers, and out this come languages which influence in their turn the theories of knowledge, making for a feedback loop type of evolution.

Google pushes this feedback loop logic to its extremes: it is an extraordinary machine that manufactures itself through the very use its users make of it. In this sense it is an 'autopoetic' machine, which accumulates all base information millions of users insert daily into the Web (such as names, pictures, e-mails, search preferences, membership of forums, blog writings and readings, filling in of forms and surveys, browsing trajectories, etc.) and uses it for targeted, 'capillary' advertising. The data furnished by users have come to represent a gigantic human, social and economic capital. A capital that surely needs protection, but constitutes also a fantastic territory for questioning, experimenting, and giving free reign to curiosity.

Google responds to users' search intents in a flexible manner, and this response surrounds itself with a bevy of ever more sophisticated and customisable services. But this multiplicity is just a facade, and its only aim is to spread a form of consumerism that conforms to the relational economy, by way of mass personalisation of products and advertisements thereof. The abundance capitalism of Google springs forth from a carefully crafted branching at all levels of the imaginary of the consumers-producers, a.k.a. 'prosumers'. This as users are delivering their personal data, but also their impressions and suggestions about the use of these services free of costs; developers, from their side, contribute to the development of 'open tools' which have been provided to them with the sole aim to spread the Google standards, and which remain under the strict purview and control of Mountain View; and finally,as the employees of the Googleplex and other subsidiary data centers fully endorse the 'philosophy of excellence' as championed by the firm.

The profiling of the imaginary is but the last phase of the process of capitalist colonisation of the networks, something we have called technological masturbation. A mercantile spirit guides statements in favor of "individual free expression", itself conditional upon being subsequentially able to exploit these "expressions" in order to sell trinkets and other useless but personalised goodies.

Google advertises its 'economy of search' as if it were a new cyber-democracy enabling hundred of millions of individuals to communicate directly among themselvesand manage their own organisation, escaping the control of the state and other institutions in the process thanks to the firm's technological implements. This simplistic message unfortunately finds support with many 'democratic' media and intellectuals the world over, who are victims of self-delusion. According to these, the Internet is essentially democratic by nature: not only are individuals stimulated to supplant institutions on the Web, but institutions themselves are becoming the better for it. Technocratic enthusiasm even goes as far as to represent the informatisation of public administration, a phenomenon

known as 'e-governance', as a form of ideology-free governance, mobilising the commitment of 'net-izens. This new political identity actually brings about first person (digital) participation, and hence the emergence of a completely diffuse public opinion. As if it was possible to remedy the 'crisis of representation' [of the classic forms of political institutions] by a networked local, but globally connected democracy! We have attempted to identify the major deficiencies of this approach, which all amount to their ideological preconceptions. The basic idea being that technologies are 'neutral' by definition, and that this alleged neutrality stands for moral virtue, in so far as it is the outcome of an objective scientific research and practice, which is able to give every individual what she wants, quickly and effortlessly.

The complex informational mediation performed by Google is presented as a transparent, high-tech skein, which guarantees the users/citizens/consumers' free choice, who use/vote/buy while surfing on the 'free' Web managed by Google for the commonwealth.

Despite all these participative dreams, which are fed by cyber-democratic fantasy but are devoid of concrete substance, it is actually impossible to put really autonomous forms of direct democracy in place by centralising information, knowledge, and power, and by putting all these in the hands of a private company (e.g. Google) - and even less, in the hands of a government body (e.g. the Telecom Regulatory Authority).

Even The more progressive margins of the alter-globalist movement have not escaped the identity trap, as they call for a reformatting of class identity through a new centralisation of work, this time of the telematic kind. But they remain far removed of the sphere of individual desire, especially when they advocate social networking, as if it were a magic solution to all personal frustrations, achieved through a ritual of global technological [auto-]solidarity.

Only a choice for self-education can really pave the way for escaping technocratic domination. And a lot of work has to be done before it becomes possible to 'put into the commons' something of one's own and create synergies. Without sound technical preparedness, the so-called community bonanza rapidly turns out into a solipsistic exercise. [Hello Pranesh! ;-)]

The people who are administering networks, on their side, must learn to protect sensitive data and start drawing the line between what they want to make public and what they wish to remain private. Moreover, they must be able to decide which information is 'correct' and which one is not, based on their subjective evaluation at any given time. This as they must be conscious that they are altering the information landscape at the very moment they browse through it.

This is the only way how individuals can develop their autonomy: by evolving rules for the journey through the virtual landscape, and by developing a personal viewpoint.

Just like all technologies, the digital ones are neither good nor bad in themselves, yet, as we have seen, they are not neutral either: it all depends on the use that is made of them and the methods that have governed their development. And since they are hybrids with a power to influence upon real life, they surely also enable to highlight the contradictions between 'nature' and 'culture'.

This makes it possible, in its turn, to conjure another danger: the idea of the Web as a de-materialised experience, devoid of physical existence, which often leads to a blind and reactionary rejection of innovation.

According to this perception, the 'virtual' reality of cyberspace is replete with insignificant interactions, triggered by an 'online crowd blatantly unawares of the material disparities of real life: gender, race, wealth, social position, all set aside in the fluid and frictionless flow of fictional identities. This totally materialist idea is usually advanced by intellectuals and other elite observers who dissect digital technologies from the height of their pulpits without ever deigning to have the modesty to ask for the opinion of the 'kids' who grew up with these (same technologies).

But quite on the contrary, 'virtual' reality is so physical as to not be able to exist without mechanical machines, the silicon sand and the integrated circuits that make it up, and without biological machines, that are the users. Connection, multiplicity and de-territorialisation are not the same as 'immateriality'. Moreover, this attitude betrays a fear to see the world change and to be left behind, together with profound misgivings about the ability of individuals to transform and enrich themselves.

Digital technologies hence are and will be an agent of liberation only if they go together with the development of complex and conscious digital alter egos who are able to interact in an unforeseen manner. It is possible to use a multiplicity of languages to bring about a place where we all can meet. Among other things, the Ippolita Collective has concluded that it was essential to have recourse to the scientific method, to turn to the inexhaustible richness of the humanistic tradition, to make use of the dialogic force of political opinions, to benefit by the coherence of the feminist narrative, and to head for the limitless curiosity that is the hallmark of hackers. Trust in the possibility to tweak technologies in accordance with the desire of individuals is essential if one wishes to create networks that are really free, and not merely digital.

The chaos of contradictory messages, the at times ear-deafening noise amidst the signal, and the near-inconceivable dimension of the Web may well instill fear - and yet the voyage of exploration has only begun.

10 Appendix

I. The End of the World in a Cookie

The end-date for Google is January 17, 2038. This date is interesting because it falls two days before the End of the World itself. January 19 2038 is a critical day all over the POSIX world, which includes all operating systems deriving from the UNIX system (GNU-Linux, BSD, Solaris, etc.), and all UNIX-style network protocols (such as http, ftp, etc.). On POSIX systems, time is measured in seconds elapsed since 1st January 1970, this for historical reasons explined later. The measuring unit for this number is a 'signed integer' traditionally of 32 bits of memory. (For a refresher see: http://en.wikipedia.org/wiki/Unix_time)

If a programmer creates a variable of the signed integer type in order to stock digital data in memory, it cannot be more than 2 147 483 648 nor less than 2 147 483 647. A very large number for sure, but rather less so if we translate it in seconds. These 32 bits indeed have room for no more than 136 years...

When Ken Thompson invented UNIX, he did so as a kind of play, and he never would have imagined that he would revolutionise information technology; and even less did he think that some of his choice, for instance the 'file system', the hour, the character codes, the functions call [?], the programming

language etc., would all in very short time become law for coders the world over. Thompson had set the point Zero of his operating system to January 1st, 1970.

2 147 483 647 seconds after that fatidic January 1st , at exactly UTC 3.14 am on January 19, 2038. will be the ultimate second of the UNIX world, of the Internet (which functions thanks to UNIX protocols) and of most of the major servers of the world (all working on UNIX-derived OSs). After that second, we will be put back to the 13th of December 1901, 8.45pm. This problem will surely have to be addressed in the years to come, and will call for a change of paradigm in time management and in all UNIX-based systems. This situation is rather more complex and serious than the much-vaunted Y2K bug that was hyped up during the 1999-2000 transition. (More refresher at: http://en.wikipedia.org/wiki/Year_2038_problem)

Google's end date hence is the longest programmable date that can be set for a cookie, which means that our OS's browser will never delete these cookies nor the information contained therein.

For more detailed info, see: http://www.2038bug.com/ And the site of http://www.theiet.org/ under 'problem dates'