tripleC 4(2): 178-186, 2006
ISSN 1726-670X
http://tripleC.uti.at

Memory versus logic: two models of organizing information and their influences on web retrieval strategies

Teresa Numerico

Department of Communication sciences, University of Salerno (Italy) V. Ponte don Melillo, 84084 – Fisciano (SA)
E-mail: t.numerico@mclink.it

Abstract: We can find the first anticipation of the World Wide Web hypertextual structure in Bush paper of 1945, where he described a "selection" and storage machine called the Memex, capable of keeping the useful information of a user and connecting it to other relevant material present in the machine or added by other users. We will argue that Vannevar Bush, who conceived this type of machine, did it because its involvement with analogical devices. During the 1930s, in fact, he invented and built the Differential Analyzer, a powerful analogue machine, used to calculate various relevant mathematical functions. The model of the Memex is not the digital one, because it relies on another form of data representation that emulates more the procedures of memory than the attitude of the logic used by the intellect. Memory seems to select and arrange information according to association strategies, i.e., using analogies and connections that are very often arbitrary, sometimes even chaotic and completely subjective. The organization of information and the knowledge creation process suggested by logic and symbolic formal representation of data is deeply different from the former one, though the logic approach is at the core of the birth of computer science (i.e., the Turing Machine and the Von Neumann Machine). We will discuss the issues raised by these two "visions" of information management and the influences of the philosophical tradition of the theory of knowledge on the hypertextual organization of content. We will also analyze all the consequences of these different attitudes with respect to information retrieval techniques in a hypertextual environment, as the web. Our position is that it necessary to take into accounts the nature and the dynamic social topology of the network when we choose information retrieval methods for the network; otherwise, we risk creating a misleading service for the end user of web search tools (i.e., search engines).

Keywords: Logic, Memory, Information Retrieval, Search Engine, Knowledge Management, Web, Hypertext, Association.

Acknowledgement: I would like to thank Roberto Cordeschi, Gordana Dodig-Crnkovic, Luciano Floridi, Marco Gori, Lorenzo Magnani, Susan Stuart for their invaluable help.

31.We hypostatize information into objects. Rearrangement of objects is change in the content of information; the message has changed. This is a language which we have lost the ability to read. We ourselves are a part of this language; change in us are changes in the content of the information. We ourselves are information-rich; information enters us, is processed and then projected outward once more, now in an altered form. We are not aware that we are doing this, that in fact is all we are doing.

Philip K. Dick, Valis

1 The philosopher and the present

In “The art of telling the truth”, Michel Foucault, commenting an essay by Kant about the French Revolution, argued that he posed a new question for philosophy, the question about the present in relation to the philosopher, participating and belonging to it. The big issue about his present according to Kant was the understanding of the Revolution. The spirit of Revolutions does not consist in the event itself; it regards, instead, the perception and understanding of them by the people who did not participated to the revolutionary events. The importance of revolutions, then, consists in the feelings and thoughts of those who were not main actors in the event itself, but experienced a strong emotion in relation to it. The interpretation of the revolutionary events means, “to know what is to be done with that will to revolution, that ‘enthusiasm’ for the Revolution which is quite different from the revolutionary enterprise itself” (Foucault 1994: 147).

We believe that some revolutionary events that took place in the young history of Informatics[1] brought radical changes to the information management strategies. We want to understand the spirit of the information and communication technologies revolution shared also among those who did not participate directly to its conception and launch, and discuss how to keep that spirit alive today, accepting the grand challenge represented by the creation of new methods for searching on the web.

2 The other analogue machine: the Memex

Vannevar Bush (1890-1974) was the anticipator of the creation of WWW. He suggested a similar application, imagining the machine as a “future device for individual use, which is a sort of mechanized private file and library” (Bush 1945). This tool, called the Memex, had to work as the “human mind” that operates by association and not as an artificial index of a real library. According to him, it was possible to mechanize the process of selection by association and analogy, and he imagined his device as a supplement for the human memory. The Memex was his second machine: during the 1930s, Bush built a powerful analogue machine, the Differential Analyzer, used to “measure” the values of complex functions, necessary to various calculation tasks during the Second World War. The associative process of creating trails between different pieces of information and connecting them was more related to the analogue paradigm than to the digital one that was at the core of the logical structure of the computer.

If someone had told Vannevar Bush that his most famous project in the twenty-first century would have been the Memex, he probably would have laughed at the idea. He is definitely one of the most successful US scientists during the 1930s and the 1940s, not only for his astounding scientific and technological achievements, but also for his success as a politician and an administrator in science. In 1940, Bush was at the centre of a powerful network of scientists that accepted to cooperate with military partners during the war: he planned and led the *Office for Scientific Research and Development* (OSRD). He was also one of the creators of the peacetime substitute of the OSRD, the National Science Foundation (NSF). His analysis of research developments from a privileged position made him envisage two major problems of science in the Twentieth Century: the specialization of scholars and the amount of literature produced in each research area. It was almost impossible at his times to “keep abreast of current thought, even in restricted fields” (Bush 1945). According to him, it was necessary to extend, share, access and consult the produced records, in order to make them useful for science progress.

3 Memory versus Logic

[1] This term is more precise here than the most common “computer science” because it has to do with the process of automate information that is more central for our discussion than the invention of the stored-program calculating machines that is the direct emphasis of the used term.

The difficulty in managing the scientific literature of every specialization has increased dramatically since 1945 and digital technologies did not offer a solution to it[2]. Bush proposed an answer that was thought provoking at that time and is still inspiring for us now. He argued that selection of relevant material was the key factor to deal with such a huge quantity of information and criticized the mechanisms commonly used by libraries to index information. It was not only a bare proposal for a mechanical improvement of the actual process used by libraries to organize bibliographic data and obtain outputs for specific researches, he suggested, instead, a complete change of paradigm in the information retrieval strategies, as well as in the knowledge management area.

> Our ineptitude in getting at the record is largely caused by the artificiality of systems of indexing. When data of any sort are placed in storage they are filed alphabetically or numerically, and information is found (when it is) by tracing it down from subclass to subclass. It can be in only one place, unless duplicates are used; one has to have rules as to which path will locate it, and the rules are cumbersome […].
>
> The human mind does not work that way. It operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain (Bush 1945: 32-33).

Bush individuated a fault in the traditional cataloguing methods of library and a new perspective in the information management field. The new idea entailed the emulation of the associative strategy adopted by memory, when it selected a trail of ideas in the mind. It was possible that the chosen associations were meaningful only for the mind that created them. Despite the arbitrariness of the links, the method could be very effective in retrieving information and making sense of raw data. Bush did not believe that machines were able to simulate memory, but he was convinced that they could "augment" the natural power of the human brain by making sound and useful associations. He is an inspiring precursor of the web and of the hypertextual world, who had an explicit or implicit influence on all the people who actively shaped that world such as Joseph Carl Robnett Licklider (1915-1990), Douglas Engelbart (1925-), Ted Nelson (1937-) and Tim Berners-Lee (1955-). In the paper *Memex II* (1959/1991), Bush raised the question about what machines can do. According to him, they could store and recollect data, which is what they were meant to do, but they were also equipped to do logical reasoning. Discussing about logic, he launched a strong attack against the abuse of it. In his view, people should use logic only when the premises were defined precisely and data were stated clearly, without these guarantees logic was meaningless. The abuse of logic consisted in the application of Aristotelian rules to undefined premises. Moreover, all sound conclusions that could be obtained from some correctly defined axioms, according to precise rules, were already implicitly contained in the premises themselves, that were valuable only because they could organize the raw data, which would difficultly achieved by other means. Machines, such as Memex, not only could apply logic reasoning without mistakes, but could also be linked to other more sophisticated devices that were able to accomplish more complicated tasks, than retrieving stored data. We can conclude that if Norbert Wiener (1894-1964), the inventor of Cybernetics, was looking for the "Human use of human beings" (1950), Bush was creating the intellectual and social space for the "human use of technology", which was a revolution in itself, and the mother of all the future revolutionary achievements. In Bush's vision, it was very clear that there were two models of representing and organizing information and his conclusion seems still sound to us now.

[2] According to the research of the School of Information Management and Systems at the University of California at Berkeley *How Much information? 2003* URL: http://www.sims.berkeley.edu/research/projects/how-much-info-2003/, only in 2002 it was produced 5 exabytes of new information mainly stored in hard disks.

The first one is based on a traditional conception of logic and mechanization of reasoning. This approach tend to eliminate all the unpredictable faculties such as ingenuity, judgement and initiative and to rely on classification tools to describe available information. We can connect this tradition to Leibniz and his project of a C*haracteristica Universalis*, a system that included a formal language (*Lingua Characteristica*) and an inference method (*Calculus Ratiocinator*). Important logicians shared this project at the beginning of the last century, such as David Hilbert, Bertrand Russell, Alonzo Church, etc.

With the recent beginning of the web, about thirteen years ago, it was necessary to collect and organize information available on the network. The state-of-the-art technology, at that time, was based on the information retrieval tools used to access the huge static digital databases, whose theoretical background belonged to relational algebra and set theory. These fields shared common roots with the tradition of logic considered as a unique, universal, coherent, individualistic method to represent data. Despite the development and the success of the information retrieval techniques in the organization and query process of the raw data, these tools were not adequate to the way information was dynamically and chaotically stored inside the web.

The other model of representing information, preferred by Bush, was influenced by the associative capabilities of the human beings. These faculties have more to do with memory and creativity, than with deductive and formal abilities. They offer a representation of data that is neither unique nor completely free from mistakes and misunderstandings; but, in spite of the difficulties it faces, this modelling attitude is more fertile in the knowledge creation process. Moreover, this method seems more promising and adequate to the understanding of the web, considering its peculiar self-organizing topology.

4 Wittgenstein and the language as a social game

Ludwig Wittgenstein (1889-1951), the great Austrian philosopher, provided us with another fruitful tool to understand the mechanism that transforms the bare information into knowledge. In his reflection on language and meaning, his main concern in the second phase of his research (1953), he formulated the interesting hypothesis that learning a language means becoming proficient in managing a series of linguistic games. Nobody can use a language that the others cannot comprehend. Language communication is a social activity that has to do with the application of rules that allow people to understand each others, even if they do not share the same vision of the world. Language needs to convey meaning to the others and the connection between meanings and words is a social procedure that happens in time and is relative to a community of human beings who learn to use the same games and share a life style (*lebensform*). In order to cope with the raw data, people need to attribute a certain meaning to the information and this is possible only by learning how to use the rules that permit to answer adequately to questions about that piece of information for a specific community of speakers. Following a rule is an activity that, according to Wittgenstein, cannot happen only once for only one human being, in order for the communication to be activated, the orders to be obeyed and the questions to be answered correctly. The understanding process that transforms information into knowledge has to do with the habits of a social group, defined in time and space. Though it is difficult to define uniformly what we mean by knowledge, we are arguing here that it is related to some way of processing information, or attributing adequacy or justification to it. In our present discussion, we are particularly interested in the kind of knowledge that is determined as a social product, created by accessing to information through the digital communication technologies, yet it is likely that a large amount of knowledge belongs to this category. When we look for information on the web, we are asking our mediators (search engines) to provide us with information, which ought to be “relevant” to the subject of our query. The principle of relevance, that is so important for the sorting out of useful results in a search (Belew 2000: 304-305), has a deep relationship with the context-sensitivity of language games used in certain situations to convey meaning to a specific sentence, as Wittgenstein envisaged. The role of mediators consists in finding out the socially relevant results according to all sort of different contexts, but their activity implies a twofold delicate status for them

because, while evaluating and delivering a context relevant list of outcomes, they have, among their various tasks, that of the production of the social context and of that socially established meaning.

5 The sociality and the topology of the network

The computer itself drives it origins from a symbolic and formal representation of information influenced by the logic of the 1930s (see Turing 1937). Its computation model works as an isolated agent, serially manipulating formal symbols according to a table of unambiguous instructions, without any relation with the external environment. The network (ARPANET, the ancestor of the Internet) was built to avoid isolation and let the scientists, working in different laboratories, communicate with each others and share the same resources. Later, the web was another step forward. Tim Berners-Lee imagined it as a big author-system in which all people could contribute and interact with each other's work, creating links between pages. According to the theory of networks, it seems that the web is highly intertwined. The web seems to behave like a dynamic ecosystem in which pages are continuously born, change their address, get removed, etc., following a distribution based on unwritten power laws (Barabási 2002: 67). Though distributed in the various hosts of the network, information is not accessible democratically due to the topology of directed graphs (Broder *et al*. 2000), that produces the characteristic bow-tie structure of nodes. Only some "continents" are easy to surf, while some others are inaccessible to the unaware user. According to Peirce, the conception of reality implies the presence of a community without specific limits, always capable of a definite increase of knowledge (Peirce 1868). We can definitely consider the web as a tool to produce socially agreed reality and probably this definition represents well the spirit of the web revolution: a collective creation of a new dynamic communication space.

The Notre Dame University research group, under the guide of Albert-László Barabási set up an experiment to map the web. Using the method commonly adopted in statistical mechanics, they analysed only a small fragment of the web and hypothesized the behaviour of the network as a whole, based on the assumption that it would have a similar topology in term of connections of nodes. Through this experiment, they obtained an astonishing result: on average, there were 19 degrees of separation between two randomly taken web pages (Barabási 2002: 32–34). As Barabási clearly stated:

> The most intriguing result of our web-mapping project was the *complete* absence of democracy, fairness, and egalitarian values on the web. We learned that the topology of the web prevents us from seeing anything but a mere handful of the billion documents out there (Barabási 2002: 56).

The fact that all information is theoretically available on the web is not a guarantee that it is easily accessible. According to the latest results of the theory of the networks, it seems that social networks (and the WWW is not an exception) are dominated by a few highly connected nodes, the so-called *hubs*, that can be considered as the strongest argument against the utopian vision of the cyberspace as an egalitarian arena.

Licklider and Taylor wrote a fundamental paper on the computer as a communication device (1968). They put forward a very optimistic scenario of technology enhancement in the human-machine interactions and in the relations between human beings, via the computer. Despite the bright future they foresaw, there was one final *caveat* note about the role of communication devices in society. The impact of the revolution would be good or bad depending on the availability of online content and services. If the access to online information were a privilege reserved to few people it might "exaggerate the discontinuity in the spectrum of intellectual opportunity", if it were a right for everybody the network will allow the population to "enjoy the advantage of 'intelligence amplification'".

We are still at same point on this delicate issue: if the information retrieval services allow a democratic access to most (if not all) of the available information, then the network will be an increase in the

intelligence possibility of the people, provided that everybody is allowed to access the technology. The role of the information filters could not be more delicate than that, and we have to concentrate on this objective of guaranteeing an open, distributed, social and cooperative access to the web.

6 The success of PageRank and its faults

Considering the structure of the web, users could experience many difficulties when they try to orientate themselves in order to find useful or relevant information; the organization and the finding of pages can frequently be frustrating due to the overwhelming and chaotic load of available data. That is why, search engines play the role of mediators for accessing information, recommending the "guided tours" to visitors and letting them enter or not in online resources. It is likely that users obtain their information via a search engine, and if a page is not listed in the results of a query, its location will remain forever unknown. According to the title of a recent paper on this subject, *Esse est indicato in Google*, existence is guaranteed only by the presence in the index of Google (Hinman 2005). This gives an enormous power and a huge responsibility to search engines, because they offer the *unique choice* of results, outside of which there is only chaos and incomprehensible noise. The questions that arise spontaneously from these premises are: how fair is this web "guided tour" by search engines, considering that we rely completely on it? Is a search engine presentation of the web trustworthy? Could we exclude that search engines are not biased by commercial reasons or casual failures in representing the integrity and the complexity of the virtual world?

We concentrate on Google, which is the most successful search engine so far[3], trying to understand the reasons for its power in retrieving data on the web. Its efficiency depends, beyond the huge number of indexed pages[4], mainly on its search algorithm, PageRank, which relies on the exploitation of interconnections of web pages, as a way to attribute authority to pages. This algorithm mixes a standard inverted index of all pages, treated as vectors of strings, with a system that allows the attribution of authority to a page according to the links that it receives from other pages. A link from page "A" to page "B" is interpreted as a "mark" given to the "B" page from the "A" page. The value of the link depends also on the authority of "A" and the definition of authority is recursive: the more a page is linked by authoritative pages, the more it is authoritative itself. Therefore, the success of Google depends mainly on its capability of taking into account the actual topology of the web, while ranking the pages (see Brin and Page 1998). Pages linked by many authoritative pages have more relevance credentials for the users. However, even if the system is adequate when the user is looking for hubs and "mainstream" pages, search engines results are less trustworthy when we look for "minority pages". When we seek for information that is not very well known, or not very popular, or even only written in a language that is not widely spoken, like Italian or Norwegian, for example, it is much more difficult to obtain it (see Cho and Roy 2004). Though minorities' protection is not at the center of the users' attention, it is a very relevant issue in the perspective of the creation of ecology of the web as a collective and distributed information tool. Thus, the ranking mechanism, that is the main reason for the "fitness" of Google performance, produces at the same time discrimination for all sorts of minorities present on the web (language minorities, scientific minorities, young communities, etc.). The solution to this problem, however, does not seem easy to find, considering the complex nature of information retrieval in a distributed environment such as the web.

Minorities visibility is not the only difficulty that we encounter when we use Google, or another similar tool, as our privileged door to the web; there is also the "Google bombing" or "link bombing" phenomena[5]. As an example, if we try to type in the query bar the string "miserable failure", we obtain as the first result President Bush's webpage on the White House website[6]. This is the outcome of the exploitation performed

[3] In March 2005 Google was the forth most accessed U.S. site according to Nielsen, with 60 million unique viewers. The result is striking also considering that according to Google more than 50% of their traffic is outside the U.S.

[4] According to the last corporate information there are more than 8 billion pages indexed in the repository: http://www.google.com/intl/en/corporate/facts.html.

[5] See http://en.wikipedia.org/wiki/Google_bomb for details and Google's response.

[6] http://www.whitehouse.gov/president/gwbbio.html

by the hackers of PageRank characteristics. To achieve this result they do not need to enter in the target page. They only need to create many pages, widely cross-linked with each others, using that term and connect them to the President Bush homepage. Notice that this activity is legal. In addition, it is not an isolated episode; all commercial companies can make the most of it to increase the ranks of their homepages relatively to the pertinent keywords (Gori and Numerico 2003).

The "second generation" ranking algorithms of search engines, such as PageRank, are based on measuring popularity of webpages, which is a successful strategy compared to the previous information retrieval methodologies, but it is still a partial representation of the web, in which only popular and well connected websites are highly ranked, and therefore visible. Moreover whatever technology is used by a search engine, the search process is not transparent, we never know the reason for each particular list of results (Hinman 2005: 22), and up to a certain level nobody is aware of all the technical evaluation steps that produced a specific outcome list in reply to a particular query. We are, therefore, condemned to ignorance about technical unintentional biases of the source. These biases are more dangerous than the deliberate ones because they are less detectable, so they require a more refined defense strategy.

7 Database information retrieval versus association retrieval strategies

In the history of information retrieval, we never dealt with unstructured data. The key process of a database creation was the interpretation of data, in order to organize them in a precise and manageable way. Steve Lawrence and Lee Giles, two of the most influential scientists in the field of search engines' performance, declared:

> The WWW is a distributed, dynamic and rapidly growing information resource that present difficulties to traditional information retrieval technologies. Traditional information retrieval software was designed for different environments and has typically been used for indexing a static collection of directly accessible documents. The nature of the web brings up questions such as: can centralized architecture of the search engines keep up with the increasing number of documents? Can they update their databases regularly to detect modified, deleted, and relocated information? (Lawrence and Giles 1999: 117-118)

Considered from this point of view, the creation of a database and its transformation of the data into a meaningful whole is not what we need when searching the web. Being a dynamic, complex self-regulating structure, the network would need more associative distributed retrieval strategies, capable of creating "trails" of meanings and of memories, than old tools coming from the long experience of large databases management tradition. This is clearly acknowledged by some of the major experts in the area:

> Many of the search engines use well-known information retrieval (IR) algorithms and techniques. However, IR algorithms were developed for relatively small and coherent collections such as newspaper articles or book catalogues in a (physical) library. The web, on the other hand, is massive, much less coherent, changes more rapidly and is spread over geographically distributed computers. This requires new techniques or extensions to the old ones, to deal with gathering information, making index structures scalable and efficiently updateable, and improving the ability of search engines to discriminate (Arasu *et al.* 2001: 2–3).

In addition, the members of Google headquarters admit that there are still difficulties to overcome in this area. In a recent paper, Monika Henzinger[7] and Steve Lawrence declare that:

> There are still many open problems and areas for future research. […] The problem of uniformly sampling the web is still open in practice: which pages should be counted, and how can we reduce biases? Web growth models approximate the true nature of how the web grows: how can the current models be refined to improve accuracy, while keeping the model relatively easy o understand and to analyze? Finally, community identification remains an open area… (Henzinger *et al.* 2004: 5190).

We can conclude that, now, the web search techniques are not capable of taking into account some of the peculiarities of the complex data organization of the network and risk to misinterpret information and conceal some relevant data, so that they will remain undiscovered for the unaware user[8]. Although web pages have a tag structure that allows a sort of database organization for information, it is inopportune to describe the web's complex topology and structure[9], using a traditional database, whose nature has much in common with the classic logic approach to knowledge representation. Fitting the web into a database means to forget the centrality of its associative, collective and narrative character, as Bush already anticipated in 1945. The database will be only an abstract snapshot of the web content that cannot preserve its integrity, its nature and, above all, its soul.

Moreover, research is a unique term used to identify a variety of activities: search through keywords, different methods to analyze structured and non-structured data, flexible ranking mechanisms, peer results evaluation, content organization, automatic, rule-based, machine-learning classification, relational taxonomies, taxonomy generation, social knowledge management, adaptive ranking based on social choices etc. (Bawa *et al.* 2003). All these different processes not only require diverse strategies to be properly performed, but are aimed at generating complementary results. For these reasons, no matter how efficient a search engine is, it is highly recommendable to invent a wide range of means to gather data from the web. There is an urgent need to invent and apply new methods for retrieving information that are more respectful of the spirit of the WWW revolution and more adequate to the collective, dynamic, distributed and analogy-based nature of the web.

We can observe that all the present centralized searching methods, based on information retrieval techniques share many characteristics with the logic approach that we have already described. Therefore, following Bush's vision, we can argue that it is necessary to work out new searching methods, adequate to the associative nature of the web. In order to activate the "memory- based", intelligence augmentation techniques that support a more creative and collective approach to information, we have to work on a different model of information management that takes into accounts the creative role of connections between different pieces of information. This method could contribute to the construction of an appropriate "web of trails" in order to help the human memory to increase its power and its control over the web, despite the fact that neither of these approaches could guarantee the exhaustiveness of the active associations. Searching methods used by Peer-To-Peer (P2P) networks are a promising new approach to the field of dynamic and associative retrieval strategies (see Androutsellis-Theotokis and Spinellis 2004 for more details on P2P networks and their searching methods). They provide dynamic replies to queries, interrogating in real time the resources temporary available within the active nodes of peers connected in a certain moment. Serendipity should be the inspirational principle of these new retrieving techniques,

[7] Monika Henzinger was a Director of research at Google until the end of 2004. From January 2005 she is professor at the École Polytechnique Fédéral of Lausanne.
[8] The discussion of the Invisible Web is outside the scope of this paper, for further information see (Sherman and Price 2001).
[9] For more information about the web topology according to power law degree distribution see (Faloutsos *et al* 1999).

whose major aims are the augmentation of creativity, the management of collective memory about the available information and increase of associative power of the hypertexual structure of the web. The new research strategy should start with a rethinking of the finding process as a dynamic never-ending creative activity and not as a static and univocal response.

References

Androutsellis-Theotokis, S., Spinellis, D. (2004) A survey of peer-to-peer content distribution technologies. In: ACM Computing Surveys. Vol. 36 No 4, pp 335-371.

Arasu, A., Junghoo Cho, Garcia-Molina H., Paepcke A., Raghavan S. (2001) Searching the Web. In: *ACM* Transactions on Internet Technology. Vol. 1 pp 2–43.

Barabási, Albert-László (2002) Linked. Cambridge (Mass.). Perseus Publishing.

Bawa, M., Manku, G., Raghavan, P. (2003) SETS: Search Enhanced by Topic Segmentation. In: Proc. of the 26th Intl. ACM Conf. on Research and Development in Information Retrieval (SIGIR). http:, , citeseer.ist.psu.edu, bawa03sets.html.

Belew, R. K. (2000) *Finding out about*. Cambridge. Cambridge University Press.

Brin, S., Page, L. (1998) The anatomy of a large-scale hypertextual web search engine. In: Computer Networks and ISDN Systems. Vol. 30 pp 107–117. http:, , citeseer.ist.psu.edu, brin98anatomy.html

Broder, A., Kumar, R., Maghoul F., Raghavan P., Rajagopalan, S., Stata, R., Tomkins, A., Wiener, J. (2000) Graph structure in the web. In: *Proceedings of the Ninth International World Wide Web Conference* (WWW9), Computer Networks and ISDN Systems. Vol. 33 pp 1-6. http:, , www.feralcows.org, tomkins, papers, www9, www9.html.

Bush, V. (1945, 1999) As we may think. In: The Atlantic Monthly, Vol. 176, July pp 101–108, reprinted in Mayer, Paul A. (Ed.) (1999) Computer Media and communication. Oxford. Oxford University Press. pp 23-36, http:, , www.ps.uni-sb.de, ~duchier, pub, vbush, vbush-all.shtml

Bush, V. (1959, 1991) Memex II. In: Nyce, James M., Kahn, Paul (Eds) (1991) From Memex to hypertext. Vannevar Bush and the mind's machine. San Diego. Academic Press. pp 165-184.

Cho, Junghoo, Roy, S. (2004) Impact of search engines on page popularity. In: Proceedings of the WWW2004. New York. ACM. pp 20–29.

Faloutsos, M., Faloutsos, P., Faloutsos, C. (1999) On Power-Law relationships of the Internet Topology. In: *Proc. of ACM SIGCOMM*, Aug, pp 251–262. http:, , citeseer.ist.psu.edu, michalis99powerlaw.html

Gori, M., Numerico, T. (2003) Social network and web minorities. In: Cognitive Systems Research. Vol. 4, pp 355–364.

Foucault, M. (1994) The art of telling the truth. In Kelly M. (Ed.) Critique and Power. Cambridge (Mass.). MIT Press. pp 139-148.

Henzinger, M., Lawrence, Steve (2004) Extracting knowledge from the World Wide Web. In PNAS. Vol.101, April, pp 5186-5191. http:, , www.pnas.org, cgi, doi, 10.1073, pnas.0307528100

Hinman, L. M. (2005) Esse est indicato in Google: Ethical and political issues in search engines. In International Review of Information Ethics. Vol. 3 pp 19-25. http:, , www.i-r-i-e.net.

Lawrence, S., Giles, L. (1999) Searching the Web: general and scientific information access. In: IEEE Communications. Vol. 37, No 1, pp 116-122.

Licklider, J. C.R., Taylor, R. W. (1968) "The computer as a communication device, In: International Science and Technology. April, pp 21-41. http:, , memex.org, licklider.pdf

Peirce, C. S. (1868) Some consequences of four incapacities. In: Journal of Speculative Philosophy. Vol. 2 pp 140-157. http:, , www.peirce.org, writings, p27.html

Sherman, C., Price, G. (2001) The invisible web: uncovering information sources search engines can't see. Medford. Information Today.

Turing, A. M. (1937) On Computable numbers with an application to the Entscheidungsproblem. In: Proc. London Mathematical Society. Vol. 42, No. 2, pp 230-265; reprinted in Davis M. 1965 The Undecidable, New York Raven Press. II Ed. with amendments, New York. Dover Publications. 2004. pp 116-154.

Wittgenstein, L. (1953) *Philosophical Investigations*, [*Philosophische Untersuchungen*] translated by G.E.M. Anscombe, Oxford. Basil Blackwell .