Bernhard Rieder:

Networked Control: Search Engines and the Symmetry of Confidence

Abstract:

Search engines have become an integral part of our Internet use. They shape the way we look at the world, they provide orientation where there is none; but the maps they draw are too often hijacked by commercial interest. Search engines are less *black box* than *black foam*; functional decoupling, parasite technologies, and the embedding in the greater context of culture and society render the *search act* subject to overdetermination. Control is thus diluted into a dense network of human and non-human "actants" and the power of the search engine is located in a *control zone* rather than a control center. In order to shift power back to the public, this paper proposes the concept of "symmetry of confidence", a new relationship between search engine companies and their users.

Agenda

Author:

Bernhard Rieder:

- Département Hypermédias UFR-6 / 2, rue de la liberté / 93526 Saint-Denis Cedex 02 / France
- ☎ ++ 33 1 40 09 11 54 , ✉ bernhard.rieder@univ-paris8.fr, 💻 http://procspace.net/berno/
- Relevant publications:
  - Algorithmische Mediatoren: Software-Agenten als Gegenstand der Medienwissenschaften. Medienjournal 1/2004 p.36-46
  - Sphère publique et espaces procéduraux. With Michaël Thévenet. Bordeaux, 2005 (forthcoming)

ISSN 1614-1687

In the middle of the nineties, when the Internet first made its entrance to a larger audience, the debate on the social, political and ethical dimension of the global network centered around two main issues: first, the question of basic access to technology and second, the effects of a global non-broadcast media on the functioning of the public sphere. After the turn of the decade and especially after the burst of the economic bubble, research on the Internet is diversifying at a rapid pace; research questions are becoming both more specific and precise – answers contain a lot more nuance. With over 700 million users, the Internet is now an established part of industrial society and the debate on access is quickly shifting from the general question of admittance to the technology in general to the problem of access to specific information once inside. As the *gatekeeper* of the digital age, the search engine has come under special scrutiny in the recent years.[1]

The latest Pew study[2] on the topic suggests that search engines are a central part of how people use the Internet. They have become *institutions*: the interactive mapmakers that chart the unstructured geography of the vast data environment that is the Web. When scouting for new information, there is practically no way around the search engine and as the Internet has become part of our daily lives, so have they; and the maps they draw are less and less representations of the public sphere, but charts of the commercial landscape. In a democratic society, the concentration of power automatically raises a series of questions and in order to gauge the size of the problem, we first need the conceptual tools to understand the phenomenon - only then can we propose a course of action. Due to the unusual complexities of the role search engines play, our understanding is still in the early stages.

This paper will add some thoughts to the discussion by making three arguments: 1) conceptualizing search engines as black boxes is increasingly inaccurate and will be more so with further technical advancement; 2) our perspective on power and control must adapt to our hybrid condition; 3) a theoretical and practical shift in our conception of the relation between user and developer should be a key element in an ethical and political stance on the question. Being a researcher as well as a developer, I will try to build these arguments on a perspective that is based on both on a technological and a cultural theory viewpoint. In a time where activities that were formerly reserved to human agents are getting automated in computer code and being delegated to machines at a fast pace, the study of technology becomes indispensable for our understanding of the forces that shape culture and society.

The search engine: from black box to black foam

The ongoing debate on the political and ethical implications of search engines[3] does rarely provide a definition of its subject; some years ago it was AltaVista and now it is Google that plays the role of a convenient *pars pro toto*. This focus on the dominant player in the market makes it difficult to generalize the technical and morphological aspects of the object of study and to understand the functional and representational choices each company makes. But it is only through a closer look on these choices that we can discuss the significance for society and culture and propose a course of action at the same time. Power structures are not confined to the social realm; they also operate inside of technical artifacts and to decipher them, we need to look at these artifacts themselves. I will therefore suggest a quick definition of the term: a (Web) search engine is a piece of software that creates an *index* of a defined set of *data*, includes a *retrieval technique* to access that index and uses a specific *mode of representation* to display the results.[4] Following this definition, we can identify four distinct conceptual layers, where each one puts the developer before a series of choices:

- *Data:* What is the scope of application? A local site/database or the entire Web? Is the data unstructured, pre-structured or structured? What is relevant? How do we extract it?

[1] E.g. Gerhart, Susan L.: Do Web Search Engines Suppress Controversy? or Introna, Lucas D. / Nissenbaum, Helen: Shaping the Web.

[2] Fallows, Deborah: Search Engine Users.

[3] While search techniques are as old as computers, the term "search engine" has come to refer specifically to retrieval software on the World Wide Web. In this article, I will use it in this context.

[4] For an introduction to the larger field of information retrieval see Chu, Heting: Information Representation and Retrieval in the Digital Age.

- *Index and indexing technique:* How is the index structured and what are the criteria to be taken in? How much of the data space is covered by the index? What is the common rate for updates?

- *Search and retrieval:* How do we query the application and how is the query related to the index? What are the criteria for relevance? In which order should we rank results?

- *Representation:* In which form does the application present the results? As a list? A clustered list? A map? A tree? A 3D-sphere?

Taken together, these four layers trace the *morphology* of a search engine – a series of choices for the developer as well as a series of questions for the investigator. Every search engine gives a particular answer on each one of those levels. Commercial success as well as political impact depends on it. While making up a functional whole in the eyes of the user, the four layers are in general built as largely independent modules rather than a monolithic application, and specialists in research and development are working on the specific problems and difficulties encountered on their level, which constitutes in fact a distinct field of research. There is actually no technical reason for packing all four layers into one application and we are already seeing specialization and diversification in the area. "Result browsers" display search engine query results using a different type of representation: Touchgraph's GoogleBrowser[5] transforms the top-down list into an animated network of nodes and Google News Map[6] projects the events from Google's news syndication service[7] onto a world map. Several research projects use post-processors to re-rank search results to add features such as weighted search[8] or document comparison based on vector models[9]. There are companies like Sensoria, whose product iSearch[10] is a basically a highly sophisticated neuromimetic search algorithm that can be used with any index and various forms of representation. Even the optimization of a Web page in order to increase its place in result ranking can be seen as part of the search process.

Such forms of modularization, functional decoupling, and "parasiting" are part of the history of systems design: while the first computers ran one program at a time, machines nowadays are a prolific environment for hundreds of processes and components, each one performing a different task. The four subsystems that constitute a search engine are therefore embedded into other, established structures that range from programming languages and frameworks to the operating systems and database applications that lay the ground for everything else. In analogy to the history of human societies, the ever more complex organization of information systems continuously produces specialization and division of labor and today, a computer may very well be called a "society of processes"[11]; habitants are highly dependent on each other and the functioning on the whole cannot be reduced to the actions of an individual.

The trend to ever-increasing organizational complexity forces us to reassess the images and metaphors we use for information technology. Search engines have often been called "black boxes"[12] – we cannot see inside (they are protected both by technical and legal door locks), the only way to judge the mechanism is therefore to analyze input and output. But the metaphor of the black box implies that we still have a clear picture of the outside shape of the object; there still *is an object* and we know where it starts and where it ends, we can clearly identify input and output. But the label is becoming increasingly inaccurate. The functional decoupling at the inside of a search engine and the integration of the system into a larger technical environment make it nearly impossible to gauge how many subsystems are actually involved in the search process. The neatness of the box is long gone; what we look at is the burgeoning assembly of black bubbles that form

[5] http://touchgraph.com/TGGoogleBrowser.html

[6] http://douweosinga.com/projects/googlenewsmap

[7] http://news.google.com

[8] For example the Fetuccino "search parasite" described in Ben-Shaul, Israel et al.: Adding support for dynamic and focused search with Fetuccino.

[9] My own project "procspace" is an example (http://procspace.net).

[10] http://www.influo.com

[11] This idea was first explored by Marvin Minsky and has practically invaded the AI community in form of multi-agent systems.

[12] E.g. in Winkler, Hartmut: Search Engines. 29

an amorphous mass: *black foam*[13]. How many layers of processing lead from the manipulated *metatags* on a webpage to the *clustermap* the user interacts with when making a search request? Through how many subsystems does the search query pass and what do they add to the result? Where does the "system" start and where does it end? There is no longer a clear answer to these questions. Functional interdependence and technical layering will only continue to grow and with search algorithms that are built on probability mathematics and connectionist approaches, even developers have no way to predict how a system will perform in a given situation.

And this is only the technical side of the process. But information systems are neither god-given nor self-sufficient. At both ends of the chain we find human beings, at the bottom as developers, system designers and information scientists and at the top as users; and in the middle there are people who optimize Web pages for optimal ranking and other developers that create meta-searchers, post-processors and parasite interfaces. All of those human beings are of course deeply embedded into the dense networks of culture and society. Taken all together, we see a great number of human and nonhuman agents that make up the *dispositif* that structures the terrain for what we might call a "search act"[14]. If we want to know how the search engine's power operates, we have to start from this hybrid complexity and cope with *overdetermination*.

The Question of Power

We are only at the beginning of our theoretical grasp of such very complex socio-technical systems. Technology has never been neutral but it is only with the computer entering the cultural practices that are so intimately tied to the production of meaning that we actually start to understand what that might actually mean. It is probably Bruno Latour that went the furthest in theorizing the hybrid practices performed in networks of human and non-human "actants". Latour goes as far as proclaiming that "action is simply not a property of humans *but of an association of actants*"[15]. Adapted to our question, it means that when a surfer uses a search engine, the human and the non-human fusion into a third, a hybrid actant that is more than the sum of both. Behind them lies the even larger, hybrid network described above: every actant, no matter if human actor or technical subsystem, plays its role in determining the outcome. The responsibility for the results cannot be labeled back to one of the components. We leave both technical and social determinism behind – at the price of loosing a stable point of origin for causation. If we take Latour's perspective seriously, the question of power suddenly becomes very complicated: "Responsibility for action must be shared among the various actants"[16]. And, as I have tried to show, there is a great number of technical and human actants at work in the *black foam* surrounding the search act and control is effectively diluted into the dense network they make up. The political choices (e.g. through ranking techniques) developers can make are actually part of a much larger, distributed space of possibility and we should not think of control *centers* but rather control *zones*.

Power runs through the capillaries of this network and with reference to Foucault[17] we have to understand power as a *productive force*, not as an inhibitor. Search engines are best understood when seen as producers, not as censors. Their product is a perspective, a topology, a map on the chaotic territory of the Web. By ranking search results, they offer a concept of *what is important* and what is less so. They are *vision machines*[18] that not only extend our perception into the masses of information that would normally be far beyond human scope, but that also *interpret* the environment they render visible. The functional morphology embedded into the four layers of a search engine might not work the same way as a human perception and interpretation, but it is nonetheless a semantic model that implies a perspective of what things *mean*. Google's

[13] The metaphor of foam has been recently explored – in a very different context – in Sloterdijk, Peter: Sphären III.

[14] The analogy to the term "speech act", first coined by Adolf Reinach and later by John Austin, intends to emphasize de pragmatic context of information search. Due to space restrictions, this line of though must be explored elsewhere.

[15] Latour, Bruno: Pandora's Hope. 182

[16] Latour, Bruno: Pandora's Hope. 180

[17] Foucault, Michel: Histoire de la sexualité I.

[18] Virilio, Paul: La machine de vision.

PageRank[19] algorithm for example is built on the assumption that every link to a site is also a popularity vote and that sites that get a linked to a lot must be very important. This is of course a sociological assumption and whether right or wrong, it implies a view of how society works. And this view is effective in every one of the millions of search operations processed each day.

There is at least one major difference between a human gatekeeper (or better: viewshaper) – a journalist for example – and an algorithmic one. The journalist is deeply situated in the culture she is working in. She is able to judge a source of information using probably hundreds of micro-criteria (some of which may very well be subsymbolic in nature) and it is clear that a PR brochure from a company will not be treated the same way as a communiqué from the United Nations. The quality of the human journalist is her *subjectivity* – her being a subject of a culture – which doesn't mean that she is not *balanced* in her work. The algorithmic gatekeeper does not have this level of immersion in culture necessary for deep semantic operation. While some level of adaptation is possible, search engines use a "one size fits all" approach: in order to produce their hierarchies, they have to decide on a set of criteria and parameters (like PageRank) that will be used on all of the analyzed data. As a result, one perspective will be favored over the others and this *worldview* is not based on the adaptive interpretation of a human being but on a short series of parameters mechanized in the form of an algorithm with little or no capacity to adapt to context. Commercial actors have the resources to adapt their Web content to the common criteria that decide on visibility and they have already hijacked large zones of the keyword terrain. Search engines have become agents of commercial interest.

But despite this critique, we have to understand that "there is no such thing as digital information *without* filters"[20], that there is not *outside* of power. The whole idea of the search engine is about providing orientation where there is none or very little and this implies higher visibility for some and less visibility for most. Foucault taught us that knowledge (and a search engine can be seen as producer of knowledge) is intimately intertwined with power and it is very clear that a commercial enterprise will chose a worldview that does not contradict the power structures of the market.

The Symmetry of Confidence

We are faced with a rather paradoxical situation: on the one hand side I have argued that search engines are powerful *vision machines* that provide cultural orientation, mostly in favor of economic interest; but just before, I suggested that information systems and the hybrid networks that surround them make it impossible to attribute accountability to a precise agent in the chain. So there is power, but nobody has it. Political and ethical choice depends however on our capacity to act and my argument for a distributed understanding of control seems to make effective action extremely difficult. At the same time we seek answers to the question of how can we guarantee that the model of knowledge, the worldview every search engines implies is compatible with the democratic values of plurality and equality and not just another outlet of special interest? Introna and Nissenbaum have pointed out[21] that the Web is a public good and that commercialization and centralization of information access through search engines is endangering the Web as an egalitarian space for civic communication and representation. They appeal to humanitarian values of fairness and restraint and urge the makers of search engines to keep an egalitarian outlook. While business ethics may be part of the solution, it is clearly not enough. As the already mentioned study[22] of the Pew project suggests, we use search engines – despite all the problems and reservations – with great confidence. It is time that this confidence was mirrored back to us.

Reduced accountability through hybridization of control and the dilution of power into a network of actants on the one hand, and the immense act of confidence in which we delegate part of our perception to search engines on the other, lead to a possible answer to the problem: the notion of "symmetry of confidence". What does this mean? I propose that instead of asking (search engine) companies not to be commercial actors, we should build on the ongoing process of modularization in order to shift more control to the public. Dilution of power does not

[19] http://www.google.com/technology/

[20] Johnson, Steven: Interface Culture. 38

[21] Introna, Lucas D. / Nissenbaum, Helen: Shaping the Web.

[22] Fallows, Deborah: Search Engine Users.

entail even distribution; there are *zones of power* and the current concentration in the zone of commercial interest can and should be countered by strengthening civil society. While it would be desirable to develop "more egalitarian and inclusive search mechanisms" as Nissenbaum and Introna suggest[23], devising policy for such a goal would be difficult and highly problematic from a political standpoint: what are the "good" values and how do we legislate them into the market? And how would we keep the commercial actors from quickly adapting their content to the new "egalitarian" search algorithm?

Instead of trying to mechanize equality, we should obligate search engine companies to perform a much less ambiguous public service by demanding that they grant access to their indexes and server farms. If users have no choice but to place confidence in search engines, why not ask these corporations to return the trust by allowing users to create their own search mechanisms? This would give the public the possibility to develop search algorithms that do not focus on commercial interest: search techniques that build on criteria that render commercial hijacking very difficult. Lately we have seen some action to promote more user participation and control, but the measures[24] undertaken are not going very far. Still, from a technical point of view, it would be easy for the big players to propose programming frameworks that allow writing safe code for execution in their server environment; the conceptual layers already are modules and replacing one search (or representation) module with another should not be a problem. The open source movement as part of the civil society has already proven it's capabilities in various fields and where control is impossible, choice might be the only answer. To counter complete fragmentation and provide orientation, we could imagine that respected civic organizations like the FSF[25] endorse specific proposals from the chaotic field of search algorithms that would emerge. In France, television networks have to invest a percentage of their revenue in cinema, why not make search engine companies dedicate a percentage of their computer power to algorithms written by the public? This would provide the necessary processing capabilities to civil society without endangering the business model of those companies; they could still place advertising and even keep their own search algorithms a secret. But there would be alternatives – alternative (non-commercial) viewpoints and hierarchies – to choose from.

Conclusion

This paper started out by arguing that search engines have become more like *black foam* than black boxes. Their highly complex hybrid technical and social composition renders clear delimitations impossible, and overdetermination dilutes power from control centers to control zones. In order to reduce commercial hijacking of search engines, we need to strengthen civil society; one way to do so would be to open the server farms or search engine companies for code written by the open source community.

This symmetry of confidence is not a concept for abolishing power structures or capitalism; it proposes a different *zoning* of power by shifting some part of control over the *vision machines* back to the public. If search engines shape the way we look at the world, the public should have the right to shape them in return.

References

Ben-Shaul, Israel et. al.: Adding support for dynamic and focused search with Fetuccino. Proceedings of the Eighth International Conference on the World Wide Web, Toronto, 1999 p. 1653 - 1665

Chu, Heting: Information Representation and Retrieval in the Digital Age. Medford, Information Today, 2003

Fallows, Deborah: Search Engine Users. Washington, Pew Internet & American Life Project, 2005 (http://www.pewinternet.org/PPF/r/146/report_display.asp)

Foucault, Michel: Histoire de la sexualité I. La volonté de savoir. Paris, Gallimard, 1976

Gerhart, Susan L.: Do Web Search Engines Suppress Controversy? First Monday, Vol 9, No 1, 2004

Introna, Lucas D. / Nissenbaum, Helen: Shaping the Web: Why the Politics of Search Engines Mat-

[23] Introna, Lucas D. / Nissenbaum, Helen: Shaping the Web.

[24] Msn search now features rudimentary user control over ranking criteria and Google grants machine access to its search (through the SOAP protocol) but limits it to 1000 requests per day, rendering effective re-ranking impossible.

[25] Free Software Foundation, http://www.fsf.org

ters. The Information Society, Vol. 16, No. 3, 2000

Johnson, Steven: Interface Culture. How New Technology Transforms the Way We Create and Communicate. New York, Basic Books, 1997

Latour, Bruno: Pandora's Hope. Essays on the Reality of Science Studies. Cambridge MA, London, Havard University Press, 1999

Sloterdijk, Peter: Schäume. Sphären III - Plurale Sphärologie. Frankfurt am Main, Suhrkamp, 2004

Virilio, Paul: La machine de vision. Paris, Galilée, 1988

Winkler, Hartmut: Search Engines: Metamedia on the Internet? In: Bosma, Josephine et. al. (Eds): Readme! Filtered by Nettime. ASCII Culture and the Revenge of Knowledge. New York, Autonomedia, 1999 p. 29-37