

Tobias Blanke:

## Ethical subjectification and search engines: ethics reconsidered

### Abstract:

This article will explore the relation of search engines to the freedom they invoke in human subjects. Away from questions about the social impact of search engines and their ethical use, it shall investigate the influence of search engines on ethical subjectifications. The article will criticise the common critique that search engines should only deliver neutral and objective results to their users, where 'neutral' and 'objective' are defined as anti-subjective. On the contrary, it will argue that search engines are designed to deliver subjective results. A possible ethical critique starts therefore where they fail to do so. Due to reasons immanent to the technology, search engines are never subjective enough in their relevance decisions. Their results collide at the same time with what their users expect them to deliver. The article will show that, far from being a disadvantage, this disagreement between the users' expectations and the search engines results is what triggers an ethical subjectification.

### Agenda

Introduction .....	34
Relevance decisions – Simulating the users' minds.....	35
The production of ethical subjectivity.....	37

### Author:

Tobias Blanke:

- Organization and contact address: Department of Computer Science, Queen Mary University of London, London, E1 4NS, United Kingdom
- Telephone, email and personal homepage: ✉ tobias@dcs.qmul.ac.uk, 🌐 <http://www.dcs.qmul.ac.uk/~tobias/>
- Relevant publications:
  - Hegels Artificial Intelligence, In: Hegel Jahrbuch (2005), Forthcoming, Berlin: Akademie Verlag 2005, 10pp.
  - Meta-evaluation of XML retrieval using Situation Logic, Master Thesis University of London, London, 2004, 101pp.

## Introduction

This article will explore the relation of search engines for the internet to the freedom they invoke in human subjects. On the one hand, the internet is representative for a concept of freedom in the very early sense of the Enlightenment; on the other it evokes well-known reactions to anarchic Enlightenment that waver between the desire for freedom and control over freedom. The freedom the internet gives rise to has been valued and praised but also feared. It is virtually impossible to control this freedom at the source. It does not take much effort for users to present any views or materials online. Search engines often take the blame for the misuse of this publishing and freedom of information, as they are the entry point to the web's freedom. They are therefore faced with the dilemma of unlimited freedom versus controllable freedom. Ethical arguments have been invoked for the control the information on the World Wide Web at its gates. New legislation cannot really help, as the web is not bound to national borders or responsible to a lawful sovereign. With modernity, if freedom and its subjects cannot be controlled by laws, moral or ethical control has provided a solution. Search engines are as such an ethical problem, because they confront us again with the modern ambivalences of ethical subjectification and its traditional question of how to make subjects behave ethically. How to control freedom by means of ethics is the question for ethical investigations of the new web technology.<sup>1</sup> This article will propose a different approach towards the complex relation between ethics and modern information technology.

Inspired by Norbert Wiener's proposal in his classical Computer Ethics book, *The Human Use of Human Beings: Cybernetics and Society*, this article will depart from traditional ethical approaches used to

assess the ethical impact of search engines.<sup>2</sup> Away from questions about the social impact of search engines and their ethical use, it shall investigate the influence of search engines on ethical subjectifications. Applied ethics is not just ethics applied by using lasting ethical ideas to gauge human behaviour in society and throughout history. It is not enough to look at the subject's ethical consciousness before and after the interaction with technology and afterwards decide on the nature of change based on situation-independent ethical values. To break loose from this type of applied ethics, it is crucial to consider the phenomena themselves and not be restricted to universally applicable ethical values. Looking at the ethical development of subjects under the influence of search engines does away with the external viewpoint. In exchange, it requires knowledge about the technology and its functioning, given that analysis starts with the phenomena. Sometimes a dispute in an argument in applied ethics is just a dispute about facts. This paper will contend that such an approach is more appropriate to explain the influence of technology in general and search engines in particular upon the processes of ethical subjectification.

In order to develop the impact of search engine technology on ethical subjectifications, the article will proceed in two stages. Firstly, it will look at how search engines decide over information. The article will demonstrate why a common ethical critique of search engines misses its target. One should not expect search engines to deliver only neutral and objective results, as the technology is not designed to do so. Secondly, it discusses the problem of reducing ethical freedom to having total information. The other argument in the common ethical position towards search engines, as criticised in this paper, is the demand not just for neutral, but also for complete information. The second part will show that an ethical decision does not derive from what is already known, but what remains to be discovered. Search engine technology will be considered in the context of the area in computer science dealing with it, Information Retrieval (IR). Here search engines are only a small field of applications and among them web search engines are the most famous. They open the way into the internet and make the internet a truly 'worldwide web'. Without them, only field experts would be able to find information, as an overview of the location of relevant sites would be a

---

<sup>1</sup> Please compare the report on 'Deutschlandfunk' about a conference „Zur Ethik und Ökonomie von Internet-Suchmaschinen - Eine Tagung der Bertelsmann-Stiftung in Berlin“. The summary is, that the world inhabitants are to be trained with ethics on the use of the World Wide Web. Deutschlandfunk, Ethik und Ökonomie von Suchmaschinen.

---

<sup>2</sup> Wiener, Norbert: *The Human Use of Human Beings*.

prerequisite for information retrieval. Search engines are the path of access for the general public into the internet world. A closer look at search engine technology will reveal that the ethical problems of search engines commence with their decision about the relevance of an information for the rest of the world.

## Relevance decisions – Simulating the users' minds

Possible ethical problems with search engines are often identified with those of the web, when they are blamed for the content they return to an information need. If the source cannot be blamed, the entrance is to be controlled. Such an approach runs into the danger of disregarding the specificity of search engines. The technology itself cannot be blamed for the content it has to consider. An item is a datum for a computer and becomes information only for a human being able to interpret it. Search engines return in this sense data and do not know anything about what this data represents. Computers are symbol-manipulating machines; they do nothing else but substitute symbols with other symbols and process data without real knowledge about what it means. Search engines work the same way. Though they are meant to retrieve information, they actually retrieve data. Computer science only calls them information retrieval engines as their matching is not absolute. Data retrieval done by database management systems relies on clearly defined objects or definitions and follows the rule: Either every object or none can be retrieved. The search conditions express necessarily and sufficiently what could possibly be retrieved by them. One incorrect match in thousands of retrieved objects means simply a total failure. Information retrieval however looks for 'relevant' information, which can still be there even if the data is inaccurate. Relevance is defined with respect to the computation of the information an object contains about another. Every user of search engines experiences relevance in the form of the order of the return list. It is with relevance that the more technologically specific ethical problems of search engines begin.

Technologically, relevance decisions of search engines work amazingly well if one takes into consideration the amount of data they have to deal with in real time. The computer makes a guess on the content of the documents. As it cannot understand what is written, its guess is not based on under-

standing. It will use statistics and calculation, its only source of information,<sup>3</sup> and start counting words, word phrases etc. By means of term frequencies, search engines estimate how useful this document could be for a query. If document A repeats the query's words more often than document B, then it will have a higher relevance.<sup>4</sup> Rather basic statistics produce what is for the users the 'mysterious' ranking of search engines. "Google" as the nowadays most famous example will take us closer to the ethical problems behind the ranking. It produces pragmatically a good match quickly. The full details of its algorithm are not known because of property rights, but an early paper of the two Google founders indicates that it is based not only on word occurrence statistics but also on a system of authorities and hubs.<sup>5</sup> Authorities are web pages that are linked by many others, while hubs link themselves to many other pages. Web pages achieve a better ranking if they optimise their relation within this system of hubs and authorities. With this algorithm, Google simulates information seeking strategies from academic contexts. In academia, it is a good habit to start research by browsing through the material referenced in an initial paper.

The Google system obviously performs so well that most users choose it instead of other systems. Google started off as one of the preferred engines among computer experts and has today almost a monopoly, as it returns reasonably good information very quickly without charge. Even if its market penetration and global dominance can be another ethical (or rather political-economical) problem, for the purposes of this article, Google is chosen as an example of how search engines work, used to investigate the impact of the basic technology upon

---

<sup>3</sup> For reasons of simplicity, the argument will be restricted to the understanding of textual information. Graphical or other multi-media matching works similarly.

<sup>4</sup> Zipf's law says that only the least occurring words hold most of a text's particular meaning while the most occurring ones are repeated throughout many texts. Mathematically, the relevance of a word is indirect proportional to its frequency. Modern information retrieval is based on this law by discarding the most and the least occurring words as not very discriminating.

<sup>5</sup> Brin, Sergey and Page, Lawrence: The anatomy of a large-scale hypertextual Web search engine.

ethical subjectifications. It is more important for the latter argument to look at the principles of how search engines deal with meaning and information. From this viewpoint the ethical problems of search engines do not begin with the fact that they decide about relevance but with how they decide about it. The technology has been developed so as to decide itself. Its decision is supposed to reveal the meaning in the data and simulate information. How is this meaning retrieval done? The first thing to notice is that it is always limited by the 'objective mind' of a machine. A search engine is designed to retrieve information relevant to a human's subjective situation. Therefore search engines are at the heart of what has been discussed for years now as the gap Artificial Intelligence will not be able to bridge.

It is the subjectivity of intelligence as a theoretical action by a human agent that makes it so difficult to simulate information and bridge the gap between the subjective and objective mind. Would intelligence be what behaviourists have thought, AI research would have already been much more advanced in making an artificial intelligence a pleasant partner to have a chat with.<sup>6</sup> Information is subjective and the aim of information retrieval is not to contradict this but to deliver something that satisfies a subjective information need.<sup>7</sup> First, however the engine has to learn to simulate the subjective decisions behind human beings' relevance associations. It has to simulate their subjective decision criteria by the objective means of mathematical manipulation. Ways of simulating the users' minds must be found. Above, I have described one example with Google. Its concept of authorities stems from what could be called the research expert option. In the praxis of research processes it is well known that the researcher trusts those information most, which come from reliable sources. One way to discriminate what a reliable source could be is to find out what other researchers quote most often. This would then be a research or information authority. Another researcher could be a hard worker so that the researcher will find in her papers many valuable links to other papers. The researcher will therefore often return to this research hub. To find a good combination of hubs and authorities is likely to

make the research process successful. Google works on this assumption that the subjectivity of its WWW's user is structured similarly to that of the professional researcher.

The target for search engines is to reproduce opinions as neutrally as possible. Developers therefore rely on the opinion of subject experts, hoping to find a neutral response matching the users' taste. While a search engine is produced, experts play an important role in the fine-tuning of parameters to deliver better results. In this sense, these subject experts are similar to Kant's artistic genius, who represents in her work something that everybody believes when faced with a structure that does not allow deducing what everybody has to think. Kant describes the genius as a talent to produce what cannot follow a rule. Just like Kant's genius the expert is supposed to deliver exemplary results.<sup>8</sup> As long as experts work for other experts something in a library for information about certain subjects, this approach may work. Nevertheless, the web search engine, the most successful information retrieval application, deals with the complete unknown average user, whose taste is different from that of experts. In this sense, the ethical problem with search engine is the form of subjectivity they simulate, which is not the subjectivity of everybody. The content they look at is not specific to them.

The key to understanding the relation between ethics and search engines is to do away that they are supposed to be neutral in their target. They attempt to be scientific in being unscientific and ask subject experts for that. This is a typical engineering move to capture something in itself impossible to capture. At the moment of developing a search engine, it cannot be known what users might think about its results, as the users are simply not present while writing the code or tuning the parameters. Even if experts were geniuses and always right in their decisions, they would still misconceive the taste of future users they cannot know. As engineering products, search engines are thus not undemocratic, rather but a-democratic. They do not intend to misconceive the preference of most people, but cannot ask most people for their opinions. Search engines are the realisation of a scientific process

---

<sup>6</sup> No machine has yet passed the Turing test and tricked a human judge that she is not talking to another human being.

<sup>7</sup> "Relevance is a *subjective* notion" Ceith Rijsbergen: Information retrieval, p. 146

---

<sup>8</sup> Immanuel Kant: Kritik der Urteilskraft, §49.

and not of a vote.<sup>9</sup> They remain to be products of engineering work and not of ethical or political action. Their aim is to help people find the data they need to satisfy their information need. They deliver meaning of texts without reading them. The next chapter will argue that it could even be regarded as an advantage in producing ethical subjectivity that they cannot provide a total view of the world.

## The production of ethical subjectivity

We have seen that the technical fact that different search engine implementations deliver different results cannot be an argument against their neutrality.<sup>10</sup> This does not make a search engine less objective and missing its targets. On the contrary, its target is to mirror subjectivity objectively. As every search engine attempts this in a slightly different manner relevance decisions are delivered differently. The critical question directed at the search engine does not attack its presupposed lack of neutrality in its results, but the form it seeks to represent the necessary a-neutral moment of subjectivity. Subject experts help to represent the unknown user. Demanding neutrality as such from search engines would end search engine technology and would mean confusing the aims of the scientific production process and the process itself. Non-rationalised subjectivity is supposed to be produced which involves different results for different search engines.

To use a famous quote from an even more famous scholar in computer ethics, Joseph Weizenbaum, what search engines still cannot do and probably will also not be able to do in the near future is to understand the content of what they retrieve and reflect that in their relevance decisions. Although

---

<sup>9</sup> Marcel Machill et al. Seem to make the argument that it should be the other way around. Marcel Machill, *Transparenz im Netz*.

<sup>10</sup> Please compare "Suchmaschinen – Bundestagsfraktion Bündnis 90/Die Grünen – 03/2005". Die Grünen: Suchmaschinen, pp. 5. Strangely enough, the paper seems to know exactly in advance what is relevant to users and what not. It often tells the reader that engines miss out relevant information, but does not show how it has come to these decisions about the engines' decisions.

current research has proposed several ways to introduce "semantics" into the syntax matching techniques of search engines, all of these are still limited. "Thesauri" for example will help the machine not get lost in similar meanings, but they will never be complete enough to help it understand the meaning.<sup>11</sup> Meta-information in different forms is only information used by the knowledge engineer or the authors of texts to summarise their understanding. In their present configuration, search engines cannot make the step towards understanding. Thus, they can never present a complete and sound overview of an information need. Their relevance decisions are never the only ones possible and they do not even attempt to be so. Research actually wants relevance decisions to be as subjective as the users' information needs.

The problem with search engines appears therefore to be of a different nature, namely that their decisions are not subjective enough. To fine-tune the parameters of the search engine's algorithm, subject experts are needed to decide which documents of a collection are relevant and which are not. Search engines therefore reproduce the subjectivity of experts. From an ethical point of view, one could argue that it is actually a good sign that users disagree with the output of search engines and do not take for granted what experts want people to believe about their subjects. There is no such thing as an unbiased relevance decision. Rather than an ethically problematic conclusion, as it has been largely advocated in current ethical and political debates, such relevance decisions constitute a valuable source of ethical subjectification nowadays by not simply reproducing what experts want people to believe about their subjects. The search engine's relevance decision – if correctly understood – offers the chance of ethical decision-making for subjects, who can with respects to ethics only be all the subjects and not a limited group like experts. Search engines open ethical considerations and autonomous decisions rather than foreclosing them by opening the knowledge of communities beyond that what is already known. The desire to have a complete overview of information is a result of the modern illusion of ethics as a rational choice made

---

<sup>11</sup> The use of thesauri still needs to prove that it will improve the effectiveness of a search engine significantly.

Grossman David A and Frieder, Ophir: *Information Retrieval*.

by the subject in transparent conditions. If one considers an approach to ethics inspired by Alain Badiou – where ethics is linked to subjective decision-making and ‘newness’ rather than reliance – the role that search engines play for ethical subjectifications appears in a different light.<sup>12</sup>

For Badiou, a subject becomes such as a result of an ethical process. It is not predefined as an ethical rational one, who only has to judge the quality of different purposes and actions. One becomes a subject in the process of ethical actions. The subject does not need to have total information before she engages in action. It is even not desirable that the subject should have a clear standing in the world. The fact that one becomes a subject through a process means that one redefines and re-configures one’s position while having a strong trust into its own ability to deal with the world. Ethical action is a process of creation by a subject that at the same time creates the subject and must not be mistaken by juridical action. With the latter, the laws are given and the facts need to be applied. Ethical action is about making new laws. Applying old ones may be better done by experts, but finding new ones is a job for everybody. That is why, Badiou is right to emphasise that ethical action is always a universalist one.

The total corpus of knowledge and rules – whether or not one doubts such a possibility – would mean the end of ethics, which is built on the fact that what the subject sees cannot be everything. Nobody wants an expert, whose expertise she does not know, to judge the relevance for her. In this sense, her disagreement with the results establishes what one could call following Badiou an ethical situation. Becoming a subject is possible only when a situation is not completely governed by laws external to the subject, since they are scientific and objective. Every user will make the experience that a search engine does not return what she wanted. The opinion of the search engine differs from that of users, which should be enough to convince them that their searches also produce just another match between the information need and the information itself. This difference creates an ethical situation and a becoming subject by changing her position in the world. Computers following experts’ opinions can simulate what everybody should believe. They must fail to find the ‘new’ an ethical situation is about.

## References

- Badiou, Alain; Ethics, An Essay On the Understanding of Evil, London, Verso 2001*
- Brin, Sergey and Page, Lawrence; The anatomy of a large-scale hypertextual Web search engine, In: Computer Networks and ISDN Systems, Vol. 30, No. 7, 1998, 107-117.*
- Deutschlandfunk; Ethik und Ökonomie von Suchmaschinen – Eine Tagung der Beterlsmannsstiftung, <http://www.dradio.de/dlf/sendungen/kulturheute/263774/>, 05/2004.*
- Grossman, David A. and Frieder, Ophir; Information Retrieval. Algorithms and Heuristics, Springer, Dordrecht 2004*
- Die Grünen; Suchmaschinen – Bundestagsfraktion Bündnis 90/Die Grünen, [http://www.gbet-tin.de/cms/files/dokbin/63/63516.reader\\_1564\\_suchmaschinen\\_das\\_tor\\_zum\\_net.pdf](http://www.gbet-tin.de/cms/files/dokbin/63/63516.reader_1564_suchmaschinen_das_tor_zum_net.pdf), 03/2005*
- Kant, Immanuel; Kritik der Urtheilskraft (1790), In: ders., Gesammelte Schriften, hrsg. von der Preußischen Akademie der Wissenschaften, Bd. V, 1902-1923.*
- Marcel Machill, Transparenz im Netz. Funktionen und Defizite von Internet-Suchmaschinen. Gütersloh: Verlag Bertelsmann Stiftung 2002*
- Van Rijsbergen, C. J.; Information Retrieval, 2nd edition, Dept. of Computer Science, University of Glasgow, 1979*
- Wiener, Norbert; The Human Use of Human Beings: Cybernetics and Society, Boston, Houghton Mifflin 1950.*

<sup>12</sup> Badiou, Alain: Ethics.