IRIE
International Review of Information Ethics Vol. 3 (06/2005)

Herman T. Tavani:

Search Engines, Personal Information and the Problem of Privacy in Public

Abstract:

The purpose of this paper is to show how certain uses of search-engine technology raise concerns for personal privacy. In particular, we examine some privacy implications involving the use of search engines to acquire information about persons. We consider both a hypothetical scenario and an actual case in which one or more search engines are used to find information about an individual. In analyzing these two cases, we note that both illustrate an existing problem that has been exacerbated by the use of search engines and the Internet – viz., the problem of articulating key distinctions involving the public vs. private aspects of personal information. We then draw a distinction between "public personal information" (or PPI) and "nonpublic personal information" (or NPI) to see how this scheme can be applied to a problem of protecting some forms of personal information that are now easily manipulated by computers and search engines – a concern that, following Helen Nissenbaum (1998, 2004), we describe as the problem of privacy in public. In the final section of this paper, we examine a relatively recent privacy theory introduced by James Moor (2004) to see whether that theory can shed any light on privacy concerns surrounding the use of search engines to acquire personal information. Although no definitive solution to the problems examined in this paper are proposed, we conclude by suggesting that Moor's privacy theory could help us to frame – via debate in an open and public forum – a coherent on-line privacy policy concerning whether, and which kinds of, personal information should be accessible to search engines.

Agenda

Author:

Herman T. Tavani:

- Organization and contact address: Rivier College, 420 Main Street, Nashua, New Hampshire 03060-5086
- Telephone, email and personal homepage: ☎ ++1 (603) 888-1311, extension 8597, ✉ htavani@rivier.edu, 💻 http://www.rivier.edu/faculty/htavani/
- Relevant publications:
  - Ethics and Technology: Ethical Issues in an Age of Information and Communication Technology. John Wiley and Sons, 2004; xxvi + 344 pages. [Second edition in progress; planned for June 2006]
  - Ethics, Computing, and Genomics: Moral Controversies in Computational Genomics. Jones and Bartlett Publishers, in press. [Planned for publication in June 2005.]
  - Intellectual Property Rights in a Networked World: Theory and Practice. (co-edited with Richard Spinello). Idea Group/Information Science Publishing, 2005; iv + 281 pages.
  - Readings in CyberEthics. (co-edited with Richard Spinello). Jones and Bartlett Publishers. Two editions. First ed. 2001; xvi + 601 pages. Second ed. 2004; xviii + 697 pages.

Introduction

Few would dispute the claim that search engines have provided an important service to Internet users – e.g., in directing users to available on-line resources for academic research, commerce, recreation, and so forth. Hence, some might be surprised to find that search-engine technology itself can be controversial from the perspective of personal privacy. Consider, however, that Internet search engines can be used to locate personal information about individuals. In some cases, personal information that is accessible to search engines resides in public records that are freely available on-line. In other cases, personal information resides in commercial databases (such as DocuSearch), and while this information is locatable via search engines, a small fee is required to access it. Also consider that some information about persons currently accessible on-line has been made available inadvertently; and in many cases, that information has become available without the knowledge and consent of the person or persons affected.

But why should these issues necessarily raise concerns for personal privacy? To answer this question, we first describe some basic characteristics of search engines in general. We then we show how access to personal information is facilitated by search-engine technology and why certain uses of this technology are controversial from a privacy perspective.

Search Engines and Implications for Personal Privacy

What, exactly, is search-engine technology, and how is this technology used to gain access to information about persons? Essentially, search engines are programs designed to point Internet users to a list of relevant Web sites that correspond a user's request for information about some topic or subject. As noted above, search engines can be used to locate information on a variety of topics – from academic research, to recreation, travel, commerce, etc. Search engines can also be used to acquire information about persons. Consider that by entering the name of an individual in a search-engine program's entry box, search engine users can potentially locate and retrieve information about that individual. For example, Marie Wright and John Kakalik (1997) note that a certain kind of information about individuals, which was once difficult to find and even more difficult to cross-reference, is now readily accessible and collectible through the use of on-line automated search facilities such as Internet search engines.

Still, we can ask why the use of search engines to gain information about persons, as opposed to other topics or subjects, raises privacy concerns. First, consider that an individual may be unaware that his or her name is among those included in one or more databases accessible to search engines. Further consider that if he or she is not an Internet user, that person might be altogether unfamiliar with search-engine programs and their ability to retrieve personal information about him. Thus individuals have little control over how information about them can be acquired by Internet users, which, in turn, has implications for personal privacy. So it would seem that questions concerning the impact that search engines have for personal privacy can indeed be raised.

Admittedly, the fact that one can search the Internet for information about one or more persons would not, at first glance, seem terribly controversial. After all, we might reasonably assume that the persons about whom information is being requested via a search engine have either placed some personal information about themselves on the relevant Web sites or perhaps have authorized someone else to do it for them. But there could also be personal information on these Web pages that an individual has neither included nor explicitly authorized to have placed on a Web site. David Kotz (1998) points out that since many email-discussion lists are stored and archived on Web pages, it is possible for a search engine to locate information that users contribute to electronic mailing lists or *listservers*. Search engines can also search through archives of *news groups*, such as *Usenet*, on which on-line users also post and retrieve information. One such group, *DejaNews*, is set up to save permanent copies of new postings. As such, it provides search engines with a comprehensive searchable database. Because the various news groups contain links to information posted by a person, they can provide search-engine users with considerable insight into that person's interests and activities. So it would seem to follow that not all of the personal information currently included on Web sites accessible to search engines was necessarily either placed there by the persons themselves or explicitly authorized to be placed there by those persons.

One might also assume that information currently available on the Internet, including information about individual persons, is, by virtue of the fact that it resides on the Internet, *public information*. And if this information is public in nature, then we can question whether it should be protected through privacy laws and policies. Of course, we can also question whether all of the personal information currently available on the Internet *should* be unprotected via privacy policies merely because it is viewed as public information. The following scenario may cause us to question whether at least some information about individuals that can be, and in some cases already has been, included on one or more Web pages or in databases accessible to Internet users should be viewed simply as public information that deserves no normative protection.

Hypothetical Scenario: Using Internet Search Engines to Acquire Information About an Acquaintance

Imagine a scenario in which an individual, named Pat, contributes to a cause sponsored by a gay/lesbian organization. Pat's contribution is later acknowledged in the organization's newsletter, a hardcopy publication that has a limited distribution. The organization's publications, including its newsletter, are subsequently converted to electronic format and included on the organization's Web site. That Web site is then "discovered" by a search-engine program and an entry about that site's address is recorded in the search engine's database. Assume that Pat has read the hardcopy newsletter that describes the various contributions that Pat and other members have made to the organization in question. It is possible that Pat has no idea that the contents of the newsletter have also been placed on the organization's Web site and that the existence of this Web site has been discovered by one or more search engines.

Now, further suppose that Pat is an acquaintance of yours from college and that you have not seen Pat since you both graduated two years ago. You then happen to cross paths briefly at a sporting event and agree to get together for dinner to catch up on events in your lives since your college days. Curious to learn more about what Pat has recently been up to, in order to be prepared to discuss some of these activities with Pat when the two of you get together for dinner, you decide to inquire about Pat via the Internet. You then access the Google search engine and enter Pat's full name in the entry box. A series of "hits" related Pat are then returned to you, one of which identifies Pat in connection with the gay/lesbian organization mentioned above. What would you likely infer about Pat on the basis of this particular "hit"?

Until now, you had no reason to wonder about Pat's sexual orientation. Pat has never disclosed to you any information pertaining to his or her sexual preferences, nor has Pat revealed through any public activities of which you had been aware any behavior traits that would link Pat to being homosexual. Yet as a result of a hit returned from the Google search engine, one might easily draw certain inferences about Pat's sexual orientation.

Perhaps Pat is, as a matter of fact, homosexual; and perhaps Pat is not. Pat's sexual orientation is not what is at issue here. Of course, even if Pat is a homosexual, and even if Pat is not troubled by the fact that others have this knowledge about him or her, the issue of how one is able to arrive at an inference about Pat's sexual persuasion is what seems problematic. What is problematic from a privacy perspective is that inferences about Pat's sexual orientation can be made in ways that Pat is unable to affect or influence.

Since Pat might have no idea that information about his or her activities involving the gay/lesbian organization is publicly available on-line to anyone with Internet access, we can ask whether the use of search-engine technology in Pat's case has raised any legitimate privacy concerns. Has Pat's privacy been violated in anyway? Or is the fact that the information about Pat was already public, at least in some sense, a relevant matter? And even if that information was publicly available in that it existed in printed material that was available to relatively few people, does it follow that there is no reasonable case to be made for why that particular information should not be normatively protected in cyberspace?

Some might argue that in the case of Pat, the fact that some personal information about him or her has been disclosed via search-engine technology is a trivial matter. After all, no one was harmed – at least not in a physical sense. However, we next examine an actual case where the use of search-engine technology (in conjunction with information brokers and off-line search facilities) to locate a person led to physical harm to an individual once that person was located. In fact, the harm ultimately resulted in that individual’s death.

Case Illustration: Internet Search Engines and Cyberstalking

In October 1999, twenty-year-old Amy Boyer was murdered by a young man who had stalked her via the Internet. The stalker, Liam Youens, was able to carry out many of the stalking activities that eventually led to Boyer's death by using on-line search facilities available to Internet users. To acquire personal information about Boyer, including information about where she worked, Youens elected to take advantage of search services provided by online "information brokers" in the commercial sector. For example, he used Docusearch.com, an on-line search agency that requires a fee for its services, to obtain the information he sought about Boyer (Grodzinsky and Tavani, 2004). So, in effect, Youens acquired much of the information he gained about Boyer through commercial on-line search facilities, as opposed to using only conventional search engines that are freely available on the Internet.

The cyberstalking incident involving Amy Boyer raises a wide range of ethical and social issues, one of which involves privacy (Tavani and Grodzinsky, 2002). For example, was Boyer's right to privacy violated because of the way in which personal information about her could be so easily gained by Liam Youens? Or was Youens simply accessing information about Boyer that was public and thus not eligible for any kind of legal or normative protection? Boyer's mother (Helen Remsburg) has since filed an invasion of privacy lawsuit (based on "commercial appropriation of personal information"), in addition to a "wrongful death" lawsuit, against Docusearch (www.epic.org/privacy/brief). And in February 2003, the Electronic Privacy Information Center (EPIC) submitted an *Amicus Curiae* brief against Docsusearch (www.epic.org/privacy/boyer/brief.html) in support of the claim that Boyer's privacy had been violated.

In assessing the Amy Boyer case from the perspective of personal privacy, we can ask: To what extent does the kind of personal information on the Internet that accessible via standard search engines, as well as through on- and off-line search facilities involving information brokers in the commercial sector, deserve some kind of legal or normative protection? In other words, to what degree is that personal information sensitive or confidential, and in what respect is that information *public* in the sense that it should be accessible to others? We next consider a framework for trying to understand and analyze the status of certain forms of personal information that would seem to span the private-public divide.

The Problem of Protecting Privacy in Public

Some forms of personal information enjoy normative protection via policies and laws because they involve data about persons that is either sensitive or intimate, or both. This kind of personal information can be referred to as Non-Public Personal Information (or NPI). However, many privacy analysts are now concerned over ways in which a different kind of personal information – Public Personal Information (or PPI), which is non-confidential and non-intimate in character – is also collected and exchanged over the Internet.

How can PPI and NPI be distinguished? As noted above, NPI can be understood as information about persons that is essentially confidential or intimate in nature. This could include information about a person's finances and medical history. PPI, which can also be understood as information that is personal in nature, is different from NPI in one important respect. PPI is personal information that is generally considered to be neither intimate nor confidential. For example, information about where an individual works or attends school, as well as what kind of automobile he or she owns, can be considered personal information in the sense that it is information about *some individual as a particular person*. However, this kind of personal information typically does not enjoy the same kinds of privacy protection that has been granted to NPI.

Until recently, concerns about personal information that was gathered and exchanged electronically have been limited mostly to NPI. And because of concerns on the part of many privacy advocates about the ways in which NPI has been exchanged, certain privacy laws and policies have been established to protect it. Many privacy advocates now worry about the ways in which PPI is routinely collected and analyzed via computer technologies. Recently, they have argued that PPI deserves greater legal and normative protection than it currently has. Helen Nissenbaum (1998) has referred to the challenge that now faces us with regard to protecting the kind of information that we refer to as PPI as the "problem of protecting privacy in public."

Why should the use of computers to collect and exchange publicly available information about persons generate controversies involving personal privacy? Initially, we might assume that there is very little to worry about with respect to the collection of PPI. For example, suppose that I happen to discover some information about Mary. I learn that Mary is a junior at Technical University, that she frequently attends her university's football games, and that she is actively involved in her university's computer science club. In one sense, the information that I have discovered about Mary is personal because *it is about Mary as a person*. However, that information is also public because it pertains to things that Mary does in the public sphere.

Should Mary be concerned that I am so easily able to find out this information about her? Certainly in the past, there would have been little reason to be concerned that such seemingly harmless and uncontroversial information about Mary was publicly available. Imagine, for example, a scenario in which eighty years ago a citizen petitioned his or her congressional representative to draft legislation that would protect the privacy of each citizen's movements in public places. It would have been difficult then to make a strong case for such legislation, because lawmakers and ordinary persons would have seen no need to protect that kind of personal information. However, some privacy advocates now argue that our earlier assumptions about the need to protect privacy in public are no longer tenable because of the way that information can be processed via computer and information technologies, especially in the commercial sphere. Nissenbaum (2004) notes that many entrepreneurs in the commercial sector currently proceed from an assumption that she believes is misleading – viz., the position that there is a "realm of public information about persons to which no privacy norms apply." It would seem that many "information brokers" who go about collecting personal information for their commercial databases find this kind of reasoning supportive of their enterprises.

From what we have seen in the hypothetical scenario involving Pat, and in the actual case involving Amy Boyer, the kind of reasoning used by information brokers can have implications that go far beyond the interests of entrepreneurs in the commercial sphere. Consider, for example, that DocuSearch.com, an on-line information company, provided Liam Youens with the information he needed to locate (and eventually murder) Amy Boyer. Yet, DocuSearch would argue that it was providing a service that was perfectly legal and that it was not responsible for Boyer's death merely because it provided information about Boyer to Youens. But even if that case had not resulted in the tragic outcome for Boyer, we can still ask whether Boyer's privacy rights were violated when DocuSearch provided information about Boyer to Youens without Boyer's knowledge and consent. To address this question and others surrounding the ability of search engines to access information about persons, we need an adequate framework of privacy.

A Privacy Scheme for Analyzing Controversies Surrounding Search Engines

Many theories of privacy have been put forth, and there is no need to review them here. James Moor (2004) has recently introduced a theory of privacy that incorporates important elements of traditional theories, which, individually, have addressed privacy concerns from the perspective of protecting individuals against either intrusion *or* interference *or* information access. According to Moor's comprehensive definition:

> *an individual has privacy in a* situation *if in that particular situation the individual is* protected from intrusion, interference, and information access *by others [Italics Added].*

One important element of Moor's definition is that it addresses issues of intrusion (into one's personal affairs) and interference (with one's personal decisions) and concerns involving access to (one's personal) information. Another important aspect in Moor's theory – especially for our analysis of privacy concerns surrounding search engines – is Moor's notion of a "situation," which is left deliberately broad so that it can apply to a range of contexts or "zones" that can be "declared private" in a normative sense. For example, a situation can be an "activity," a "relationship," or the "storage and access of information" in a computer or on the Internet. Thus, practices involving the use of search engine-programs would meet the criteria of a *situation* in Moor's scheme.

Central to Moor's privacy theory is another important distinction – viz., one between *naturally private* and *normatively private* situations. This distinction enables us to differentiate between a *loss* of privacy and a *violation* of privacy, thus showing that not every loss of privacy necessarily results in a violation of privacy. Consider that in a naturally private situation, individuals are protected from access and interference from others by "natural" means, such as physical boundaries in natural settings that might preclude one from being seen. Consider, for example, a situation where one is hiking alone in the woods. In this case, if the person is seen at some point while hiking, his or her privacy can be *lost* but not *violated*. It is not violated because there are no norms -- conventional, legal, or ethical -- according to which one has a *right* or even an expectation to be protected (i.e., not to be seen hiking). In a *normatively private situation*, on the other hand, individuals are protected by conventional norms. An individual's privacy can be violated only in "normatively private situations" because it is only in those kinds of situations that zones or contexts that merit some kind of normative protection have been formally established.

When a search engine is used to locate information about some person, *X*, has *X's* privacy necessarily been violated? Arguably, *X* may have lost some of his or her privacy in the process, but it is not yet clear whether any privacy violation has also occurred. But consider once again the hypothetical scenario involving Pat, where information returned from a search query about Pat suggested that he or she is likely a gay or lesbian. Was Pat's privacy violated in – i.e., in a *normative* sense – in this scenario? Pat may indeed have lost some privacy in the *natural* (or descriptive) sense of privacy because information about Pat's volunteer work on a project was disclosed to a wider audience. However, Pat's privacy is violated only if search engines are (i.e., have been formally declared to be) *normatively private situations*.

Should practices involving the access of personal information on the Internet via search-engine technology be declared a normatively private situation? If we begin to think of personal information on the Web as constituting (Moor's notion of) a normatively private situation, we can also begin to think about some ways that this information can be protected in certain ways while other kinds of information – i.e., non-personal information – currently accessible to search engines can continue to flow easily. To help us decide this matter, Moor provides a framework for debating issues such as these. For example, he recommends that there be open and "rational" debate on questions involving privacy policies, and this is clearly articulated by Moor in his *Publicity Principle*. According to this principle:

> *Rules and conditions governing private situations should be clear and known to persons affected by them.*

Thus there is an important element of transparency or openness in Moor's principle, which also supports the notion of informed consent in policy decisions. This would certainly apply in the case of Internet users, who first would be made aware of the issues involving the access of personal information on-line and who would then have a say in how the policy would be determined. As Moor states:

> *…we can plan to protect our privacy better if we know where the zones of privacy are and under what conditions and to whom information will be given.*

In Moor's scheme, privacy policies need not be cast in concrete, since they are always subject to refinement and revision. Moor also points out that privacy policies can, under certain conditions, be justifiably breached – via his *Justifications of Exceptions Principle*. And they can also be modified and revised through his *Adjustment Principle*. So, Moor's privacy theory would seem to provide plenty of flexibility within a structure that sets up zones of privacy called *normatively private situations*.

Applying this model of privacy to practices involving the use of search engines to acquire information about persons would perhaps be an ideal way of testing out Moor's privacy theory in the area of public policy involving the Internet. It could also prove very useful in an effort to resolve some of the concerns we have identified with respect to the problem of privacy in public, particularly as that problem applies to the use of search engines in online activities.

Concluding Remarks

We began this essay by examining some reasons why the use of search engines to acquire information about persons raises privacy concerns. We then considered a hypothetical scenario and an actual case, both of which were controversial because of the way that search engines were used to acquire personal information about individuals. Next, we considered how privacy issues arising in these cases were similar to those surrounding the "classic" problem of determining the private vs. public character of personal information; and we saw how this concern is exacerbated on the Internet by what Nissenbaum calls the problem of privacy in public. Finally, we examined Moor's theory of privacy to see how we could better understand, and perhaps even begin to resolve, some privacy issues associated with the use of search engines to gain information about persons by framing a comprehensive privacy policy that explicitly addresses this issue.

Acknowledgements

In composing this paper, I have drawn substantially from material in two previously published works: Tavani (1998, 2004).

References

Amicus Curiae *Brief of the Electronic Privacy Information Center. The State of New Hampshire Supreme Court. Case No. C-00-211B (Estate of Helen Remsburg v. Docusearch, Inc., et al.). Available at http://www.epic.org/privacy/boyer/brief.html.*

Electronic Privacy Information Center. "The Amy Boyer Case: Remsburg v. Docusearch*." Available at http://www.epic.org/privacy/boyer.*

Grodzinsky, Frances S., and Herman T. Tavani (2004). "Ethical Reflections on Cyberstalking." In R. A. Spinello and H. Tavani, eds. Readings in CyberEthics*. 2nd ed. Sudbury, MA: Jones and Bartlett, pp. 561-570.*

Kotz, David (1998). "Technological Implications for Privacy." A paper presented at the Conference on The Tangled Web: Ethical Dilemmas of the Internet, Dartmouth College, Hanover, NH, August 7-9.

Moor, James H. (2004). "Towards a Theory of Privacy for the Information Age." In R. A. Spinello and H. T. Tavani, eds. Readings in CyberEthics*. 2nd ed. Sudbury, MA: Jones and Bartlett, pp. 407-417.*

Nissenbaum, Helen. (1998). "Protecting Privacy in an Information Age," Law and Philosophy*, Vol. 17, pp. 559-596.*

Nissenbaum, Helen (2004). "Toward an Approach to Privacy in Public: Challenges of Information Technology." In R. A. Spinello and H. T. Tavani, eds. Readings in CyberEthics*. 2nd ed. Sudbury, MA: Jones and Bartlett, pp. 450-461.*

Tavani, Herman T. (1998). "Internet Search Engines and Personal Privacy." In M. J. van den Hoven, ed. Proceedings of the Conference on Computer Ethics: Philosophical Enquiry*. Rotterdam, The Netherlands: Erasmus University Press, pp. 214-223.*

Tavani, Herman T., and Frances S. Grodzinsky (2002). "Cyberstalking, Personal Privacy, and Moral Responsibility," Ethics and Information Technology*, Vol. 4, No. 2, pp. 123-132.*

Tavani, Herman T. (2004). Ethics and Technology: Ethical Issues in an Age of Information and Communication Technology*. Hoboken, NJ: John Wiley and Sons.*

Wright, Marie, and John Kakalik. (1997). "The Erosion of Privacy," Computers and Society*, Vol. 27, No. 4, pp. 22-25.*