

Thea van der Geest

Big business, Big Brother

User profiling on the Internet

Keywords: user profiling, personalization, customization, one-to-one marketing, data mining

When doing our on-line business, we scatter snippets of personal information around. Data mining companies aggregate the information in user profiles and sell those to commercial and public organizations. The data in the profiles can be used to adapt on-line content and presentation, thus making communication and interaction easier and more effective for users. But the aggregated personal information can also be used to single users out for less noble purposes. Most Internet users are not aware of the spyware that tracks their behavior and collects personal information. It is a responsibility of information and document designers to make web communication safe and private.

In the science fiction movie *Minority Report*, director Steven Spielberg shows us Washington DC in the year 2054. By that time, computing is ubiquitous. When the movie's hero, crime fighter John Anderton, moves through the streets of the city, he is continuously recognized on the basis of his biometric data; his iris is scanned. In the shopping mall, holographic shop assistants and interactive posters address John on a first-name basis, whispering irresistible offers based on his preferences and his previous purchases. In other words, they use personal information about John that must be

stored in a huge database. His user profile data are used to personalize the communication and target him as a 'group-of-one'. Is this indeed what we can expect to happen in the near future of technology-mediated communication? Is this science, or is it fiction?

Web services and user profiling

User profiling is a hot topic in Internet technology development circles. With networked computers at our disposal wherever we are, Internet technology is intertwined with our daily life, be it work or leisure. Increasingly, we use it for specialized services, such as buying plane tickets, doing our banking affairs, obtaining the special parking permit that we are entitled to, sending in our tax forms. All these services require ICT applications, running within web sites. The standards that are developed for making all those applications work and communicate with each other behind the screens are called Web Services or Application Services standards.

Advances in computing and database technology are making it technically possible and relatively easy to (for example) connect your financial information, stored in the bank's database, with the figures that you provide in the electronic tax form application. Or to connect your purchase behavior, as documented on your personal

customer card of your favorite super market, with the health information you provided on-line to your health insurance company. We scatter snippets of personal information about ourselves continuously, and by now the technology has become available to connect all those snippets in a personal user profile. Both commercial and public organizations can't wait to act upon those user profiles, and in some cases have started using them without asking our informed consent. The year 2054 is not far away at all; in some regards it might even have arrived already!

User benefits of profiling

In a way, user profiling is old wine in a new bottle. Rhetoric, the art of communicating effectively and persuasively, has always focused on how to adapt the message in its form or content to the intended audiences, in order to achieve the intended effects. Effective communication, and hence effective information and document design, takes into account what the receivers would like to hear, what they need to know, how well they comprehend information, what they already know about the topic, and how they prefer to be addressed. Being a good communicator meant applying user profiling long before 'audiences' became 'users'.

In the past, acquiring all kinds of services or products meant that you were in contact with many different organizations, at different times and at different locations, providing each with the data they needed to deliver the service and the product. Whatever your special preferences or experiences, you had to fit the mould of the service provider. Nowadays, since many of these services are offered electronically, the actual contact with organizations is often realized through your computer, be it a desktop computer or a more mobile application such as a phone or PDA. Computers have become single access points to a wide variety of organizations, services and

products. But if you do your business through that single access point, you do not want to be treated in each contact as if it were a first-time contact, with the organizations requiring your individual data as if you have not provided them already at earlier occasions.

Single access is particularly efficient for your contacts with sets of related organizations or departments within organizations. For example, for an impaired citizen who is moving to a new hometown, the request for a special parking permit is part of a 'scenario' of moving from one town to another. That scenario also includes activities such as registering with the municipality and finding out about waste collection times at the new address. Although these services might be offered by different departments or even by different organizations, the citizen will perceive them as part of one 'event' and might easily become frustrated when playing out the complete scenario would require providing specific data over and over.

Adapting the content, presentation or interaction of web sites sets on the basis of your user profile, for example your preferences, needs and prior surfing behavior, sounds like a good idea. Companies can provide better service and offer you more suitable products when they know who you are and what you want. That's why 'data aggregators' have so much interest in collecting those snippets of data you scatter around when using the web. There is big business, hidden in the small details of your personal life.

Targeting and personalizing communication: Big business

Once an organization has collected data about you, it can try to overcome the negative effects of 'one size fits all' in its communication, services and product offers. It can make 'intelligent' use of your profile data for the planning and adaptation of further messages, information or actions. It uses the data about your characteristics

or behavior to adapt information and communication to your needs and interests and to predict your behavior in the future. But it won't do so, unless it thinks it can achieve its organizational goals better: sell you more products, make sure that you comply better with the law, convince you better of the quality of the information.

Many of us have received those promising envelopes on our doorstep, telling us that we personally are the winners of a small fortune. We know that most direct mail letters are just barely personalized; our name is inserted into a text slot, but the content is exactly the same for the millions of other fortunate winners. The senders bought a database with our names, but not much more, from a data aggregator. On-line stores do a cleverer job than those direct mailers. They not only address us with the names we provided, but also adopt the content to our interests. They ask us to provide information (a user profile) about our preferences for their stores and products. Then they make us personalized recommendations, adapting the *content* of the web page on the basis of the data we provided. Figure 1 shows an invitation to provide user profile data, as it can be found in the Amazon online store.

Targeting and personalizing is not new, especially

not in marketing communication. But with increasing computing power and digital snippets of personal information scattered around, it has become feasible to target individuals rather than groups. Data-mining companies have started collecting information from the numerous databases that contain the digital traces of our daily life: credit card companies, on-line stores, library book orders, etc. They analyze the data for patterns to distinguish groups (market segments) or try to identify individual users whose profile shows particular characteristics. And particularly when group patterns are connected to our individual names and addresses, we can easily get the feeling that Big Brother is watching us.

Big Brother is spying

A relatively new means to secretly collect information about your web surfing behavior is spyware. Spyware is software that is put on your computer (as a virus, or when you are installing other software) and then tracks your web surfing behavior. That information is sold to advertisers or other interested parties. Whereas the legal status of unwanted e-mails and viruses is becoming clearer, the awareness of spyware among Internet users and lawmakers is still low. Besides all the information about ourselves that we willingly (but often naively) provide, spyware helps data aggregators to fill our user profiles.

Is user profiling already widely used and if so, to what ends? Companies are reluctant to give any information about their data mining and user profiling efforts. But we know more about the US Federal Government's data mining efforts. The General Accounting Office (GAO), the controlling agency of the US federal government, has investigated the government's use and application of data mining (GAO, 2004). Of 128 federal departments and agencies surveyed, 52 reported that they are using or planning to use data mining, in 199 different projects. The top reasons to use

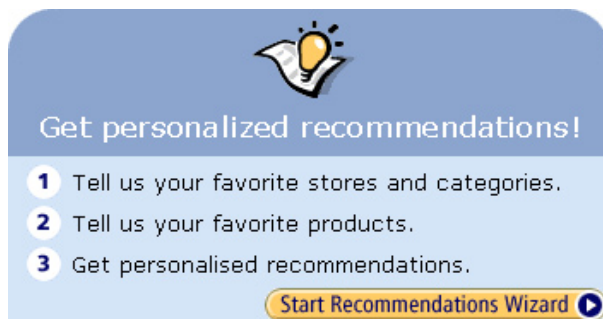


Figure 1. Recommender application in the Amazon on-line store

data mining, as reported by the federal agencies, are improving services or performance (in 65 projects) and detecting fraud, abuse, terrorism and other criminal activities or patterns (in 53 projects). Out of the 199 projects mentioned, 122 used information that could be traced back to an individual. Among the data processed were credit histories, credit card transactions, student loan applications, and bank account numbers. These personal data were primarily used to detect criminal and terrorist activities and to increase tax compliance.

A range of laws limits how governments can collect and use information on its citizens. But the private sector in the US and many other countries operate under fewer restrictions. If governmental or other organizations can collect data on you without your explicit permission, than it is time to worry about privacy. Not only as citizens, customers or clients, who wonder what is happening with those easily filled in names and e-mail addresses, but also as designers of the web pages and services. It is part of our responsibility as designers to think carefully about what kind of data from our customers we actually need to design optimal communication and interaction, and what guarantees are in place to make sure that those data are used safely and in the interests of our audiences.

References

- Awalt, S. Minority Report (2002).
<http://www.spielbergfilms.com>. Consulted January 2005.
- General Accounting Office (2004). Data Mining. Federal Efforts Cover a Wide Range of Uses.
Report GAO 04-548. Retrieved January 2005 from
<http://www.gao.gov/new.items/d04548.pdf>

ABOUT THE AUTHOR

Thea van der Geest is an associate professor in the Communication Studies/Technical Communication Department of the University of Twente, the Netherlands. She teaches courses in interface and interaction design, web design and research methodology. Her research focuses on the process of web design, methods for evaluating web sites, usability and accessibility, and information and document design for new media.
Contact: t.m.vandergeest@utwente.nl