

How Much do Digital Natives Disclose on the Internet – A Privacy Study

Erik Buchmann, Klemens Böhm
 Karlsruhe Institute of Technology (KIT), Germany
 {erik.buchmann|klemens.boehm}@kit.edu

Abstract—With the advent of social services on the Internet that encourage the disclosure of more and more personal information, it has become increasingly difficult to find out where and for which purpose personal data is collected and stored. The potential for misuse of such data will increase as well, e.g., due to the ongoing extension of social sites with new features that make it more appealing to reveal personal details. In order to research and develop approaches that give way to privacy on the Internet, it is important to know which kind of information can be found, who has been responsible for publishing it, the age of the information, etc. This paper describes a user study about the personal information available about Digital Natives, i.e., young people who have grown up with the Internet. In particular, we have guided 65 undergraduate students to search the web for personal information on themselves by using various search engines. Our students have completed 302 questionnaire sheets altogether. We have analyzed the questionnaires by means of statistical significance tests and cluster analysis. As a part of our results, we have found out that today's personal search engines like 123people.com do not find much more information than general-purpose search engines like google, and that today's Digital Natives are surprisingly aware of the information they are willing to disclose.

Keywords—privacy; user study; digital natives.

I. INTRODUCTION

At this moment, more than 200 social networking sites exist [1] that encourage the disclosure of personal information about the daily life, hobbies and interests, work-related information, etc. Furthermore, more and more classical web forums, photo-sharing portals, news portals and other sites offer “social” features, e.g., allow to share and comment digital objects of interest. In consequence, it has become more and more tempting for individuals to disclose personal details, with unpredictable consequences for privacy.

The potential for misuse of such data depends on how it can be found. While general-purpose search engines like Google or Bing search for any kind of information, people-search engines like 123people or Yasni are tailored to search for personal details, e.g., from web forums, social network sites or commercial portals.

In this article, we investigate which personal information is typically available on the Internet. We study who has published this information and how our participants assess the impact of its availability on their privacy. From related

research (see Section II) and privacy issues on public media, we have developed four classes of research questions:

A1: Information characteristics This class subsumes the extent of personal data available on the Internet, the age of the data and parties who have uploaded it.

A2: Search characteristics This class covers the influence of the search engine and the search terms on the search result, and the amount of ambiguous information.

A3: Impact on privacy This class considers if people are content with the fact that the personal information they have found on themselves has been uploaded and is available on the Internet, and how sensitive they deem this information.

A4: Patterns and rules The fourth class addresses relationships between the characteristics of the information. In particular, we look for clusters of people with similar privacy attitudes and privacy behavior.

We have conducted a user study with educated Digital Natives, i.e., with people grown up with the Internet. Digital Natives are relevant for our study, because those people have integrated the Internet into their daily life. A very large share of the adolescents and young adults belong to this group. We have decided in favor of a qualitative study, i.e., a very detailed questionnaire and an intensive supervision of the study participants. Over a period of three years we have guided 65 undergraduate students of computer science to search for personal information about themselves by using various search engines. We have asked them to state who has uploaded this information, the age of the data, who would be able to find it, etc.

As a part of our results, we have found out that today's Digital Natives are very aware of the information they are willing to disclose. Nevertheless, despite the fact that names can be ambiguous (“John Smith”), all of our participants found at least some information about themselves on the Internet, and they disagreed or strongly disagreed with the availability of about one fourth of this information.

Paper Structure: The next section reviews related work. Section III describes our study methodology. The study is presented in Section IV, followed by a discussion of the study results in Section V. Section VI concludes the paper.

An extended version of this paper that includes more analyses and discussions is available as a technical report [2].

II. RELATED WORK

In this section, we explain the privacy paradox, we outline studies on Internet privacy in different use cases, and we discuss privacy perception and user categories.

Privacy Paradox: Our survey is motivated by the privacy paradox [3]: This paradox means that the attitude towards privacy and the daily behavior of individuals is inconsistent in many cases. For example, a study about anonymous and personalized gift cards [4] shows that people tend to assign a high price to the protection of a certain information, but in fact accept a much lower price to actually sell the same information. In contrast, we want to find out if there is a discrepancy between the personal information Digital Natives have explicitly published and the information they would tolerate to be disclosed. The privacy paradox can be modeled as a function of costs and benefits, which is maximized by each individual [5]. The costs include the risks of identity theft, marketing, stalking or negative reputation. Benefits include social aspects like relationships, collaborations, friendships or positive reputation in general. Related to the privacy paradox is the privacy awareness, i.e., the individual attention and motivation regarding the whereabouts of personal data. Privacy awareness influences individual decisions about publishing data [6].

Studies on Internet Privacy: Comparative privacy studies consider different use cases on the Internet:

Social Networks A study on information disclosure in social networks like Facebook or Myspace relates experience and behavior of users to the amount of private information that is disclosed [7]. Another study focuses on the privacy settings that control which information from the personal profile is shown to others [8].

eCommerce Privacy studies on customer data in eCommerce focus on the relationship between privacy and sales. A customer cannot observe if an online dealer follows the privacy policy on the shopping web site. Thus, a study [9] investigates the trust of the consumers in the willingness and ability of the dealer to handle personal data with care.

Personalization Many commercial web sites generate customer loyalty by personalization. This requires the customer to reveal personal details. The tradeoff between personalization and privacy is known as the online consumer's dilemma, which has been studied according to user value [10], transparency and willingness [5], and other impact factors [5].

The studies show that users tend to reveal information only if they see a direct use for it. For example, customers of a web shop do not disclose religious information [11]. This is important for our survey, because it shows that Internet users do not publish information indiscriminately.

Privacy Behavior and Privacy Perception: Surveys [12] about privacy behavior investigate the relationship between the perception of risks [13], e.g., identity theft, and the use of privacy-enhancing technologies. The studies show

that the perception of privacy risks varies widely, but privacy behavior has been comparable among all participants.

Categories of Users: We are interested in identifying user groups that differ with respect to the personal information available on the Internet. An email survey of Internet users [14] has computed a score for privacy concerns on the Internet from questions about typical situations, e.g., if an individual registers for a company web site when receiving an unsolicited email about a new product. The survey has identified the categories "unconcerned user" (16%), "circumspect user" (38%), "wary user" (43%) and "alarmed user" (3%). Studies that directly inquire the privacy behavior from the users are prone to the privacy paradox. Our study in turn looks at this problem from a different perspective: we analyze personal information disclosed on the Internet.

III. METHODOLOGY

In this section, we compile concrete research questions and we describe our study methodology.

A. Research Questions

To investigate the personal information that is available on the Internet, we have come up with specific research questions, as follows:

A1: Information characteristics

- How much personal information is available?
- How old is the information?
- Who has made the information available?

A2: Search characteristics

- Which search terms have yielded most information?
- How much does the search result depend on the search engine?

A3: Impact on privacy

- Have our participants been surprised to find a particular piece of information?
- Had our participants given permission to upload the information?
- How sensitive do the participants deem the information they have found?
- Do the participants approve that this information is available on the Internet?
- Who is able to find which kind of information?

A4: Patterns and rules

- Do groups of individuals with different privacy perception and behavior exist?

B. Study Participants

We have tested our research questions on educated Digital Natives, i.e., on people who have grown up with the Internet, for two reasons. First, these individuals use the Internet frequently, and they are aware of the social benefits of sharing personal information, e.g., to keep contact with friends and relatives, or to find individuals with similar interests and

attitudes. Second, Digital Natives can be assumed to be able to develop strategies, e.g., using different pseudonyms and email addresses for different purposes, to prevent someone from learning personal details which are not for the eyes of others. We have conducted our study with 65 German undergraduate students of computer science. Since we had announced an anonymous study and demographic data is a quasi identifier [15], we did not collect such information.

C. Study Procedure

We have conducted our study in three tranches with different participants over a period of three years. In the first step of each tranche, we have described the purpose of the study to our participants. Furthermore, we have handed out a guideline how to search for personal details on the Internet by using different search engines, and by refining the search term if a search returns only results that do not have any relationship to the searcher.

In a second step, we have handed out a number of identical questionnaires to each participant. We have guided our participants to search for personal information, i.e., we have provided hints and support if necessary. We have asked our participants to answer one questionnaire sheet for each distinct search result, i.e., each answer sheet has been obtained using a different set of search terms and/or a different search engine. To avoid erroneous data, we have told our participants to omit questions when they do not feel comfortable to provide us with correct answers. Guideline and questionnaire are available at [2].

D. Questionnaire

In this subsection, we briefly introduce our questions. All questionnaires contained the following questions:

- Q1** Which search engine did you use?
Q2 Which search terms did you use?
Q3 Which people know the search terms used?
Q4 Does the search term itself contain private information?
Q5 Which kind of information is on display on the first 20 hits of the search results?
Q6 How much information about yourself is displayed?
Q7 How old is the information found?
Q8 Who has uploaded the information?
Q9 Estimate the sensitivity of the information.
Q10 Do you approve that this information is available?
Q11 Have you been surprised to find this information?
Q12 Did you allow that this information was published?
Q13 What is shown on the images in the search results?

Depending on the question, we had allowed predefined categories of answers, free-text or a five-point Likert scale. We explain our questions in detail in the next section.

IV. STUDY

We have obtained 58 questionnaires from 10 participants in 2009, 137 questionnaires from 21 participants in 2010

	Question	Number
Q6	How much information about yourself is displayed?	123
Q9	Estimate the sensitivity of the information.	65
Q10	Do you approve that this information is available?	60
Q12	Did you allow that this information was published?	49
Q11	Were you surprised to find this information?	44

Table I

TOP-5 OF THE QUESTIONS THAT HAVE REMAINED UNANSWERED

and 107 questionnaires from 34 participants in 2011. Thus, 65 participants provided us with 302 questionnaires, and each questionnaire contains information about one distinct search result. 152 questionnaires contained one or more questions that have not been answered. 51 participants always answered all questions on each questionnaire. Table I shows the top-5 of questions that have not been answered, and the number of questionnaires where the question has been left unanswered. Note that **Q11** and **Q12** were not part of questionnaires from 2009.

A1: Information Characteristics

How much personal information is available on the Internet?: We have asked our participants to categorize the information that was on display on the first page of the search results or within the first 20 hits (**Q5**). We have provided the following categories: “Memberships” means that the search results indicate that the person concerned is a member of an online community or a social network. The class “Postings” refers to content generated by the person, e.g., a product review on Amazon. “Photos” means that the search result contains pictures showing or made by the searcher. “Locations” refers to places related to the searcher, e.g., the place of living. “Addresses” means telephone numbers, email addresses, etc. “Hobbies” and “Employment” indicates leisure and professional activities, and “Friends” refers to information about social contacts. For the years 2010 and 2011 we have also asked for the content of images in the search results (**Q13**).

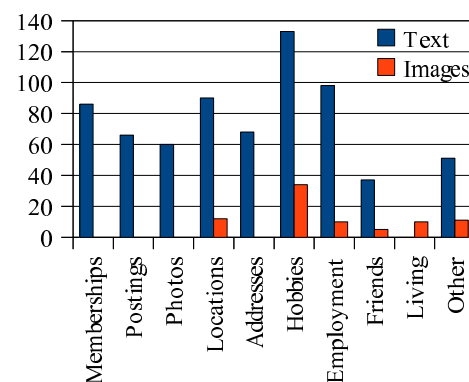


Figure 1. Which and how much personal information is found?

Figure 1 shows that the majority of the textual information available on the Internet refers to hobbies, followed by

Year of the Study	2009	2010	2011
Minimal Age	0.97 (1.67)	1.19 (2.76)	1.54 (2.12)
Average Age	2.07 (1.60)	2.38 (1.85)	3.16 (2.13)
Maximal Age	3.18 (1.80)	3.91 (2.74)	4.65 (2.54)

Table II
AGE OF THE INFORMATION

Uploader	Myself	Friends	Colleagues	Unknown
Number	170	59	101	44
Percent	45%	16%	27%	12%

Table III
UPLOADER

employment, locations and memberships. The distribution of the information categories found on images is similar to the textual results. The information found is well balanced over almost all categories we have provided. Only hobbies seem to be over- and friends underrepresented. Besides our questionnaires, we asked our participants where this information has come from. Important sources of information were student-research papers (recall that our participants were students), web sites of schools and sport clubs that publish awards, placements and team lists, and private homepages.

How old is the information?: Since we were interested to find out if the information found might be out of date, we have asked our participants to write down the range of the age of the information displayed on the first page of the query result (Q7). Table II shows the minimal, average and maximal age of the information found, together with the standard deviation (in parentheses). The table shows that, from year to year, the oldest information in the search result gets older. We speculate that publishing personal information regarding our participants at a large scale might have started around 2007, e.g., as a result of online communities like Facebook becoming more and more popular.

Who has made the information available?: From a privacy perspective, it is different if the individual concerned or someone else has uploaded the information. We have asked our participants which category of people might have been responsible for uploading (Q8). “Myself” means that our study participant has uploaded the information she has found. “Friends” subsumes friends, acquaintances and relatives. “Colleagues” means that the information has been uploaded with a relation to professional activities, e.g., education, employment or studying. Table III reveals that most of the information our participants have found on the Internet has been uploaded by themselves. Furthermore, a lot of information has been uploaded from colleagues. This observation complements Figure 1, which tells us that “Employment” is the second most-frequent category of information found. A small part of the information has been uploaded by unknown parties.

A2: Search Characteristics

Which search terms have yielded the most information?: Our participants have searched for personal details by

using various search terms. In particular, we have encouraged them to search for combinations of the following terms: first name, last name, parts of the postal address of their home and workplace, employment details, email addresses and login names used for instant messaging services or online communities. Note that our participants have filled out questionnaires only for combinations of search terms that have returned at least some personal information. Thus, we have not collected information about inconclusive searches.

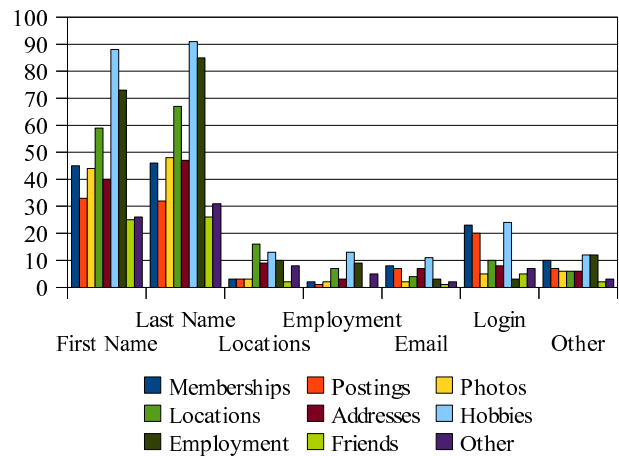


Figure 2. Search terms and information found

Figure 2 shows components of the search term together with the category of information returned. The figure indicates that search terms including at least a part of the real name find most personal details, i.e., user pseudonyms, nicknames, login names etc. play a less important role. However, knowing the login name of a person might be helpful to obtain information related to hobbies, online community memberships and online forum postings which might not be associated with his or her real name.

How much does the search result depend on the search engine?: Our participants were free to use various search engines. To provide a starting point, we have suggested Google.de (general purpose search), Images.Google.de (image search), Yasni.de and 123People.com (person search).

Figure 3 shows which search engines have found which information. In contrast to general-purpose engines, person-search engines should produce more personal results, since they are able to search for semantic information in structured databases, e.g., address registers, indexes of social networks, and in electronic market places like Amazon.com or eBay. Thus, we have expected that person-search engines would be heavily used during our study. However, we have observed the opposite: Provided with the search term “first name last name” (cf. Figure 2), Google has found more information than person-search engines. Only one participant did not find any personal information with Google, but with a person-search engine. The search results of the person-search engines Yasni and 123People were strikingly similar.

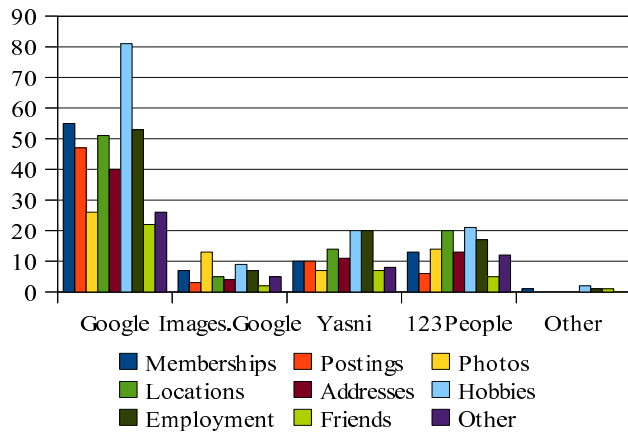


Figure 3. Search engines and information found

A3: Privacy Issues

Have our participants been surprised to find a particular piece of information?: To find out if our participants were able to control which personal information is shown to others, we have asked them if they were surprised by the availability of the information found (Q11). This question was answered on 200 questionnaires. Figure 4 shows that, in most cases, the participants found the information they had expected. However, in 20% of all searches our participants were at least surprised by the result, i.e., a significant share of information is available on the Internet without the individuals concerned knowing about it.

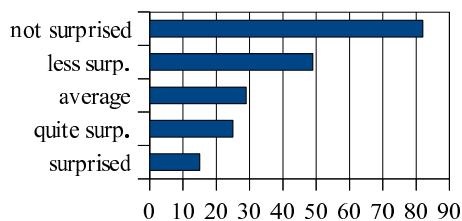


Figure 4. Surprise to find an information

Had our participants given permission to upload the information?: In 2010 we have asked if the information our participants have found has been uploaded with their permission (Q12). In 2011, we have refined this question. Now we have asked *which share* of the information has been uploaded with permission. Our participants were asked to regard information they have uploaded themselves as “upload with permission”. We have obtained 99 answers to this question in 2010, and 96 in 2011. Table IV shows the results for both years. 15% to 20% of any information found has been uploaded without consent of the individuals concerned. The more detailed results from 2011 indicate that approximately 50% of all searches returned at least a few results where the data has been uploaded without consent. These findings also correspond to Figure 4, where

our participants were at least surprised about 20% of the information found.

Year	With Permission	Number	Percentage
2010	no	15	15%
	yes	84	85%
2011	none	20	21%
	few	4	4%
	average	11	11%
	many	13	14%
	all	48	50%

Table IV
UPLOAD WITH PERMISSION

How sensitive do the participants deem the information they have found?: In order to estimate the impact of the information available, we have asked our participants about the sensitivity of the information they have found (Q9). This question was answered on 237 questionnaires. As Figure 5 shows, approximately one-fifth of the information found was deemed to be either private or secret, i.e., the participants appraised a significant impact on their privacy.

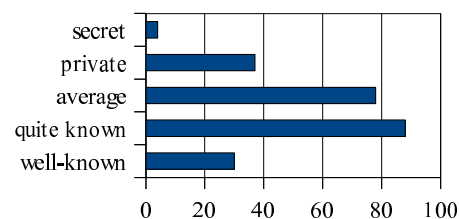


Figure 5. How sensitive is the information?

Do the participants approve that this information is available on the Internet?: As a follow-up question to the last one, we have asked if our participants could tolerate that the information found was on display on the Internet (Q10). This question was answered on 228 questionnaires. Since one-fifth of the information has been uploaded without permission, and the same share of information has an impact on the privacy of the participants, we expect that our participants disagree with the availability of at least one-fifth of the information.

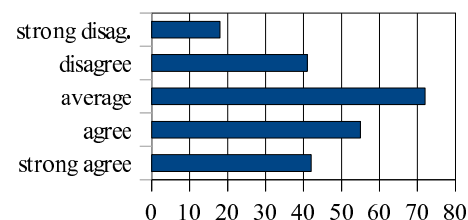


Figure 6. Agreement with availability

As Figure 6 points out, our participants disagree or strongly disagree with the availability of one-fourth of the information. We have calculated the empirical correlation coefficient between the sensitivity of the information (Q9)

and the approval of its availability (Q10) by regarding the answers to these questions as interval-scaled variables. Both variables are correlated; the correlation coefficient is 0.78. This means that in many (but not in all) cases a participant who thinks that an information is sensitive does not want this information published on the Internet.

Who is able to find which kind of information?: Since the information found depends on the search term, it is important to know who would be able to find which kind of information, i.e., who knows which search term. For example, we know from personal observations that many people do not tell vague acquaintances details about their employment or their place of living, which would enable them to find some information (cf. Figure 2). Figure 7 shows the search terms used together with the categories of people who know these terms. The figure shows that first name and last name are generally known to many categories of people. Locations, employment details, login names and other kinds of identifying information are known to much fewer people. Furthermore, the figure tells us that our participants have shared email addresses and login information with more friends and acquaintances than relatives or other people. We see this as an indication to prevent people like parents or colleagues from learning some information.

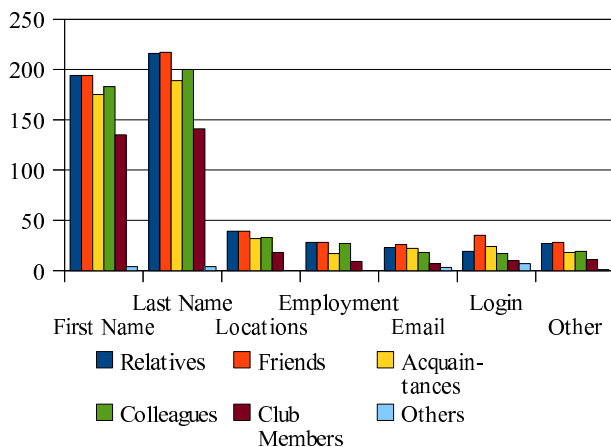


Figure 7. Who knows which search term?

A4: Patterns and Rules

Do groups of individuals with different privacy perception and behavior exist?: In order to design appropriate privacy mechanisms, it is important to identify groups of people with similar attitudes towards personal information on the Internet. Considering that participants have returned different numbers of questionnaires, we have decided for a two-stage procedure: First we apply a clustering approach on all questionnaires. In the next stage, we assign people to each cluster. In particular, we derive a feature vector from each of our 302 questionnaires. The feature vector models the answers of questions about search terms, privacy attitudes

and the amount of information available (Q2, Q4, Q6, Q9, Q10). We have regarded the answers as interval-scaled features where unanswered questions are an additional interval. Since it allows us to inspect clustering results of varying size, we have applied a hierarchical clustering approach. In detail, we have used between-group linkage [16] that starts with one cluster for each feature vector and iteratively combines two clusters with the smallest average distance between all group members in each step. We have used the square Euclidean distance. Due to the hierarchical clustering approach, all questionnaires will be assigned to a cluster. Finally, we have assigned a participant to a cluster if all but at most one questionnaire are a member of the same cluster.

We have manually interpreted cluster sets from 10 to two clusters. According to our interpretation, the most meaningful set consists of four clusters:

Cluster 1: Restrained Publishing This group (105 questionnaires, 18 participants) has found only little information on the Internet, and has not found anything that would have had a severe impact on their privacy: From a privacy perspective, all search results were deemed harmless. The data available has been published with the consent of the individuals. We conclude that this group of people controls very well which information is published on the Internet.

Cluster 2: Incomplete Questionnaires The second group (103 questionnaires, 6 participants) has returned questionnaires that have been filled out incompletely. Because of the anonymity of the study itself, we could not ask the participants for further explanations. We have spent much effort in supervising our participants, and we suppose that they have understood the questionnaire. However, the participants might have found nothing about themselves on the Internet, or they might not have wanted to disclose their results.

Cluster 3: Surprised Individuals from the third group (62 questionnaires, 8 participants) have been negatively surprised about the kind and the extent of personal information they have found about themselves on the Internet. The information has been published without consent of the individuals, or they have published the information without remembering that the data would be available for anybody later on. We assume that this group is less careful in managing their personal data than the first group.

Cluster 4: Generous Publishing This group (32 questionnaires, 3 participants) did find a lot of information about themselves, but does not see this as a problem. The members of this group were not surprised about the kind and extent of the information available. We conclude that this group has a less restrained attitude towards publishing personal information, but manages very well which information may be available to others.

The clustering indicates that many of today's Digital Natives control very well which personal information is published on the Internet. Only eight participants (Cluster 3) were negatively surprised about most search results.

V. DISCUSSION

We were surprised to see that the personal information found on the Internet is well balanced over almost all categories we have provided. Only hobbies seem to be over- and friends underrepresented (cf. Figure 1). Furthermore, we were surprised to see that our participants found nothing unexpected in about 80% of all searches (cf. Figure 4), at least nothing they would deem problematic from a privacy perspective. Nevertheless, we have observed that a certain fraction of information has been uploaded by unknown people and without consent and knowledge of the individuals concerned. Our participants also disagree with the general availability and traceability of some information on the Internet (cf. Figure 6). Furthermore, we have observed that the privacy paradox holds, but to a limited extent: Although most information has been uploaded either by or with consent of the study participant, they disagree with the availability on the Internet of only one fourth of the information.

VI. CONCLUSION

Due to the advent of social networking sites on the Internet it has become increasingly tempting for individuals to disclose personal details. The potential for misuse of such information is high. To facilitate the design and realization of future privacy approaches, it is important to know the extent and the characteristics of personal data available on the Internet. In this article we have studied which personal information Digital Natives can find about themselves on the Internet. In particular, we have guided 65 undergraduate students of computer science to search for personal information. We have studied the influence of the search engine on the search result, and we have inquired the impact of personal information publicly available on the privacy of the individuals concerned.

As one result, we have gained evidence that Digital Natives are well-adapted to the privacy problems of the Internet. Our explorative study has shown that privacy perception and privacy behavior is different from individual to individual. In particular, we have observed that different search terms return different results, but different search terms are also known to different people. We interpret this as a trend towards managing different digital identities in order to stay in contact with different persons. It might be an interesting topic of research to design privacy approaches that support different digital identities.

ACKNOWLEDGEMENT

We like to thank Victoria Kayser for her data analyses.

REFERENCES

- [1] D. M. Boyd and N. B. Ellison, "Social Network Sites: Definition, History, and Scholarship," *Journal of Computer-Mediated Communication*, vol. 13, no. 1, 2007.
- [2] E. Buchmann and K. Böhm, "How Much do Digital Natives Disclose on the Internet – a Privacy Study," in *Karlsruhe Reports in Informatics 2012/15, ISSN 2190-4782, Karlsruhe Institute of Technology*, 2012. [Online]. Available: <http://digbib.ubka.uni-karlsruhe.de/volltexte/1000029029>
- [3] P. Norberg *et al.*, "The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors," *Journal of Consumer Affairs*, vol. 41, no. 1, pp. 100–126, 2007.
- [4] A. Acquisti, L. John, and G. Loewenstein, "What is Privacy Worth?" in *Proceedings of the 21th Workshop on Information Systems and Economics (WISE'09)*, 2009.
- [5] N. F. Awad and M. S. Krishnan, "The Personalization Privacy Paradox: An Empirical Evaluation of Information Transparency and the Willingness to be Profiled Online for Personalization," *MIS Quarterly*, vol. 30, 2006.
- [6] S. Pötzsch, *Privacy Awareness: A Means to Solve the Privacy Paradox?*, ser. IFIP Advances in Information and Communication Technology, 2009, vol. 298, pp. 226–236.
- [7] C. Fuchs, "StudiVZ: Social Networking in the Surveillance Society," *Ethics and Information Technology*, vol. 12, no. 2, pp. 171–185, 2010.
- [8] K. Lewis, J. Kaufman, and N. Christakis, "The Taste for Privacy: An Analysis of College Student Privacy Settings in an Online Social Network," *Journal of Computer-Mediated Communication*, vol. 14, no. 1, pp. 79–100, 2008.
- [9] J.-Y. Son and S. S. Kim, "Internet Users Information Privacy-Protective Responses: a Taxonomy and a Nomological Model," *MIS Quarterly*, vol. 32, no. 3, pp. 503–529, 2008.
- [10] R. K. Chellappa and R. G. Sin, "Personalization versus Privacy: An Empirical Examination of the Online Consumer's Dilemma," *Information Technology and Management*, vol. 6, pp. 181–202, 2005.
- [11] Universität St. Gallen, "Abschlussbericht des Projekts Sicherheit vs. Privatheit - Vertrauensfaktoren im Umgang mit Daten und Konsequenzen für die digitale Identität," <http://isprat.net/isprat-projekte>, 2011.
- [12] M. Kurt, "Determination of in Internet Privacy Behaviours of Students," *Procedia - Social and Behavioral Sciences*, vol. 9, no. 0, pp. 1244 – 1250, 2010.
- [13] I. Oomen and R. Leenes, *Privacy Risk Perceptions and Privacy Protection Strategies*, ser. IFIP International Federation for Information Processing, 2008, vol. 261, pp. 121–138.
- [14] K. B. Sheehan, "Toward a Typology of Internet Users and Online Privacy Concerns," *The Information Society*, vol. 18:21/32, 2002.
- [15] L. Sweeney, "k-Anonymity: A Model for Protecting Privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 5, pp. 557–570, 2002.
- [16] L. Rokach, "A Survey of Clustering Algorithms," in *The Data Mining and Knowledge Discovery Handbook*, 2nd ed., L. Rokach and O. Maimon, Eds., 2010, ch. 14, pp. 269–298.