2009

Media Coverage of Online Social Network Privacy Issues in Germany: A Thematic Analysis

Ramzi Rizk
Humboldt-University, Berlin, rizk@wiwi.hu-berlin.de

Daniel Marx
Humboldt-University, Berlin, daniel.marx@cms.hu-berlin.de

Matthias Schrepfer
Humboldt-University, Berlin, srepferm@wiwi.hu-berlin.de

Janina Zimmerman
Humboldt-University, Berlin, zimmerja@wiwi.hu-berlin.de

Oliver Guenther
Humboldt-University, Berlin, guenther@wiwi.hu-berlin.de

Media Coverage of Online Social Network Privacy Issues in Germany: A Thematic Analysis

Ramzi Rizk
Institute of Information Systems
Humboldt-Universität zu Berlin
Spandauer straße 1 D-10178
Berlin, Germany
rizk@wiwi.hu-berlin.de

Daniel Marx
Institute of Information Systems
Humboldt-Universität zu Berlin
Spandauer straße 1 D-10178
Berlin, Germany
Daniel.marx@cms.hu-berlin.de

Matthias Schrepfer
Institute of Information Systems
Humboldt-Universität zu Berlin
Spandauer straße 1 D-10178
Berlin, Germany
srepferm@wiwi.hu-berlin.de

Janina Zimmermann
Institute of Information Systems
Humboldt-Universität zu Berlin
Spandauer straße 1 D-10178
Berlin, Germany
zimmerja@wiwi.hu-berlin.de

Oliver Günther
Institute of Information Systems
Humboldt-Universität zu Berlin
Spandauer straße 1 D-10178
Berlin, Germany
guenther@wiwi.hu-berlin.de

ABSTRACT

The massive growth of online social networking platforms in recent years has brought with it increasing privacy concerns. Rarely do long periods of time pass without some major privacy issue being brought to light. In this paper we present our study regarding media coverage of privacy issues in social networks, and the effects thereof. We gathered all news articles from over 30 German media sources, published between September 1st, 2007 and August 31st, 2008. Using 'thematic analysis', we categorized and coded those articles, and identified the major themes among them. The results showed the main themes of interest in the media, and, more importantly, the very significant effect media coverage has on providers, and on the redesign of privacy-compromising features. The results emphasized the need for more regulation of privacy-preserving practices. This study also set the stage for further inquiries into the topic in U.S. media.

Keywords

Online social networks, privacy, thematic analysis, German media, news, privacy design

INTRODUCTION

In February 2009 Facebook published their new terms of use (TOS) (Facebook Terms of Use, 2009). A blizzard of media coverage (The Consumerist, 2009; Scrobelizer, 2009) and user activism immediately followed, protesting, in particular, a passage in the new TOS that granted Facebook ownership of all user content uploaded to the platform, even after users deactivate their accounts. Following this, Facebook's CEO, Mark Zuckerberg, issued a statement on the SNS official blog (Zuckerberg, 2009) announcing that the old TOS (Facebook Terms of Use, 2008) had been reinstated. This was followed by an announcement asking the platform's users to participate in drafting the new TOS.

Just like Beacon (Facebook Beacon, 2007; Zuckerberg, 2007) and Newsfeed (Zuckerberg, 2006) before it, this incident was yet another case of privacy detrimental changes and features quickly being retracted following user activism and media coverage.

On the other hand, many other cases of misappropriation, contested ownership, information aggregation and separation of identity breaches (Guerses, Rizk and Guenther, 2008) remain unresolved. Facebook's same TOS also state that whenever a user views an advertisement, Facebook shares the user's IP addresses with that advertiser (Facebook TOS 2008). Flickr and (parent company) Yahoo have conflicting terms of use when it comes to sharing user information with 3rd parties (Flickr promises not to do that, Yahoo reserves the right to!). Yet another issue pertains to photo albums on Facebook. Through some conflicting privacy settings, a user might be able to browse a photo album of another (non-friend) user. Facebook applications, another privacy-nightmare, actually have access to all the information available to the users that install them. This means that simply blocking an application does not prevent it from gathering information, assuming a "friend" has that application installed.

Even this current issue with the Facebook TOS is not a new one (Guerses et al. , 2008), the old TOS were not very different from the initial version of the new ones, the only difference is that they, for some reason, did not catch the media's attention.

These cases of privacy breaches clearly show that SNS providers are not proactively seeking out grey areas in the privacy policies and clearing those out, but rather, only react under pressure. Media coverage pressures providers into rethinking their strategies and redesigning their features in order to appease angry users. In a way, this is understandable. Seeing as SNS have had a hard time creating a viable revenue model, their only real asset is user information. Naturally, they would try to own or control as much of that information as they possibly could, in the hope of generating revenue through it.

We believe that media coverage is a primary driver of privacy redesign. In this paper, we set out to analyze media coverage of privacy issues in SNS, in the hope of identifying patterns that would prove our hypothesis that an increase in media coverage of a particular privacy-related issue has the effect of pushing SNS providers to improve their privacy preservation and assurance mechanisms.

The rest of this paper is structured as follows: Section 2 introduces thematic analysis, our methodology of choice for conducting this study. The sources we used and our data corpus and data set are presented. This is followed by generating the codes, defining themes and creating the thematic map. Section 3 analyzes the resulting thematic map, presents general statistical results and discusses interesting trends and specific peaks in media coverage. This leads to our conclusion and next steps in section 4.

RESEARCH METHODOLOGY AND SETUP

Our research is based on the thematic analysis methodology (Braun and Clarke 2006, Aronson 1994). Thematic analysis is a method of analyzing data by identifying patterns and organizing them into themes. The thematic analysis process as described by Braun and Clarke (Braun and Clarke, 2006), simplified, consists of gathering all relevant news articles within a specific time frame and pertaining to the research study. This raw data constitutes the *data corpus*. The data corpus is then checked for relevance, and all invalid or irrelevant content is weeded out. The remaining data is now called the *data set*. From this data set, initial codes are derived and categorized. "Codes identify a feature of the data (semantic content or latent) that appears interesting to the analyst…" (Braun and Clarke, 2006). This coding allows us to organize the data into meaningful groups. Themes are then introduced, acting as umbrella terms for certain codes. Codes are assigned to a theme if the theme semantically combines the meaning of these codes. Themes may also be assigned to other themes, thus becoming sub-themes.
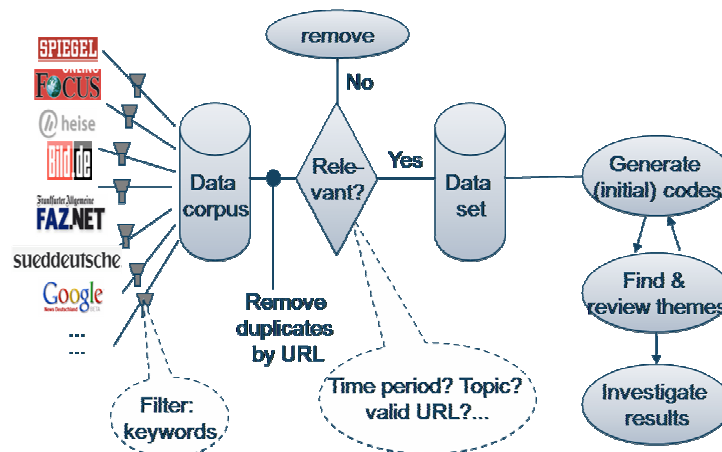

Figure 1. The thematic analysis process of this paper

At this point, all introduced themes are merely candidates. An iterative process follows for checking the validity of the assignments made, and whether a candidate theme actually does cover all the codes assigned to it. Candidate themes may be abandoned, and codes may be moved from one theme to another during this phase. This process continues until all wrongly assigned codes have been relocated to the relevant themes, and/or all invalid candidate themes have been abandoned.

Finally, a thematic map is generated to visualize the themes and sub-themes and their interrelations. This map is then used for our analytical purposes. The figure above visualizes the process followed to conduct this study.

Data Gathering and Data Corpus

In our study, we focused on news articles available through online sources. For that purpose, we gathered articles from aggregated media search engines like Google News (http://news.google.com) and the individual search engines of the publishers. The process included the selection of keywords (as search terms), identifying the search engines to use, and specifying the necessary metadata. All this information served additionally as part of the data corpus.

Selecting Keywords

Selecting keywords was the simplest task that we undertook for this study. We decided to limit our search to Facebook.com, Myspace.com as both are global players and well-established in Germany, in addition to studiVZ.net, Germany's largest online social network and the general term "social network". In order to better define the scope of this study, business networks (such as Xing and LinkedIn) were excluded, as were other large recreational networks. Additionally, we decided on the terms "Privatsphäre"(Privacy) and "Datenschutz"(Data privacy) to filter out non privacy-related articles.

Our search keywords thus became all the possible permutations of the above terms (ex: "Facebook Datenschutz", "MySpace Datenschutz", "social network Privatsphäre", etc…)

Choosing Sources and Gathering Meta Information

We decided on Google news as our aggregated search engine of choice, since it offered more German media results for our keywords than any other engine. Additionally, we searched the online archives of Germany's top daily newspapers, magazines and weekly newspapers and specialized (computer) magazines in terms of circulation. For each of those categories, we chose the top ten whose search engines returned results to our keywords. Although many results here were already made available through Google news, we decided to expand our study to the specific archives for completeness.

Meta information was appended to each article that resulted from our searches. This meta information included the URL (uniform resource locator), date of publication, access date, search engine used, keywords used and source. The purpose of this information is to aid in creating the data set, and eventually offer more dimensions for our analysis of the thematic map later on.

Using eight different search terms, and 31 different search engines, the resulting data corpus consisted of 2412 articles. In the next step, these articles were analyzed for inclusion into the data set.

Generating the Data Set

The meta information and content of all the articles present in the data corpus were assessed for inclusion into the data set. As a first step, redundant articles had to be removed. These redundancies could have occurred due to different search terms returning the same results (identical URLs), or various news sources using the same article (different URL but same date of publication and title). This step allowed us to exclude 830 articles from our data corpus.

The next step consisted of checking the remaining 1582 articles for relevance. Articles may be found irrelevant since they don't fall within the specified time frame of our study (some search engines do not allow filtering by date), since they might mention our search terms but not really be an article on that topic, or finally since the link might be dead or require a fee.

259 results did no lie within the time period (September 1st 2007 – August 31st 2008). Another 208 results were excluded because they linked to invalid pages or required a fee to be accessed, and finally, 782 did not match the topic at hand.

The remaining 333 articles (14% of the data corpus) were admitted into the data set and are relevant for our analysis.

The data set articles came from 85 different sources, and were found using 20 different search engines. Spiegel.de, with 12% of all articles is the largest source, followed by Computer woche (9%), and PC Welt and Heise.de (with 5% each). 52% of the articles were found using the Google News archive.
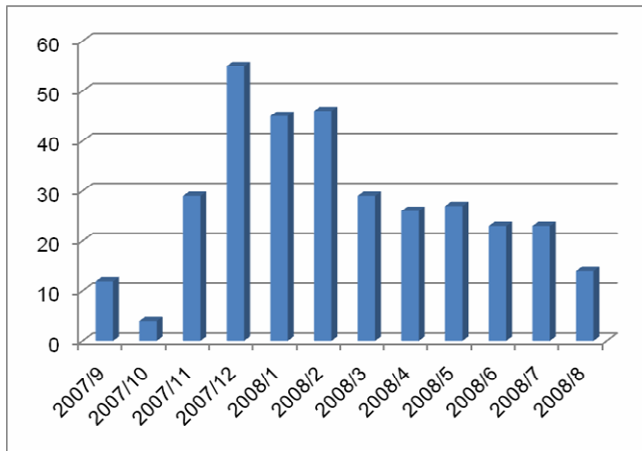

Figure 2. Articles per month   **Figure 3. Percentage of articles per search term**

StudiVZ is mentioned in 56% of the articles in our data set. Facebook is a close second, with 53% of all articles mentioning it. MySpace earns a mention in 26% . 10% of the articles mention social networks without specifying a particular one, and 18% mention other (German) platforms such as MeinVZ.net and Lokalisten.de.

On average, we identified 28 relevant articles per month. The months of December and February had more articles. Additionally, there were significantly more results for StudiVZ and Facebook than for MySpace and social network in general. These basic results already imply that our hypothesis had some truth to it, as we will discuss later.

Generating Codes and Categorizing Articles

At this point, all articles have been scanned or read in order to create the data set. The articles are now read again to generate the codes. According to (Braun and Clarke, 2006) these codes underline interesting features of the articles, and are mapped to meaningful groups which describe the semantic content of these features. Since we are dealing with the data set, all articles analyzed here are relevant and must generate at least one code. Naturally, an article may generate multiple codes, since each code describes one interesting feature, and is documented individually.

Following this, the articles were categorized according to the online social network they discuss. This allows us to correlate privacy issues with platforms, and thereby emphasize what issues the German media associates each platform with. That information (platform discussed, extracted codes) is then added to the data set for each article. The possible platforms are: Facebook, MySpace, StudiVZ, other platform, and social networks in general.

In order to assure that the codes are complete, three researchers analyzed each of the articles and generated their own codes. These codes were then compared and unified before starting the next phase.

Defining Themes and Creating the Thematic Map

The codes generated above are then used to identify candidate themes. At first, each code introduces (or, in fact, becomes) a new theme, however, patterns and associations soon emerge, sub-themes are created, and more codes are associated with existing themes. It is critical, when choosing themes, to assure that their semantic context does not overlap.

The thematic map initially contains only the root node "privacy issues in social networks", our study topic. The first code initializes the map by introducing a new theme. Following codes may be assigned to the root node, an existing node, or become a new node. Extending the thematic map, in the early phases, often leads to a complete restructuring. Further down the process, a traversal of the map from the root nodes to the lower ones shows an increasing level of detail in terms of the privacy issues discussed.
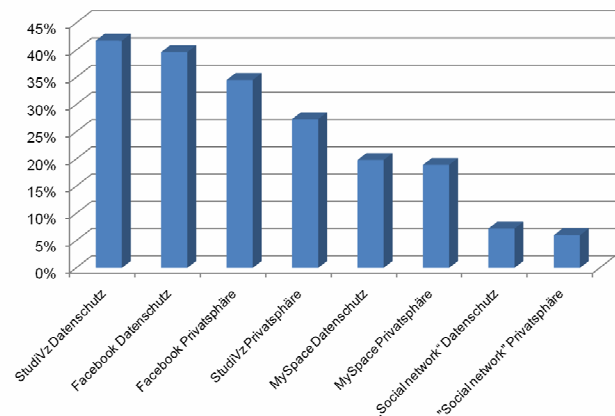
Once all the codes have been mapped, an iterative process starts, in which each article in the data set is read once again, and the extracted codes are matched against their corresponding themes. If the matching holds, the relationship is considered accurate. Mismatches lead to new themes, or modifications of existing ones, such as the new theme is a better semantic match to the code/article.

Data protection	350	Data collection	49	Storing of user activities	5		
				No possiblility to delete personal data	26		
				"Internet vergisst nichts"	18		
		Data abuse	44	Data theft	12		
				Publication of private information of others	10		
				Identity grabbing	19	Hacking of accounts	8
						Abuse of identities	11
		User information	11				
		Law	14	Country specific	5		
				Illegal activities	9		
		Forwarding of personal data	129	Selling personal data	32		
				Forwarding information to authorities	7		
				Beacon	40		
				Interfaces	46	3rd party software	3
						APIs	6
						Accessibility for search engines	27
						Platform independant single-sign-on	10
		Data protection (technical view)	65	Spam on platform	2		
				Profile settings	38	Unknown or complex profile settings	2
						Lack of profile settings	27
						Insecure initial settings	9
				Software bugs	19	Malware	4
						Hacking	2

Figure 4. Data protection and associated sub-themes

Advertisement	121	Personalized Advertisement	110		
		Behaviour targeting	11		
Personal behaviour	80	Misleading impression of security	10		
		Propagation of personal data	70	Falsification of data	2
				High willingness to propagate data	56
				Trend to better protection of data	12

Figure 5. Advertisement, Personal behavior and associated sub-themes (with number of matches)

In all, 43 themes were identified (including the root node). Some of these themes are self-explanatory; others are a little more detailed. For brevity, we cannot include a detailed explanation of each theme in this paper; however, we will discuss the main topics identified, and some interesting observations. The complete results and descriptions will be made available online as an appendix.

The three main privacy issues from the media perspective, according to our results are: "data protection", "advertisement", and "personal behavior" (which cover the actions and expectations of users). In figures 4 and 5, the sub themes for each of these major themes are listed.

24% of the articles discussed personal behavior (such as high willingness to propagate personal information). Advertisement and its sub-themes are discussed in 35% of all articles. Additionally, every article in our data set mentions data protection.

RESULT ANALYSIS

The previous section already presented some basic statistical results pertaining to the data set and thematic map. In this section, we will take on the bigger picture of trends and peaks of media coverage, and identify the instances that support our hypothesis. First, figure 6 shows the main themes in our thematic map, and the percentage of articles in which they are discussed.
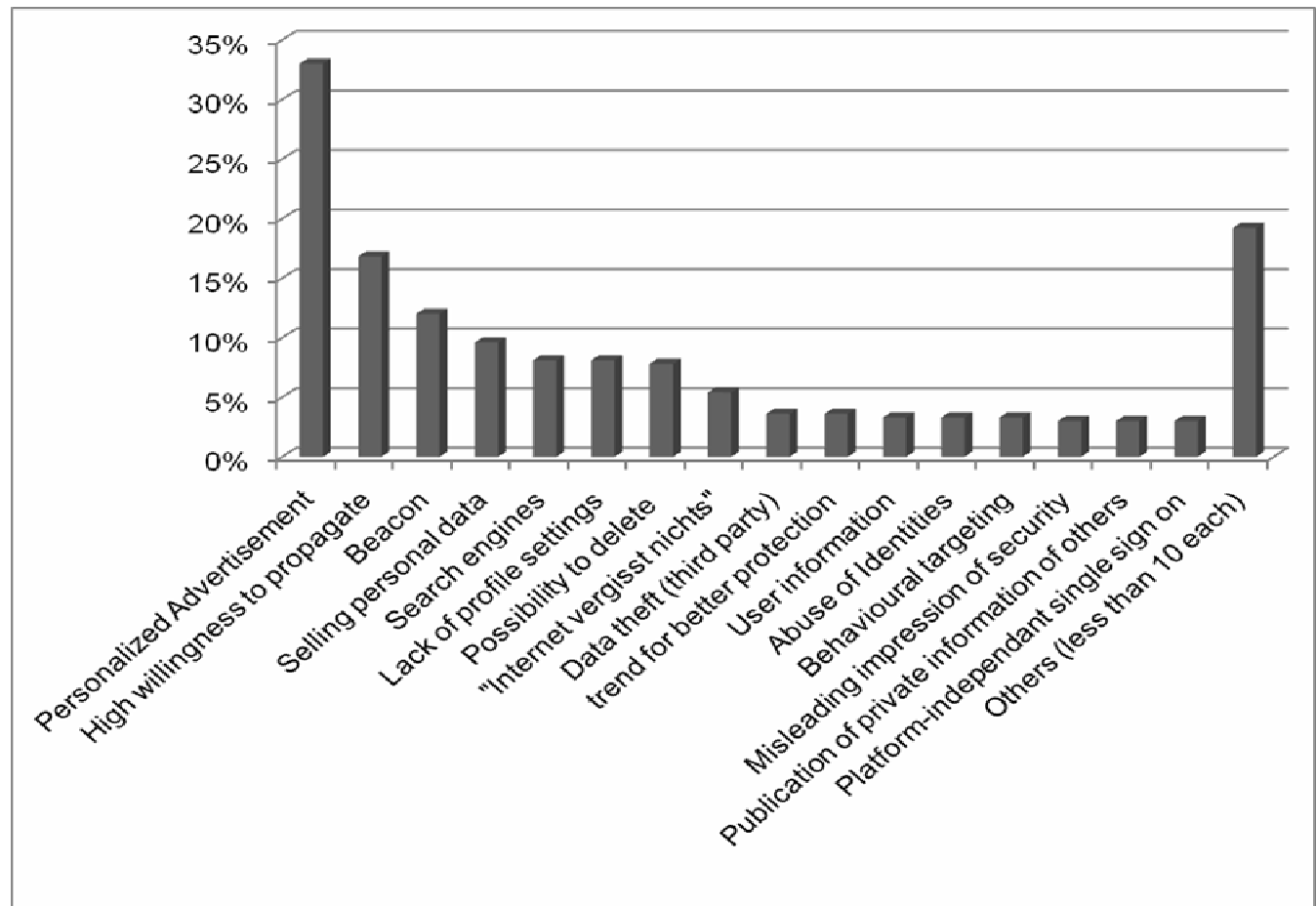

Figure 6. Percentage of articles per sub-theme

Interesting Trends

"Personalized Advertisement" appears in one third of all the articles. In most cases, this theme appears alongside discussions regarding the new general terms and conditions StudiVZ in December 2007. The high willingness of users to propagate personal information ranks second, followed by discussions of privacy issues related to Facebook's Beacon. The last sub-theme in figure 7 ("Others") is a catch-all for the remaining sub-themes. Each of those themes are mentioned in less than 3% of the cases (or 10 articles).

Another way to view these results is by associating each sub-theme with the related online social network. Most discussions of Facebook involve the sub-themes "Beacon", "Personalized advertising", "No possibility to delete", "Lack of profile settings" and "Search engines". StudiVZ brings up the issues "Personalized advertisement", "Selling personal data", "User information", "Illegal activities", and "Insecure initial settings". MySpace conjures up thoughts of "High willingness to propagate", "Hacking of Accounts (Identity grabbing)", "Malware", "Lack of profile settings" and "Search engines". Finally, social networks in general are associated with the sub-themes "High willingness to propagate", "Personalized advertisement", "Internet vergisst nichts" (literally: the internet does not forget, in reference to the permanence of information posted online, and young user's apparent unawareness of that fact), and "Selling personal data".

Finally, we made two more observations of note. First, "High willingness to propagate" is a constant theme that appears regularly in news articles (see figure 7), and seems to be a general concern when it comes to privacy in social networks. On a more optimistic note, towards the end of the time period we examined, we found an increase in articles related to users adopting better privacy practices. This may very well also be due to the extent of media coverage of the topic educating users and leading them to take advantage of the privacy features on offer.

Specific Peaks

In addition to identifying the general trends of media coverage, we wanted to find evidence of our hypothesis that said coverage actually leads to providers redesigning their privacy settings. The peaks with respect to certain sub-themes and the timing of those peaks definitely support our hypothesis. In particular, we identify two peaks related to Facebook Beacon's release and subsequent phasing-out, and StudiVZ's new general terms and conditions, and their subsequent modification. Figure 8 shows the most discussed topics over the time period of our study.
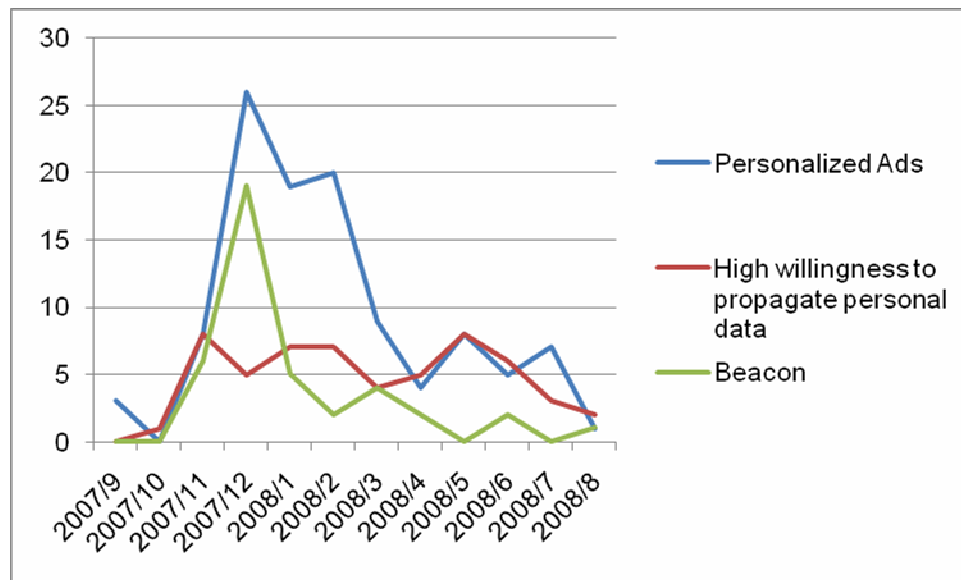

Figure 7. Peaks and most discussed sub-themes over time

Facebook Beacon was a tracking system, integrated on external platforms. It reported user activities to Facebook and informed a user's contacts of those activities. Facebook originally marketed Beacon to potential businesses by stating 'The user can choose to opt out of the story, but the user doesn't need to take any action for the story to be published on Facebook' (Facebook Beacon 2007). Their policy later changed, and that statement was replaced by 'The user must proactively consent to have a story from your website published' (Facebook Beacon 2008). Facebook's CEO Mark Zuckerberg issued an apology on his blog in mid December 2007 and announced the possibility to deactivate Beacon (Zuckerberg 2007). The service was gradually phased out.

Observing figure 7, media coverage of Bacon runs quite obviously parallel to this sequence of events. Coverage of the features starts soon after its launch, and reaches a peak in mid-December. Following Zuckerberg's apology, and the restrictions that followed, coverage decreased steadily.

The same pattern can be observed with regards to StudiVZ and "Personalized advertisement". StudiVZ changed their general terms and conditions in December 2007. Users were required to accept the new terms if they were to be allowed to remain on

the platform. Complaints regarding these new conditions hovered around the clause that gave StudiVZ the right to sell user data to third parties for marketing analyses (Schulte, 2007; WELT, 2007). Another issue was the fact that all users were now forced to accept personalized advertising. Under pressure, StudiVZ updated these terms in mid December 2007 (Kleinz, 2007). The updated terms explicitly stating that they are not allowed to sell user data, and offering users the option to opt-out of personalized advertisements.

Again, figure 7 shows a peak in media coverage right before StudiVZ's change of heart.

One could argue that this coverage is merely a reaction, and not a driver. We will discuss that in our conclusion.

Due to space limitations, we selected only a subset of our results to present here. The entire data set with links to all the analyzed articles, in addition to the thematic map, definitions of the sub-themes, and various tables and charts pertaining to the results will be made available online.

CONCLUSION

In this paper, we presented a thematic analysis of German media coverage of privacy issues in social networks. We retrieved 2412 articles from 31 sources matching our keywords, whittled that down to 333 relevant articles that constituted our data set. We then coded the articles according to their points of interest and created a hierarchical thematic mapping of those codes. The results were then analyzed across the various networks, by theme and over time. Results of special interest, particularly to our hypothesis were further presented in detail.

In the introduction we claimed that media coverage is a major driver for privacy design. The results of this study show that the redesign of privacy features does indeed coincide with increased media coverage of those features. One could chalk this up to the media simply reacting to ongoing issues and covering the news, as it is. However, that argument has one major flaw. Over the past two years, the only major redesigns when it came to privacy settings were preceded by a lot of media coverage. Among them, Facebook's Beacon and new Terms of Use, and StudiVZ's general terms and conditions. Other issues, such as Facebook's original terms of use, their sharing of user IPs with advertisers, and the ability to view photos of non-friends, as well as Flickr and Yahoo's conflicting Terms of Use have either gone unnoticed, or did not gain traction in the media. These issues remain to this day unresolved.

A recent entry on Consumerist.com presents the case of a deceased journalist and his Facebook page (The Consumerist, 2009a). After his passing and even though the family requested otherwise, Facebook kept his profile page, turning it into a "Memorial". Such situations would not arise in the offline world, where families and individuals are protected by laws. However, now that this story is being pushed in the social web, we believe it is a matter of time before Facebook succumbs to the family's wishes.

Manufacturers, upon discovering flaws in their products, automatically recall the affected items. The legal systems has very clear rules when it comes to consumer products. Online social networks, due to their relative youth, have much less regulation. Due to the simple nature of this evolving system, the importance of user advocacy and media pressure cannot be underestimated.

Social media on the internet is merely a reflection of users and their issues. In that sense, media coverage is driven by users expressing their opinions and advocating for their own rights. This means that users can directly influence changes that better protect their privacy.

In conclusion, we have started conducting this same study with a focus on US media coverage of said issues. We believe that the results of that study will further confirm statements made in this paper, and help us form an even better picture of the issues discussed here. Additionally, we are in the process of launching a collaborative website where privacy topics and user concerns relating to privacy in social networks can be voiced and analyzed by experts. Through that website, we hope to harness the power of social media, shed light on privacy breaches and violations, and play a proactive role in driving the redesign of privacy features for the benefit of users.

REFERENCES

1. Aronson, J. (1994) A Pragmatic View of Thematic Analysis, The Qualitative Report, 2, 1. retrieved on 09.09.2008 from http://www.nova.edu/ssss/QR/BackIssues/QR2-1

2. Braun, V. and Clarke, V. (2006) Using thematic analysis in psychology, Qualitative Research in Psychology, 3, 77-101.

3. Facebook (2008) Terms of Use, retrieved in September 2008 from http://www.facebook.com/terms#/terms.php?ref=pf

4. Facebook (2009) Terms of Use, retrieved on Feb 6, 2009 from http://www.facebook.com/terms#/terms.php?ref=pf

5. Facebook Beacon (2008) retrieved in March 2008 from http://www.Facebook.com/business/?Beacon.

6. Facebook Beacon (2007) retrieved in Dec. 2007 from http://www.Facebook.com/business/?Beacon.

7. Guerses, S., Rizk, R. and Guenther, O. (2008) Privacy Design in Online Social Networks: Learning from Privacy Breaches and Community Feedback, in proceedings of ICIS 2008, http://aisel.aisnet.org/icis2008/90.

8. KLEINZ, T. (2007) StudiVZ ändert AGB: Doch keine SMS- und Messenger-Werbung retrieved on 01.19.2009 from http://www.heise.de/newsticker/StudiVZ-aendert-AGB-Doch-keine-SMS-und-Messenger-Werbung-Update--/meldung/100642, 12/15/2007

9. SCHULTE, M. (2007) Was StudiVZ-Nutzer jetzt beachten müssen retrieved on 01.19.2009 from http://www.welt.de/webwelt/article1459528/Das_Geschaeft_mit_den_Nutzerdaten.html, 12/14/2007

10. Scrobelizer, R. (2009) User data ownership on Facebook and why it doesn't matte, posted on February 17th 2009, retrieved on February 20th 2009 from http://scobleizer.com/2009/02/17/ownership_facebook/

11. The Consumerist (2009) Faceboook's New Terms of Service posted on February 15th, 2009, retrieved on Feb 25th, 2009 from http://consumerist.com/5150175/facebooks-new-terms-of-service-we-can-do-anything-we-want-with-your-content-forever

12. The Consumerist (2009a) Facebook Won't Let You Remove Dead Relative's Page, Per "Policy", posted on February 20th 2009, retrieved on February 22th 2009from http://consumerist.com/5157481/facebook-wont-let-you-remove-dead-relatives-page-per-policy

13. WELT (2007) Das Geschäft mit den Nutzerdaten, retrieved on 01.19.2009 from http://www.welt.de/webwelt/article1459528/Das_Geschaeft_mit_den_Nutzerdaten.html, 12/14/2007

14. Zuckerberg, M. (2007) Thoughts on Beacon, posted on December 5th 2007, retrieved on 01.19.2009 from http://blog.facebook.com/blog.php?post=7584397130

15. Zuckerberg, M. (2007) An Open Letter, posted on September 8th 2006, retrieved on 01.19.2009 from http://blog.facebook.com/blog.php?post=2208562130