

QUT Digital Repository:
<http://eprints.qut.edu.au/>



Nicolai, Thomas and Kirchhoff, Lars and Bruns, Axel and Wilson, Jason A. and Saunders, Barry J. (2009) *The self-Googleing phenomenon : investigating the performance of personalized information resources*. First Monday, 14(12).

© Copyright 2009 [please consult the authors]

The Self-Googling phenomenon: Investigating the performance of personalized information resources

Thomas Nicolai, Lars Kirchhoff

Institute for Media and Communications Management, University of St.Gallen, Switzerland

lars.kirchhoff@unisg.ch, thomas.nicolai@unisg.ch

Dr Axel Bruns, Dr Jason Wilson, Barry Saunders

Creative Industries Faculty, Queensland University of Technology, Brisbane, Australia

a.bruns@qut.edu.au, j5.wilson@qut.edu.au, b.saunders@qut.edu.au

ABSTRACT

This paper investigates *self-googling* through the monitoring of search engine activities of users and adds to the few quantitative studies on this topic already in existence. We explore this phenomenon by answering the following questions: To what extent is the *self-googling* visible in the usage of search engines; is any significant difference measurable between queries related to *self-googling* and generic search queries; to what extent do *self-googling* search requests match the selected personalised Web pages? To address these questions we explore the theory of narcissism in order to help define *self-googling* and present the results from a 14-month online experiment using Google search engine usage data.

INTRODUCTION

Searching for information about oneself has become fairly popular with the rise of modern search engines and social uses of the Web in recent years. Due to a constant growth in social networking sites and other forms of social applications on the Internet and mobile devices, more and more people publish personal (and to some extent even very personal) information about themselves on the World Wide Web (WWW).

This development has increased awareness of the possibility of tracking personal and related information by using search engines. Even social networking sites like *Xing.com* and *LinkedIn.com* have adapted their information architectures to be in line with such self-reflective usage in order to optimize the targeting of their audiences on search engines. Additionally, people search engines (new search engines focusing on information about people) have gained popularity by recognizing people's interest in information about other people on the WWW (Arruda and Dixson, 2007). Therefore, end-users have increased their efforts in actively monitoring and shaping their personal information online.

However, only a few studies explore this phenomenon and the underlying motivations behind such behavior. One approach to explain and understand this practice of monitoring the Web often, which is described as *self-googling*, includes the theory of narcissism, which helps explain the phenomenon of people searching and browsing the Web for information about themselves (Contrada, 2004). *Egosurfing* and *egogoogling* are different names describing the same “practice of harnessing the Internet's vast data-collection powers to dig up information about oneself,” as Glasner (2001) puts it. This practice involves a way of observing the social construction of personal reputation, and a means of managing the personal “brand” through self-marketing (Lampel and Bhalla, 2007).

THEORY OF NARCISSISM

Narcissism is generally perceived as a growing social phenomenon in western society, as Lasch describes it in his widely cited book *Culture of Narcissism* (1991). Recent technology shifts have yet again started a discussion about the increase of narcissistic behavior in western culture (Kopelman, 1984; Mullins and Kopelman, 1984; Nelson, 1977), such as the technologies that enable social interaction in general, and social networking sites in particular (Halavais, 2007). In spite of any possible negative connotations of this phenomenon, narcissism can provide a functional and healthy personal strategy for making sense of our increasingly fast and techno-oriented world. Emmons (1987) and Mullins and Kopelman (1984), then, argue that narcissism operates as a cultural and social entity.

Defining Narcissism

Narcissism has its origins in the ancient Greek mythology and is usually perceived as a personality disorder in mainstream psychology and psychiatry literature. Katz (1993) defines narcissism as an insensitiveness to other people’s interests and emotions. He further argues that people showing narcissistic behavior focus excessively on their own image and their perception by others. Emmons (1987) characterizes narcissism in the way that individuals easily appreciate success but do not accept defeat and refuse jointly defined social objectives. Campbell (2001) has criticized this perception of narcissism as a dynamic construct with negative connotations for an individual. Rather, he interprets narcissism as a positive phenomenon, and suggests that we need to make a distinction between abnormal and normal narcissistic behavior.

In this paper we accept Campbell’s argument; narcissism can be a functional and healthy strategy for dealing with the growing complexity of our modern technological world, and therefore narcissistic behavior is accepted as a cultural and social entity (Emmons, 1987; Mullins and Kopelman, 1984). This definition is consistent with Freud’s interpretation of narcissism; he explains the motivation of narcissistically behaving individuals as driven by their survival and self-preservation instinct (Alford, 1987).

However, we distinguish between a clinical and a cultural narcissism (Kopelman, 1984; Lasch, 1991; Nelson, 1977). Our research follows the argumentation that cultural narcissism is based on the hypothesis that narcissistic behavior is a highly common personality characteristic found in all modern societies. Therefore

narcissism is not necessarily perceived as a personality disorder, but (in the online context) rather also as a healthy strategy to deal with the increased complexity of everyday life through social media. The introduction of social networking platforms like *Facebook*, *Bebo* and *LinkedIn* has led to new forms of self-representation in certain online areas. In order to keep up with this development in business or private life, users increasingly need to adapt their behavior to this complexity in everyday life. People working in the service sector cannot remain absent from platforms like *LinkedIn*; even industry is using these platforms more and more to recruit staff. Online self-representation is becoming ubiquitous.

The Narcissistic Personality Inventory

The most specific approach to understand narcissism is proposed by Raskin *et al.* (1981; 1988; 1979) through their efforts to develop an empirical method to measure the level of narcissism in society. The *Narcissistic Personality Inventory* is a 54-part questionnaire-based methodology to measure the narcissistic evolution of society. The research methodology is based on the definition of narcissism as a self-focused concentration on one's own behavior and image as outlined above. Their 2006 research project examined 16,475 college students using the *Narcissistic Personality Inventory* methodology. While comparing the findings with results from 1982, they found that two-thirds of the students showed a score above average. Furthermore they found evidence that narcissistic behavior had increased by more than 30 percent between 1982 and 2006.

These consolidated findings and definitions of narcissism lead us to conclude that ***self-googling can be defined as a self-focussed concentration of the attention of an individual to themselves by actively monitoring and shaping their persona and perception online.*** The increase in narcissistic behavior which Raskin *et al.* (2006) measured may be explained at least in part by the ongoing evolution of social media: this evolution has shifted attention to people who usually have remained outside of the scope of traditional media. Consequently, more people experience and must learn how to deal with this attention on single individuals. The theory of narcissism can help to explain and understand the *self-googling* phenomenon.

RESEARCH METHODOLOGY

Drawing on our definition of *self-googling* and the motivations behind it, we investigate changes in search engine usage patterns and the adoption of *self-googling* practices. The methodology used to explore these assumed changes is based on an analysis of search engine usage by online users. Previous attempts to explain and analyze the search behavior of users fall into three distinct categories: (1) those that primarily use transaction log analysis, (2) those that involve users in a laboratory survey or other experimental setting, and (3) those that examine issues related to or affecting WWW searching (Jansen and Spink, 2006; Spink, Jansen, and Pedersen, 2004). Our research methodology and the results presented clearly belong to the first category. However, none of the previous attempts made an explicit differentiation between generic search term keywords and personal names in order to investigate the *self-googling* phenomenon. Only a vague category of "people, places or things" has previously been analyzed by Jansen and Spink (2006), which shows a growth of search terms in that category from 21.5 percent in 2001 to 41.5 percent in 2002.

To investigate the *self-googling* phenomenon we have gathered and analyzed 2.46 million search engine requests over a time period of 14 months. On the basis of this data we have conducted an in-depth analysis of the search terms used, in combination with the links eventually selected and clicked on by the user from the search engine results lists. We address the gap in the available research literature on *self-googling* by more specifically comparing search and click-through trends for *personalized information resources* and *non-personalized information resources*. *Personalized information resources (I)* are Web pages which specifically refer to information about individual persons, whereas the Uniform Resource Identifier (URI) of the Web page includes a version of the person’s name in the URI path. *Non-personalized information resources (II)* refer to pages which do not contain information about a specific person, whereas the URI does not contains any name-specific element in the URI path. Instead they are generic in nature. The detailed construction of the two different URI schemas is presented below (figure 1).

Type	Protocol	Domain	Path
I	http://	www.example.com/	sergey_brin
II	http://	www.example.com/	298055

Figure 1: Personalised and non-personalised information resource identifier

In particular, we examine if there is evidence that generic or personal name queries show proof for a measurable difference in the click-through performance of search engines and we analyze the quality of search queries for personalised resources. For this, we have created 7 million personalised Web pages and 20 million non-personalised Web pages. By having these pages retrieved by the Google Web crawler and included into the Google’s search index we were able to retrieve search engine usage data for analysis purposes. We used the referrer mechanism of the HTTP protocol to collect the required search engine usage data. The referrer mechanism submits the URI of a source Web page and additional information to the destination server each time a user clicks on a link. Web servers can then detect a user’s location and which external Web page “refers” to a Web page on the destination Web server. The referrer information is typically stored within the log files of Web servers and in most cases also made available to server-side applications by the Web server.

A typical log file entry, which contains a referrer, is presented below.

```
218.41.xxx.xxx - - [15/Sep/2006:02:50:16 +0200] "GET /sergey_brin HTTP/1.1" 200 5516
"http://www.google.com/search?num=50&hl=ja&q=Sergey+Brin+1997+correlation+association+rules&lr="
"Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322; .NET CLR 2.0.50727)"
```

This log file example shows that a user viewed a Web page, which they accessed on 15 September 2006 under /sergey_brin on our Web server. This particular Web page provides information about Google founder Sergey Brin. The second line shows the referrer information in this log file entry and contains the full URI of the Web page on which the user found a link to the page on our Web server.

```
http://www.google.com/search?num=50&hl=ja&q=Sergey+Brin+1997+correlation+association+rules&lr=
```

In this case, the Web page was a page generated by the Google search interface. As Google uses the URI to pass the search query parameters of each search to the search engine, all information about the search can be found in the URI and is therefore also passed on to our Web server as part of the referrer information. From the above URI, we can determine that the user was searching for “Sergey Brin 1997 correlation association rules”:

```
q=Sergey+Brin+1997+correlation+association+rules
```

Additionally, the log file record contains the user’s IP address (made anonymous in this paper), and some information about their browser and operating system as well as an access timestamp.

ANALYSIS

To investigate the *self-googling* behavior we analyzed 2.46 million of these search engine referrals to our Web servers. We limited our data analysis to the Google search engine. Because of the current dominance of Google in the search market, we ignored referrers from other search engines, which accounted for less than one percent. Analysis started in August 2006 and ran for 14 months, to October 2007. Initially we identified and extracted all Google referrals from more than 100 million log file entries containing URIs from Google. From these extracted referrals we retrieved the search terms and the chosen Web page (landing page) on our server.

In analysing these data we first examined the distribution of personalized and non-personalized search queries over the 14-month time period of the experiment. Second, we determined the absolute growth in search requests for both categories (personalized and generic) over the whole period. Finally, we selected those log file entries that contained a personalized Web page and a referrer. We then assigned a score that quantifies the match between the search query and the Web page that the user accessed on our Web server. This analysis was motivated by the question: to what extent does a search request for personalized Web pages match the search query and therefore the name of the person?

RESULTS

Our presentation of the research results is divided into three sections which consider the three questions outlined in the introduction. First, we address the question to what extent the *self-googling* phenomenon is traceable through the usage of search engines, and whether we can measure a significant difference between *self-googling* and generic search queries.

Figure 3 presents the distribution of search requests and clicks on the hyperlinks of personalised Web pages in the search engine's results lists that point to personalized Web pages. In the first 11 months of our study we saw a steady growth in retrieved personalized Web pages. At its peak, users selected 14,000 personalized Web pages each day. This number declined to 8,000 search requests by August 2007. The trend for non-personalized Web pages looks similar for the first months. However, access to non-personalized Web pages (figure 4) shows a different trend for the later months of our experiment, peaking at 6,000 search requests per day only. We also witnessed a regular decline in search requests for personalized and non-personalized Web pages on weekends. Additionally, there is a difference in the total amount of personalized and non-personalized Web pages that were accessed: personalized Web pages were accessed more frequently, by an average factor of two.

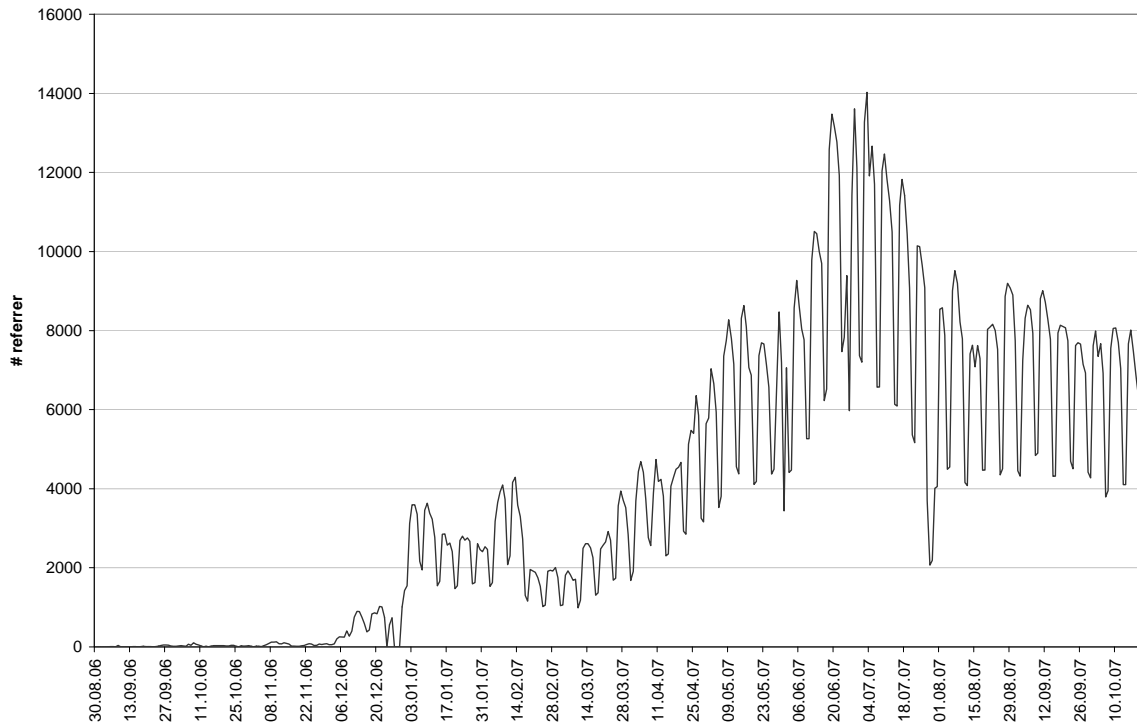


Figure 3: Access to personalized web pages following a referral from Google

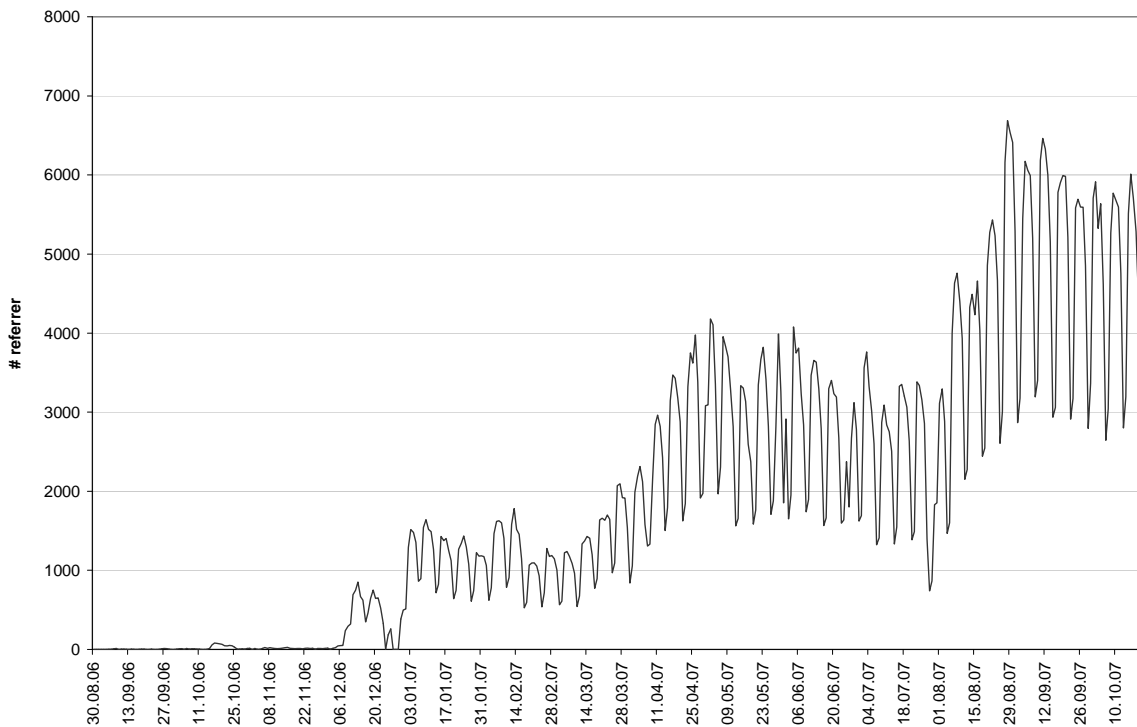


Figure 4: Access to generic web pages following a referral from Google

For a comparison of the number of search engine requests for personalized and non-personalized (generic) Web pages over time, we have integrated both page types into a single chart. Figure 5 presents the difference between both page types by showing the total number of Web page requests over time. The chart clearly demonstrates a difference in growth over the time period of our experiment.

Demand for personalized Web pages grew more strongly than demand for non-personalized Web pages.

Over the lifetime of the study we measured 1.66 million search requests for personalized Web pages and 0.8 million search requests for non-personalized Web pages. The number of requests for personalized Web pages is double the number for generic Web pages, and the 1.66 million search requests for personalized Web pages are distributed across 0.8 million different Web pages. On average, every personalized Web page was requested 2.07 times. Non-personalized Web page requests show a distribution over 0.36 million different Web pages and an average of 2.22 requests per Web page, which is slightly higher.

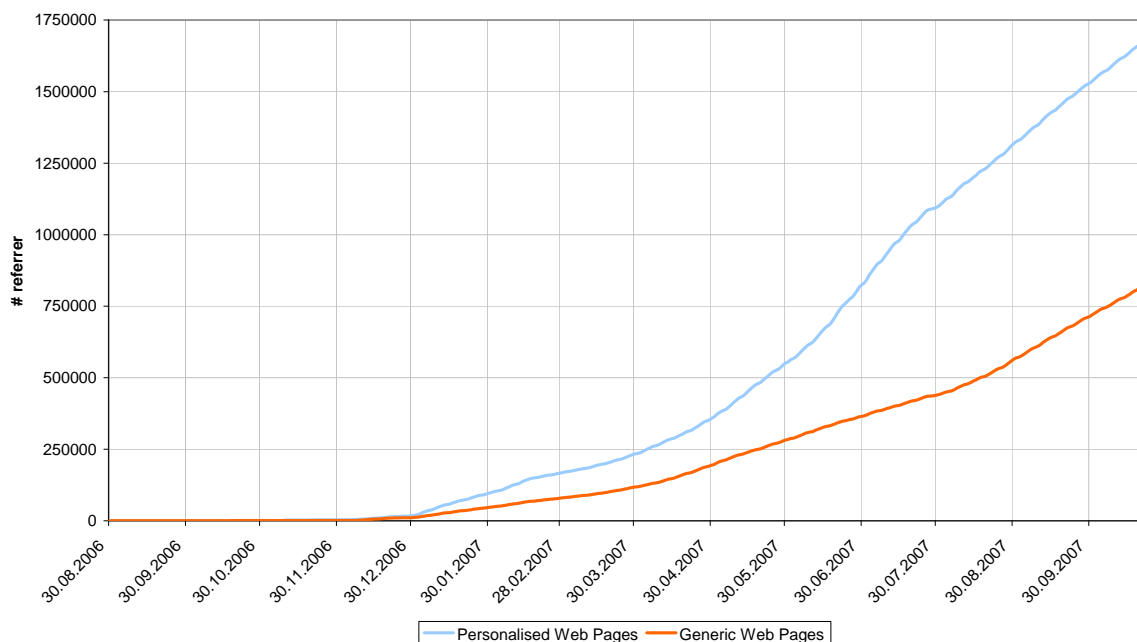


Figure 5: Growth of personalized and non-personalized page requests

Finally, we investigate the quality of correspondence between the personalized search queries and the hyperlinks selected from the result lists. We analyzed the difference between the search query and the selected URI path (see figure 1) to see what level of correlation the selection shows. Based on an algorithm proposed by Oliver (1993), we analyzed the 1.66 million search engine referrals related to personalized web pages. Figure 6 presents the distribution of matching scores against percentiles of search queries.

The graph shows a significant distribution towards a matching score of 100 percent between a search query string and the selected URI path string. Almost 0.3 million of the analyzed 1.66 million referrer show a matching score of 100 percent, which means that search engine users entered the name of the person they were searching for exactly and without errors. The second significant characteristic revealed a matching score

between 95 and 92 percent. However, between 99 and 98 percent the data shows an anomaly in the distribution. This deviation is a potential effect of Oliver’s matching algorithm itself, and of its dependence on the length of the strings being compared. A perhaps more interesting and divergent explanation points to a “self-correcting” mechanism in human cognition that prevents us from very slightly misremembering a person’s name – a mechanism which leads us to make either no error at all, or a more substantial error which in our rating manifests as a deviation of more than 4 percent from perfect recall. Further analysis of our data would be required to uncover evidence supporting this assumption.

Overall, we note that 1.1 million search engine referrer show a matching score above 50 percent between the search query and URI path string. On average we measured a matching score of 64.74 percent, with a standard deviation of 32.52.

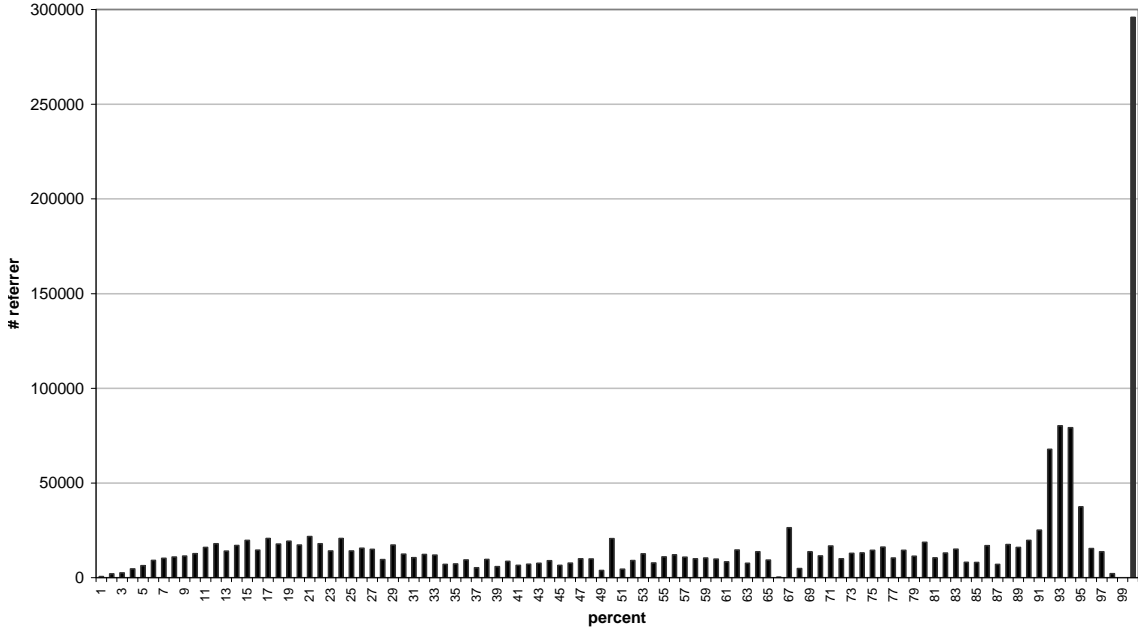


Figure 6: Similarity between search terms and selected personalized Web page URIs

DISCUSSION

To investigate the *self-googling* phenomenon, we have extracted and analyzed 2.46 million search engine referrers from a set of 100 million log file entries in a 14-month experiment. The results point to a higher demand for personalized information resources in comparison to generic Web pages. Furthermore, we analyzed the quality of matches between personalized search queries and selected hyperlinks (URI paths) and found a correlation between them. More than 0.3 million referrals from searches show a matching score of 100 percent out of 1.66 million search engine referrals for personalized web pages, and we see a very high quality of interaction with personalized Web pages through search engines.

The results from our experiment indicate that users show a higher interest in personalized Web pages. The high degree of correlation between search query and selected hyperlink (URI) for personalized Web pages, in contrast to access to non-personalized Web pages, provides evidence of a significant *self-googling* phenomenon. Having defined *self-googling* as **a self-focussed concentration of the attention of an individual to themself by actively monitoring and shaping their persona and perception online**, we conclude that there is a growing interest in personalized Web pages. The significant difference in interest in personalized versus generic Web pages might explain the overwhelming success of social media applications on the WWW.

We argue that our investigation supports the hypothesis that the WWW is transforming itself from a Web of documents and hyperlinks into a Web of social relationships (Kirchhoff, Stanoevska-Slabeva, Nicolai, and Fleck, 2008).

Limitations

While measuring the performance of personalized and non-personalized information resources in terms of search query/Web page access correlations, it is clearly impossible to separate between *people-googling* and *self-googling* requests using our methodology. While *self-googling* describes the action of searching for and monitoring information about oneself, *people-googling* refers to the action of searching for information about somebody else. From the point of view of our server logs, both are indistinguishable, and only a combination of quantitative and qualitative approaches can reliably separate these two types of search requests for personalized Web resources. The scope of our research project may be too limited for generalization, and further research needs to be undertaken. Nevertheless, we argue that this quantitative approach provides new insights into the adoption of social media, and in particular into the *self-googling* phenomenon.

CONCLUSION

Exploring the *self-googling* phenomenon by applying a quantitative approach has provided new insights into the adoption of social media. Our research results point to the rise of narcissism as a cultural entity in society that may explain in part the evolution of social media services, and vice versa. Our results add some evidence to the hypothesis that a growing narcissism, related to the rise of social media, is visible in the *self-googling* phenomenon. Of course some of this 'narcissism' may also be externally enforced, as the perceived need to maintain a public persona drives users to self-google (and thus keep track of what information about them others may be able to find) more regularly. We note again in this context that our study does not investigate the *motivations* behind the self-googling phenomenon – only the evidence for its existence.

We investigated the *self-googling* phenomenon by measuring the level of access to personalized and generic Web pages from the search engine result lists. Our results show some significant differences between the two page types. The performance of the two types shows that personalized Web pages are twice as popular as generic Web pages. Search engine results lists indicate user bias in favour of personalized Web pages, which could be the subject of future research.

ABOUT THE AUTHORS

Thomas Nicolai was Researcher at the University of St. Gallen while conducting this analysis of the Self-Googleing Phenomenon. In relation to Internet studies, he has recently worked with Lars Kirchhoff and Axel Bruns on monitoring the Australian political blogosphere. Building on the work presented in this paper he recently founded the sociomantic labs based in Berlin.

E-mail: thomas [dot] nicolai [at] sociomantic [dot] com

Lars Kirchhoff was Researcher at the University of St. Gallen while conducting this analysis of the Self-Googleing Phenomenon. Among his research interest are new media, social networks and their impact on information retrieval. Building on the work presented in this paper he recently founded the sociomantic labs based in Berlin together with Thomas Nicolai.

E-mail: lars [dot] kirchhoff [at] sociomantic [dot] com

Axel Bruns is a Senior Lecturer in the Creative Industries Faculty at Queensland University of Technology in Brisbane, Australia, and a Chief Investigator in the ARC Centre of Excellence for Creative Industries and Innovation (CCI). He is the author of *Blogs, Wikipedia, Second Life and Beyond: From Production to Prodsusage* (2008) and *Gatewatching: Collaborative Online News Production* (2005), and the editor of *Uses of Blogs* with Joanne Jacobs (2006; all released by Peter Lang, New York). He blogs at *snurb.info* and *Prodsusage.org* and contributes to the *Gatewatching.org* group blog with Jason Wilson and Barry Saunders.

E-mail: a [dot] bruns [at] qut [dot] edu [dot] au

REFERENCES

- Alford, C. F. (1987). "Eros and Civilization" after Thirty Years: A Reconsideration in Light of Recent Theories of Narcissism. *Theory and Society*, 16(6), 869-890.
- Arruda, W., & Dixson, K. (2007). "Build Your Brand in Bits and Bytes". ChangeThis, 40.04, from <http://www.changethis.com/40.04.BuildYourBrand>
- Campbell, W. K. (2001). "Is Narcissism Really So Bad"? *Psychological Inquiry*, 12(4), 214-216.
- Contrada, J. D. (2004). "UB communication professor calls 'self-Googling' shrewd form of personal brand management". *University of Buffalo Reporter*, 35(29).
- Emmons, R. A. (1987). "Narcissism: Theory and measurement". *Journal of Personality and Social Psychology*, 52(1), 11-17.
- Glasner, J. (2001). "Your Ego Just Took a Blow". *Wired*, 4.
- Halavais, A. (2007, 02.03.2007). "Rising narcissism among college students". Retrieved 16.09.2007, 2007, from <http://alex.halavais.net/rising-narcissism-among-college-students/>
- Jansen, B. J., & Spink, A. (2006). "How are we searching the World Wide Web? A comparison of nine search engine transaction logs". *Information Processing and Management*, 42(1), 248–263.
- Katz, L. G. (1993). *Distinctions between Self-Esteem and Narcissism: Implications for Practice*. Urbana, IL: ERIC Clearinghouse on Elementary and Early Childhood Education.
- Kirchhoff, L., Stanoevska-Slabeva, K., Nicolai, T., & Fleck, M. (2008). *Using social network analysis to enhance information retrieval systems*. Paper presented at the Applications of Social Network Analysis (ASNA), Zurich.
- Kopelman, L. S. M. R. E. (1984). "The Best Seller as an Indicator of Societal Narcissism: Is There a Trend"? *Public Opinion Quarterly*, 48(4), 720-730.
- Lampel, J., & Bhalla, A. (2007). "The Role of Status Seeking in Online Communities: Giving the Gift of Experience". *Journal of Computer-Mediated Communication*, 12(2).
- Lasch, C. (1991). *Culture of Narcissism. American Life in an Age of Diminishing Expectations: American Life in an Age of Diminishing Expectations*. New York, N.Y.: W. W. Norton & Company, Inc.
- Mullins, L. S., & Kopelman, R. E. (1984). "The Best Seller as an Indicator of Societal Narcissism: Is There a Trend"? *Public Opinion Quarterly*, 48(4), 720-730.
- Nelson, M. C. (Ed.). (1977). *The Narcissistic Condition: A Fact of Our Lives and Times*. New York, N.Y.: Human Sciences Press.
- Oliver, J. J. (1993). *Decision Graphs An Extension of Decision Trees*. Paper presented at the Fourth International Workshop on Artificial Intelligence and Statistics, Clayton, Victoria, Australia.
- Raskin, R., & Hall, C. S. (1981). "The Narcissistic Personality Inventory: Alternative Form Reliability and Further Evidence of Construct Validity". *Journal of Personality Assessment*, 45(2), 159-162.
- Raskin, R., & Terry, H. (1988). "A Principal-Components Analysis of the Narcissistic Personality Inventory and Further Evidence of Its Construct Validity". *Journal of Personality and Social Psychology*, 54(5), 890-902.

Raskin, R. N., & Hall, C. S. (1979). "A narcissistic personality inventory". *Psychological Reports*, 45(2), 590.

Spink, A., Jansen, B. J., & Pedersen, J. (2004). "Searching for people on Web search engines". *Journal of Documentation*, 60(3), 266-278.