

Google Yourself!

Measuring 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

INTRODUCTION

Searching for information about oneself has become fairly popular with the rise of modern search engines and the social 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 led to an increasing 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 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 & Dixon, 2007). Therefore end-users have increased their efforts in actively monitoring and shaping their information online.

However, only few investigations explore this phenomenon and the underlying motivation behind such behaviour. One approach to explain and understand this practice of monitoring the web often, which is described as *self-googling*, can be the theory of narcissism, which may help to 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. In addition, however, this practice can also be seen as a way of observing the social construction of personal reputation, and a means of managing the personal 'brand' through self-marketing (Lampel & Bhalla, 2007).

By investigating *self-googling* through the monitoring of search engine activities of users this paper adds to the few quantitative studies on this topic currently in existence. The paper explores this phenomenon by answering the following questions: (1) To what extent is the *self-googling* visible in the usage of search engines? (2) Is any significant difference measurable between queries related to *self-googling*, and generic search queries? (3) To what extent do *self-googling* search requests match the selected personalised web pages?

To address these questions the next section provides an introduction to the theory of narcissism in order to define *self-googling*. Subsequently, the paper presents the research methodology and the results from a 14-month online experiment using Google search engine usage data. Finally the results are presented and discussed.

THEORY OF NARCISSISM

Narcissism is generally perceived as a growing social phenomenon in western society, as Lasch puts in his widely cited book *Culture of Narcissism* (1991). Recent technology shifts have yet again started a discussion about the increase of narcissistic behaviour in western culture (Kopelman, 1984; Mullins & Kopelman, 1984; Nelson, 1977). Among these technology shifts are 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 also be seen as a functional and healthy personal strategy for making sense of our increasingly fast and techno-oriented world. According to this argument, narcissism can be considered as a cultural and social entity (Emmons, 1987; Mullins & Kopelman, 1984).

Defining Narcissism

Having its seeds in the ancient Greek mythology the term narcissism is usually perceived as a personality disorder in the 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 behaviour 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. Furthermore it can be observed that people refuse jointly defined social objectives. Campbell (2001) has criticized this perception of narcissism as a dynamic construct with negative connotations for an individual. He rather argues to interpret narcissism as a positive phenomenon where we need to separate between an abnormal and normal narcissistic behaviour.

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

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 narcissic

behaviour is a highly common personality characteristic found in all modern societies. Therefore narcissism is not necessarily perceived as a personality disorder. Rather it is a healthy strategy to deal with the raising complexity in everyday lives through social media.

The Narcissistic Personality Inventory

The most specific approach to understand narcissism is proposed by Raskin et al. (1981; 1988; 1979) and their efforts in developing 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 in our society. The research methodology is based on the acceptance and definition of narcissism as a self-focused concentration on one's own behaviour and image as defined above. In their most current research project in 2006 they examined 16,475 college students using the *Narcissistic Personality Inventory* methodology. While comparing the findings with the results from 1982 they found that two-thirds of the students were showing a score above average. Furthermore they found evidence that narcissistic behaviour has increased by more than 30 percent between 1982 and 2006, if we follow their definition.

Summarizing these consolidated findings and definitions of narcissism, we 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 behaviour measured by Raskin et al. may be explained in part by the ongoing evolution of social media. This evolution has been shifting attention to people usually outside of the scope of traditional media. Consequentially, people are starting to experience and learn how to deal with this growing social media attention on single individuals. Thus, 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 analyse the search behaviour 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 et al. 2006; Spink et al. 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 analysed by Jansen et al. (2006), showing a growth of search terms in that category from 21.5% in 2001 to 41.5% in 2002.

To investigate the *self-googling* phenomenon we have gathered and analysed 2.46 million search engine requests over a time period of 14 months. On the basis of these 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 *personalised information resources* and *non-personalised information resources*. *Personalised information resources (I)* are web pages which specifically refer to information about individual persons, and whose URIs include a version of the person's name. *Non-personalised information resources (II)* refer to pages which do not contain information about a specific person, and whose URIs do not contain any name-specific elements, but are instead generic in nature. The detailed construction of the two different URI schemas is presented below (figure 1).

	Protocol	Domain	Path
I	http://	domainname.com/	sergey_brin
II	http://	dommainname.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 analyse 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. This way a web server can detect where a user comes from 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 on our server, which he accessed on 15 September 2006 under /sergey_brin on our web server. This particular web page provides information about Google founder Sergey Brin. The referrer information in this log file entry can be seen on the second line.

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

It contains the full URI of the web page where the user found a link to the page on our web server. 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 URI above, 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 (anonymised in this paper), and some information about their browser and operating system as well as an access timestamp.

ANALYSIS

To investigate the *self-googling* behaviour we analysed 2.46 million of these search engine referrals to our web servers. Data analysis was limited 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 1 percent. Analysis started in August 2006 and ran for 14 months, to October 2007. In a first step, we identified and extracted all Google referrals from more than 100 million log file entries which contained 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 personalised and non-personalised search queries over the 14-month time period of the experiment. Second, we determined the absolute growth in search requests for both categories (personalised and generic) over the whole period. Finally, we selected those log file entries which contained a personalised web page and a referrer and assigned a score that quantifies the match between the search query and the web page that was accessed on our web server. This analysis was motivated by the question to what extent a search request for personalized web pages would match the search query and therefore the name of the person.

RESULTS

Our presentation of the research results is divided into three sections, addressing 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 personalised web pages. In the first 11 months of our study we

saw a steady growth in retrieved personalised web pages. At its peak, users selected 14,000 personalised web pages each day. This number declined to 8,000 search requests by August 2007. The trend for non-personalised web pages looks similar for the first months. However, access to non-personalised web pages (figure 4) shows a different trend for the later months of our experiment, peaking at 6,000 search requests per day only. Further, we can see a regular decline in search requests for personalised and non-personalised web pages on weekends. Additionally, there is a difference in the total amount of personalised and non-personalised web pages which were accessed: personalised web pages were accessed more frequently, by an average factor of 2.

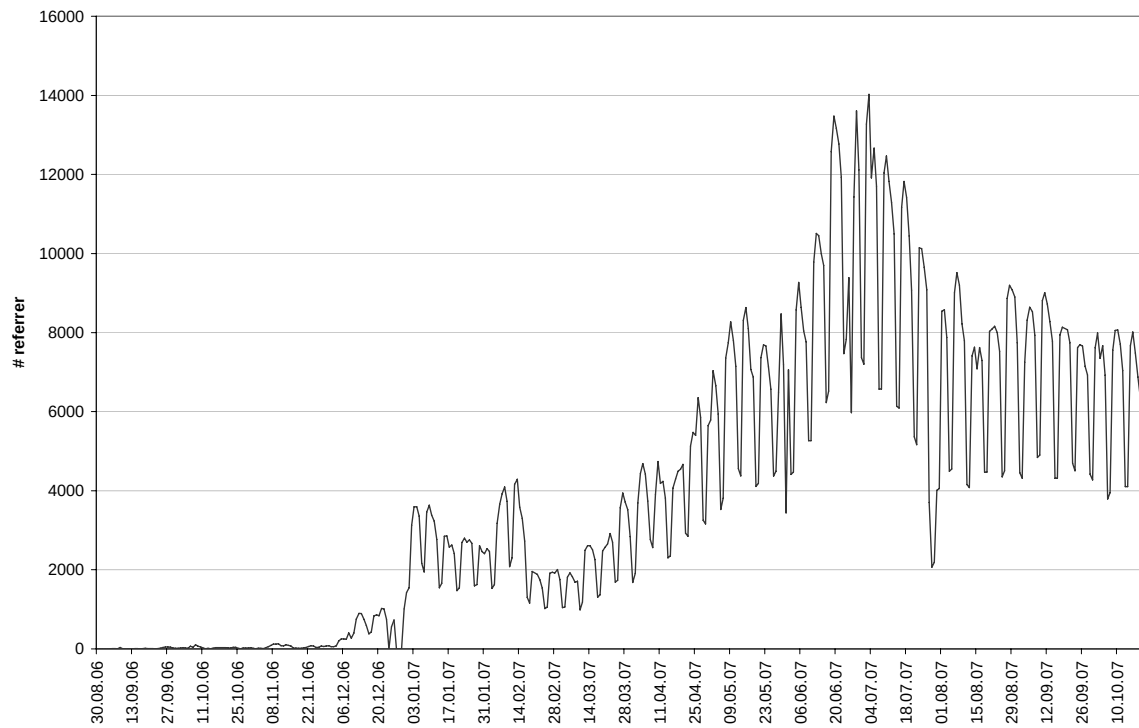


Figure 3: Access to personalised web pages following a referral from Google

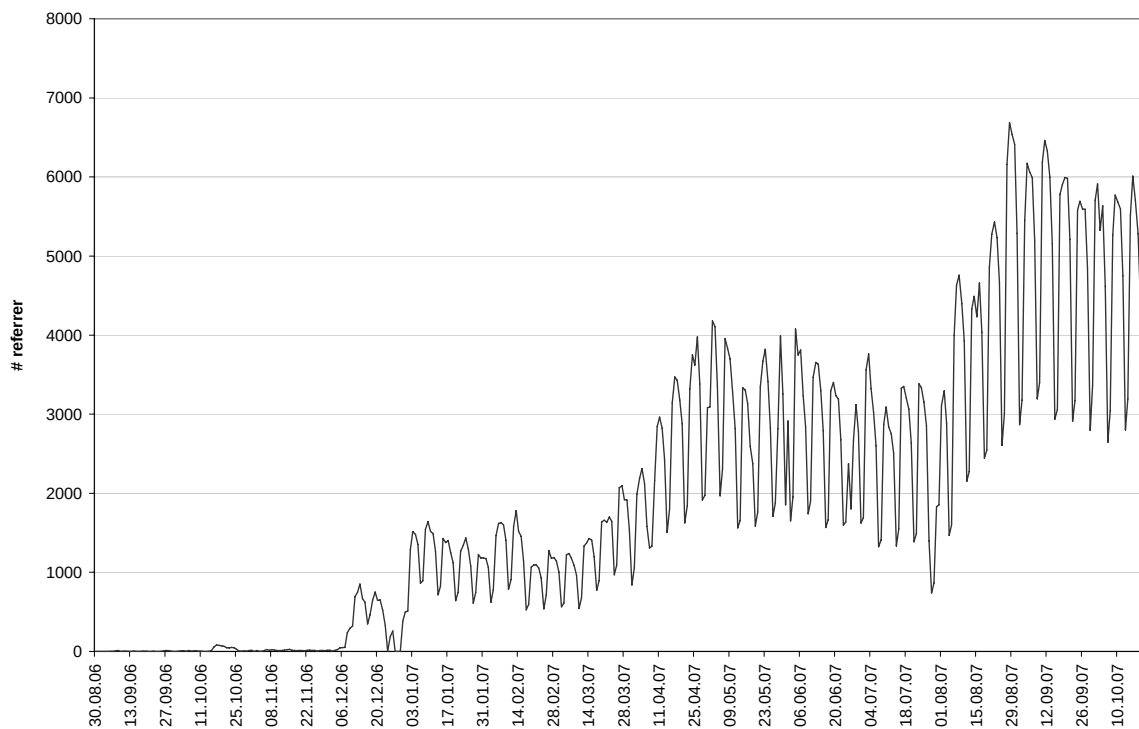


Figure 4: Access to generic web pages following a referral from Google

For a comparison of the number of search engine requests for personalised and non-personalised (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 numbers of web page request over time, and clearly demonstrates a difference in growth over the time period of our experiment.

Demand for personalised web pages grew more strongly than demand for non-personalised web pages.

Over the lifetime of the study, we measured 1.66 million search requests for personalised web pages and 0.8 million search requests for non-personalised web pages – the number of requests for personalised web pages is double the number for generic web pages. Furthermore, the 1.66 million search requests for personalised web pages are distributed across 0.8 million different web pages. Hence, on average every personalised web page was requested 2.07 times. Non-personalised 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.

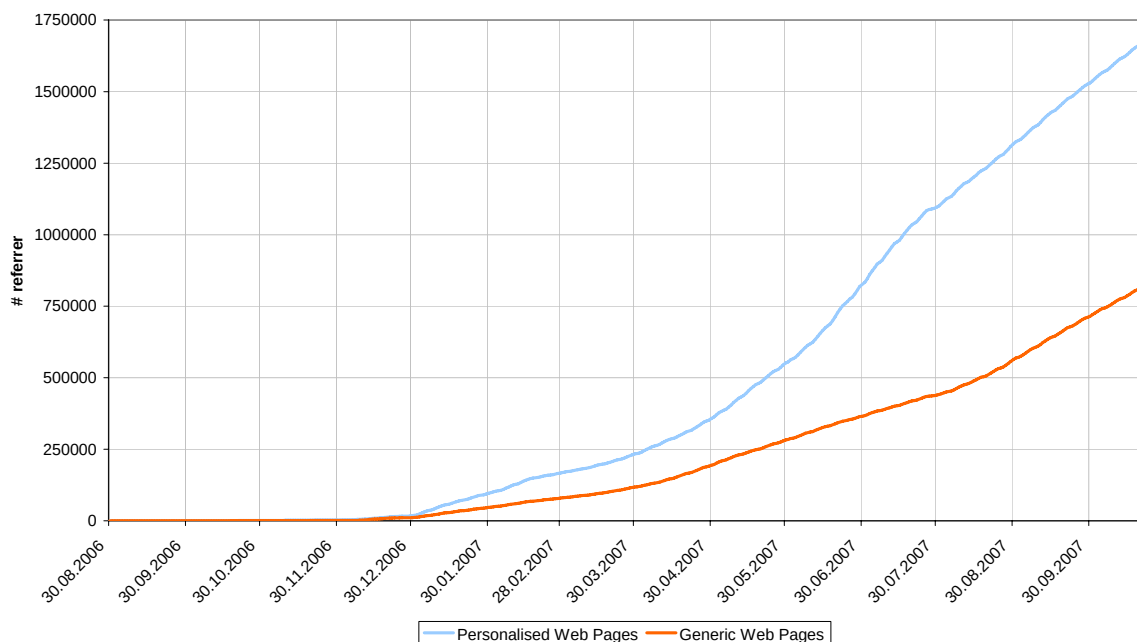


Figure 5: Growth of personalised and non-personalised search queries

Finally, we investigate the quality of correspondence between the personalised search queries and the hyperlinks selected from the result lists. We analysed 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 analysed the 1.66 million search engine referrals related to personalised 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 analysed 1.66 million referrer show a matching score of 100 percent meaning that search engine users entered the name of the person they were searching for exactly and without errors. Furthermore, there is a second significant characteristic at a matching score

between 95 and 92 percent. However, between 99 and 98 percent the data shows an anomaly in the distribution. Potentially, this may be able to be explained as an effect of Oliver’s matching algorithm itself, and of its dependence on the length of the strings being compared. A different, and perhaps more interesting explanation would point to a ‘self-correcting’ mechanism in human cognition that prevents us from misremembering a person’s name by a very small margin, and 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% 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.

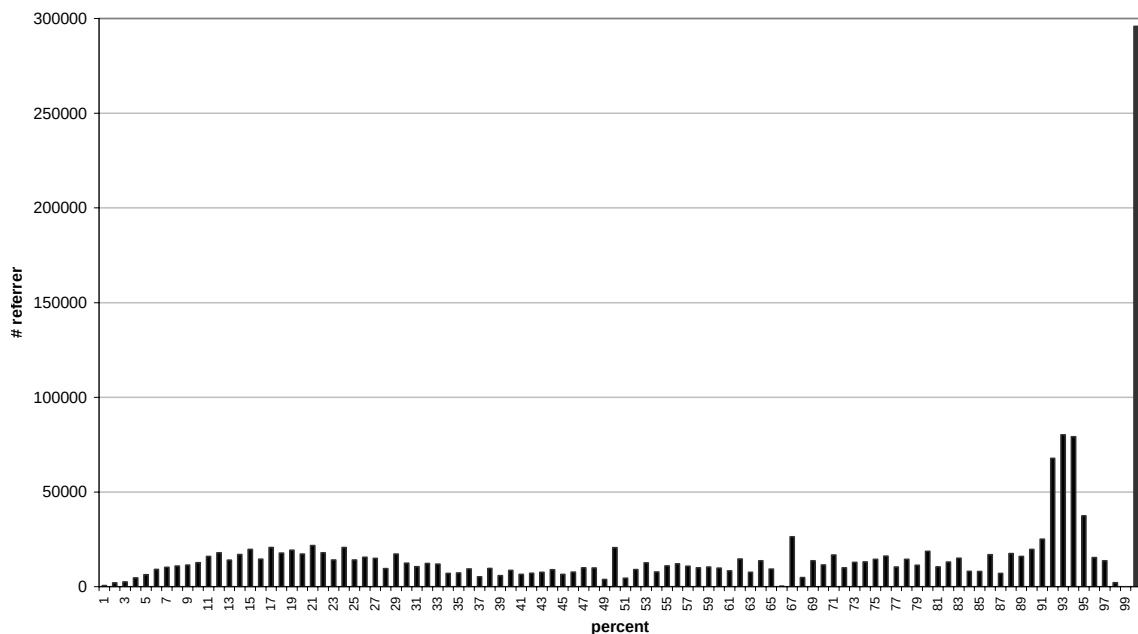


Figure 6: Similarity between search terms and selected personalised web page URIs

DISCUSSION

To investigate the *self-googling* phenomenon, we have extracted and analysed 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 personalised information resources in comparison to generic web pages. Furthermore, we analysed the quality of matches between personalised 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 personalised web pages. Thus, we see a very high quality of interaction with personalised web pages through search engines. Based on our experiment we are

questioning that the average of everyday search queries provides an equivalent quality score of 64.74 percent as presented above.

The results from our experiment indicate that users show a higher interest in personalised web pages. The high degree of correlation between search query and selected hyperlink (URI) for personalised web pages, in contrast to access to non-personalised 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 themselves by actively monitoring and shaping their persona and perception online**, we conclude that there is a growing interest in personalised web pages. The significant difference in interest in personalised as against generic web pages might explain the overwhelming success of social media applications on the WWW.

Finally 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 to some extent (Kirchhoff, Stanoevska-Slabeva, Nicolai, & Fleck, 2008).

Limitations

While measuring the performance of personalised and non-personalised 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 personalised web resources. Additionally, the scope of our research project may be too limited for generalisation, 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. Summarising our research results, we might conclude that the rise of narcissism as a cultural entity in society 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.

Our approach to investigating the *self-googling* phenomenon by measuring the level of access to personalised and generic web pages from the search engine result lists has shown some significant differences between the two page types. The performance of the two types shows that personalised web pages are twice as popular as generic web pages. Even when both types show up in the search engine results lists there may be some user bias in favour of personalized web pages which could be subject to future research.

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.
- Campbell, W. K. (2001). Is Narcissism Really So Bad? *Psychological Inquiry*, 12(4), 214-216.
- Contrada, J. D. (2004). UB communication professor calls "self-Googleing" 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/>
- 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.