Searching for Myself:
Motivations and Strategies for Self-Search

Catherine C. Marshall
Microsoft Research
1065 La Avenida, Mountain View, CA, 94093, US
cathymar@microsoft.com

Siân Lindley
Microsoft Research
21 Station Road, Cambridge, CB1 2FB, UK
sianl@microsoft.com

ABSTRACT

We present findings from a qualitative study of self-search, also known as ego or vanity search. In the context of a broader study about personal online content, participants were asked to search for themselves using their own computers and the browsers and queries they would normally adopt. Our analysis highlights five motivations for self-search: as a form of identity management; to discover reactions to and reuse of user-generated media; to re-find personal content; as a form of entertainment; and to reveal lost or forgotten content. Strategies vary according to motivation, and may differ markedly from typical information-seeking, with users looking deep into the results and using image search to identify content about themselves. We argue that two dimensions underpin ways of improving self-search: controllability and expectedness, and discuss what these dimensions imply for design.

Author Keywords

Self-search; ego search; vanity search; autosurveillance; identity; aggregation; doppelganger; archive.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

As social networking sites evolve and personal content is increasingly hosted online, search engines are a central way users seek online content about themselves. In 2004, a Harris poll revealed that 40% of adults in the US had performed an ego or vanity search; in 2009, a Pew Internet poll found that the majority of adults (57%) had searched for their own name [16]; and in 2008, to support their work on query log privacy, Jones and her colleagues used Yahoo's query logs to establish that as many as 30% of

users conducted a vanity search over the 70-day period of their study [13]. *Thus, people do conduct vanity searches, and some may conduct them regularly*.

When people search for themselves this way, they may encounter content they have generated themselves, but they may also find public records, stories from local newspapers, items from commercial data repositories, and unfamiliar material about themselves. This diverse material is brought together as a search engine results page (SERP), which may act as a form of dynamic profile or self-presentation [17]. If favorable, the SERP can serve as a résumé; anecdotal reports point to the success of having a business card that exhorts the recipient to 'google me', rather than providing a static URL [25]. Indeed, at one time Google issued business cards to support the phenomenon [22]. However, the SERP may also present links to a variety of confounding sites: those about doppelgangers, pages assembled by aggregators and data miners, and social media content the search's subject hoped would remain hidden.

In this paper, we report findings from a qualitative study that begins to investigate this type of search, which we will refer to as *self-search,* focusing specifically on participants' interaction with the SERP. Our research aims to understand the motivations and strategies behind self-search, as well as the obstacles searchers face and surprises they encounter. Building on these results, we consider what our analysis implies for the design of search engine interaction.

Thus far, self-search has been the province of personal branding books, a sideline of the digital identity literature, and an implicit aspect of named entity information retrieval. In this paper, we shift self-search and the SERP from the periphery to the research focus. We begin with a brief discussion of related work before turning to a description of the overarching study, focusing on the part of data collection and analysis that is relevant to self-search. After we present the self-search results, we explore their implications: how can self-search be rendered more effective? How can we better support people's ultimate aims in seeking this type of personal information?

RELATED WORK

Research on self-search lies at the intersection of digital identity management, information-seeking, personal information management (PIM), and privacy; we will draw

on all four areas to inform the interpretation of our results. Self-search (also known as vanity search, ego search, and autosurveillance [7]) has also captured the popular imagination; thus we will touch on recent privacy and identity management books aimed at non-researchers.

Where self-search has been examined, it is often in the context of improving search results or protecting privacy. In the case of the former, researchers have sought to predict whether a query is a name [2] or have explored self-search in the context of rare queries [26]. In the latter case, researchers have pointed to the risks of revealing identity through self-search, highlighting the implications for privacy and data surveillance [13,23]. Other researchers have examined self-search in the context of narcissism (albeit healthy narcissism), situating it in the development of digital identity [20]. Their aim was twofold: to develop a method for distinguishing between self-searches (as part of a broader phenomenon of people searches) and other types of search engine-based information seeking, and to characterize the phenomenon's relative prevalence in overall search engine use. A technique of the sort these researchers describe is likely to be important in providing computational support for self-search.

Self-search sometimes falls under the operational rubric of PIM: people may search for themselves as an implicit part of managing their outward-facing personal information [14]. The key distinction is that we are focusing on how people encounter themselves and their footprints in the process of self-search rather than in common PIM tools. By looking in some detail at how people negotiate self-search, we hope to provide a new perspective on managing the visible traces of one's own online identity. Such ideas resonate with Marchionini's notion of proflection [17], the construction of a digital presence in online resources via explicit action and monitoring implicit online activities.

While research on self-search is fairly sparse, it features prominently in non-scholarly guides to digital identity management, where it is highlighted as a key activity in 'personal branding' [1,7] and 'protecting your internet identity' [4]. People are increasingly worried about "*how much sensitive information about ourselves we voluntarily consign to others in exchange for social interaction, a discount, or simply to access a product or service*" [4, p. viii]. Self-search is a diagnostic tool that can be used to debug a digital identity, to answer the question of what's out there, possibly out of most people's control [10]. These guides distinguish self-search from traditional information seeking in two important ways: First they emphasize that self-search is aimed at uncovering one's digital footprint, the impression made as results are combined and taken as a whole (as opposed to information seeking, which is aimed at a per-document notion of relevance). Secondly, they highlight the distinction that the results presentation (the SERP) is directed at an imagined audience, rather than at the original searcher. Taken together, these two emphases make this aspect of self-search unique.

RESEARCH AIMS

The findings reported in this paper are from a study that aimed to understand how people conceptualize online content about themselves, be this through collections that they have curated (e.g. on sites such as Flickr and Pinterest), traces they have left (e.g. on social network sites), or content created by others that is about them. We were interested in people's awareness of the breadth of this online content, how they kept track of it, and whether they made any efforts to back it up or archive it. A key part of this research relates to self-search. Findings from this study that relate to archiving and managing online content are covered in [15]; in this paper, unpublished findings that focus on self-search are presented. We aim to understand how people search for themselves, what they expect to find, where surprises lie (both pleasant and unpleasant), and what they would like to do with the results. We then consider what our analysis implies for the design of search engines.

METHOD

We interviewed 14 people, 8 in the UK and 6 in the US. To reach participants with a spectrum of external web footprints, we recruited two types of people: those whom we expected to have a significant online presence, including creative output (e.g. amateur musicians), and those with normal social networking habits (e.g. regular Facebook users). Two digital natives (young adults) were among our recruits, since we are aware they might have distinctive online practices and relationships with their 'virtual possessions' [21]. Participants were given a £40/$65 gift voucher in appreciation of their time. Table 1 (included from [15]) summarizes participants' characteristics. The same pseudonyms are used in both papers.

Participant	Age	Gender	Based	Details
Harry	35	M	UK	Amateur musician, consultant
Charlie	36	M	UK	Amateur musician, academic
Jacob	27	M	UK	Amateur musician, sound designer
Oliver	48	M	UK	Amateur photographer, academic
Ava	38	F	UK	Blogger/tweeter, e-learning professional
Thomas	32	M	UK	Blogger/tweeter, lobbyist
Jane	32	F	UK	Stay-at-home mum
Sophie	18	F	UK	Student embarking on a gap year
Ann	50	F	US	Tech writer, amateur musician
Kim	50	F	US	Novelist, blogger and stay-at-home mom
Lynn	37	F	US	Admin, videographer
Todd	33	M	US	Journalist, fact-checker
Mary	31	F	US	Non-profit administrator
Vincent	20	M	US	Student, math nerd

Table 1. Participants' pseudonyms and backgrounds.

The interviews were semi-structured and open-ended, and lasted between 90 minutes and 3 hours. They consisted of

three segments: an online tour, a self-search exercise, and reactions to a set of design sketches for an integrated personal archive. The self-search exercise was motivated by our desire to turn up content the participant had forgotten, or was unaware of. However this portion of the interview often became multi-faceted, as the participants encountered unexpected content, and considered how they appeared to others through the lens of the SERP. Self-search data also appeared in other interview segments, primarily during online tours, e.g. when participants forgot where they had stored something, and needed to search for it explicitly.

To initiate a discussion of self-search, we asked participants if they had ever searched for themselves on the Web; all had done so for a variety of reasons. We then asked them to search for content about or by themselves as they normally would; they were able to choose the search engine, the starting point (e.g. the whole Web or a social media site like Facebook), and use input terms that seemed appropriate to them. The searches were conducted on participants' own computers, so normal personalization may have occurred. We urged them to discuss the value and appropriateness of individual results, usually through multiple results pages.

The interviews were audio-recorded and transcribed for analysis. We also collected screen snaps and photos to document portions of the interviews. To re-analyze the data, we extracted portions of the transcripts concerned with self-search, and used thematic analysis to code them and identify common themes and underlying patterns.

FINDINGS

Although the interview provoked the self-searches that we observed, our discussions with participants highlighted various circumstances in which the activity would normally occur. These ranged from self-search as a relatively common online housekeeping task, to a focused effort in preparation for an event like a job search. Self-search also fell under the rubric of more traditional information seeking tasks, like re-finding familiar content, such as a blog post the participant wrote, or discovering new material, such as commenters' reactions to that blog post.

We do not wish to make any strong claims about the regularity or frequency of self-search on the basis of our qualitative findings. Indeed, it is quite possible that participants may have forgotten specific instances of self-search, especially if it was embedded in a broader task (such as re-finding). Furthermore, since self-search is apt to carry with it a stigma of narcissism, some participants were quick to assert that for them, it was done infrequently. However, none of our participants claimed they had never tried a self-search, most seemed fairly familiar with the results, including content that was not about themselves (such as content related to doppelgangers), and some admitted to conducting regular self-searches. So while Ann said she'd last searched for herself "*years ago*", Mary said she did it "*every couple of years… to make sure there's*

nothing out there that I don't know about", and Lynn who was an avid Facebook user, said "*I hate having stuff about me online*". She performed self-searches regularly.

This brief overview points to a number of different motivations for conducting self-searches. In the following, we consider in more detail what these motivations are. We also relate these to self-search strategies, since how people look for themselves online and interact with the resulting SERP is contingent on why they are doing it. Finally, we consider how searchers' motivations shape what they expect to find, and what types of results they consequently perceive as surprising. We sum up these findings in a framework that juxtaposes familiarity of the material with how and whether it can be controlled.

Self-Search Motivations

What provokes people to search for themselves? In the following, we tease apart five motivations that emerged in our analysis. We focus on the first (identity management) in particular, as this practice is associated with strategies that are distinct from other information-seeking practices; in this case the SERP may function as a form of dynamic profile. Other motivations we discuss more briefly include the need to re-find familiar content or discover new content; these are familiar tropes in the information retrieval literature (e.g. [6, 12]). We also consider self-search in the name of entertainment, and as a way of curating archival material.

Self-search in pursuit of identity management
As past research has pointed out, managing one's online presence and multiple personas is a significant and consuming endeavor [9]. Self-search may act as quality control on proflection; participants use self-search to check how well they are curating their online identities, and whether intervention is necessary. Are various personas adequately represented? Are they reasonably distinct? If a unified presentation of self is desired, is this integrated online presence available via the SERP? Finally, if circumstances have changed, is one's SERP up-to-date?

While some proflective self-searches are performed with online identity itself as the focus, others are performed prophylactically: participants place themselves in the role of a specific searcher—a prospective employer or an ex-boyfriend or girlfriend—to speculate on what that person will find. This way a person can potentially avoid exposing content they don't want the prospective searcher to see, or they can identify unwanted porosity between distinct and purposefully separate portions of an online identity.

Two aspects of proflective self-search emerged in our data: (1) managing the appearance of the SERP as a whole; and (2) identifying specific pieces of content that the participant wanted to weed out. While these aspects are inter-related, they were associated with different strategies and suggest different implications. We deal with each separately, first discussing motivations, then covering associated strategies.

Managing the appearance of the SERP as a whole. As they searched for themselves, participants reacted to the SERP as an artifact in its own right. Specifically, they seemed concerned with the relative rankings of different sites, content repetition and its effect, and doppelgangers. While these elements may be difficult to control, participants had adopted strategies to influence their SERP, especially with an eye to particular audiences.

The aspect of self-search that changed most frequently, and was often the source of some mild surprise, was the order in which the SERP presented results and the auto-complete terms the search engine suggested. Todd, a journalist, noted that although nothing was missing that he expected to see on his semi-annual self-search *("it's always been the same stuff")*, there was nevertheless *"more stuff that I would rather see higher up, like that story from China was a story that I enjoyed doing. ... It's a good example of something in my portfolio that I wish was higher."* He went on to contrast the article he had written with other genres of material he was encountering on his SERP: *"So then when I see this weird Quora user page so prominent, or that MyLife thing, it just seems like something is not working quite correctly there. Because it's not relevant to me. And if you clicked on it, and were curious about me, I don't know how relevant it would be to somebody who wanted to find out something about me."* Similarly, Charlie (an amateur musician) noted that his band's name did not appear as an auto-complete, saying *"that's sad."* Like SERP position, the presence or absence of auto-complete terms was a part of participants' search engine-based identity.

A second, related, feature of the SERPs was the presence of doppelgangers. While search engine features such as Google Profiles have been introduced to support people in distinguishing themselves from their doppelgangers in search engine results [8], our analysis revealed that users have a more complex relationship with their doppelgangers than simply wishing to be separate from them. Most participants were aware of who their doppelgangers were and some were playfully competitive with them. So in addition to showing an interest in the ranking of their own content, they also pitted themselves against bartenders, homeopaths, writers and musicians who shared their names. Indeed, some participants adopted strategies to underpin this competitive element of self-search; Oliver had never searched for his full name, only entering his rather distinctive surname, which allowed him to *"compete with a curry"*, as well as with his brother, to add interest to his self-search results. Doppelgangers are also essential in obscuring one's own activities. We discuss this later, as a strategy of sorts for maintaining anonymity. By contrast, doppelgangers can also stand in as a proxy for one's own presence. Jane, who did not appear on the SERP for her own name, noted *"sometimes it's quite nice to see that there's successful Jane Taylors out there doing well, I think, ah that must be a good name, a successful name."*

Finally, participants noted that novel content can be lost or devalued through repetition (in spite of search engines' use of sophisticated duplicate detection algorithms). Ann, who like several of our other participants, had an academic persona and an artistic one, commented on the degree to which the same content recurs: *"You publish a few things, and it's just one giant echo chamber... you have to try really hard to not have a million things on the Web. Because you put one thing, and it gets copied."*

Strategies for managing online identity. Clearly, it is difficult to control the ranking of search results, and even more so to influence the presence of doppelgangers in a SERP. However, participants did adopt various strategies in an effort to influence the SERP as a whole. These included adopting online identities that could be distinguished from prominent doppelgangers, using pseudonyms to create distinct online identities, and even embracing sites that were likely to appear high in the rankings. We consider these in turn.

Some participants specifically tailored online content that they were in control of with search in mind. Often this meant specifically curating online identities. For example, Kim had adapted her online identity to distinguish herself from a well-known actress by adding her middle name. Searching for her full name would then pull up her *"writing persona"*, which meant her Amazon author's page would be included prominently in the results. Where participants wished to separate out different elements of their online activities, multiple identities were often adopted. Charlie pointed out the absence of any of his musical efforts in an initial search for his name: *"I don't have any music associated with my name. I have a different name for that, my artist's name or whatever"*, and Ann used several different pseudonyms to control which sites would appear on the SERP, using *"BitBucket for the nerd persona"* and SoundCloud for her musician persona. Online identities were often managed with a specific audience in mind (e.g. prospective employers). Thus, Charlie described self-search as *"having a little look in the mirror in terms of what a potential employer's going to see, so I want to be aware of, ok if they search my name they're going to see this this this and this"*. For him, separating out his musical persona was key to managing his professional identity.

Interestingly, for some participants anonymity trumped any desire for a SERP-based online identity. Lynn said: *"I am thankful for [a woman whose name is spelled similarly], who is a jewelry designer, so when you search my name, she [comes up]. There's like three pages of [links to her pages]. I'm going, 'thank god this woman exists.'"* Her concern with anonymity was in large part motivated by the dissonance between a previous identity in the entertainment industry and her current professional identity in the corporate world: *"I make sure I don't have an online presence. Especially when some things that I work on can be misconstrued as offensive. I worked on a film called*

[provocative name redacted]. So if I'm job hunting and people look me up, I don't want them to say, 'oh this woman's worked in porn. Let's hire her.'" She later clarified, *"When you're gay, you don't know. Because I always look at it from a perspective of, who's going to search for me? People who want to hire me, or work with me, or something. So I try and keep it clear. ... Because I can't control it, and I'm reliant on somebody else's belief system."*

Some participants saw embracing social media as a way of having some control over the way a SERP appears. Because certain social media sites were apt to appear high in the rankings, using them to curate a presence meant that a self-search would produce content that was under their control, and would represent them in a satisfactory way. Ava commented, *"If you use social media in the right way then if someone searches for you, you can make sure that the results that they find are about you and not about someone else, and therefore you are in control of it"*. By making public the profiles she was happy for her professional peers to see (such as Twitter), while hiding others (such as Facebook) from search results, she felt in control of the image she was presenting via the SERP related to her name.

Similarly, Thomas invested heavily in his own website, preferring to put content there as a way of participating in debates about local issues, rather than on other forums or blogs. This was partially about retaining control of the content, and partially about building up his own "*asset*" rather than adding value to someone else's. *"I'll write it on my site and link to it and that's mainly about having control over it and getting the search results, so if I've written something interesting then people may as well come and read it on my site."* His view of the SERP was *"if I didn't have my own website then it wouldn't look like that would it? So there is definitely a reason for having a website of your own [... and] another reason for having a Twitter feed."*

As a final strategy in managing a SERP-as-profile, some participants sought to avoid personalized results, since they felt they would be misled about what others would see. Although several participants were aware of personalization and disliked it, they were uncertain how to turn it off. Todd said, *"I don't know whether it's set right now. I think it might be the default..."* and Kim said, *"But Google's strange, because it will pull up things you've already pulled before. ... I don't know what normal people would see."*

Weeding out unwanted personal content. Presenting a coherent identity—either as a unified person or as an activity-centered persona—via a SERP is difficult. Participants relied on certain sites being visible only to certain audiences as a way of preserving their privacy in an online (and very public) space. Making sure there were no rogue elements within a SERP entailed weeding out unwanted content, and this meant a robust approach to search. Unlike traditional information-seeking where the

SERP is abandoned if the desired content isn't found on the first few pages, participants went deep into the search results and often repeated searches using different modifiers, as a way of doing a form of what Charlie termed "*auditing*". Important types of unwanted content included social network sites that unexpectedly showed up in search results, pages that aggregators constructed about participants from multiple sources, and references to past careers and deeds that participants would prefer to forget, particularly in the form of photos.

Even those participants who used social network sites carefully and with a view to enhancing their SERP were sometimes surprised by what search engines presented to them. Ava, who noted that controlling social media means a degree of control over the SERP itself, was surprised to find Pinterest near the top of her SERP. She described it as "*completely different from anything else that I do online … I don't even know if I would really like to engage either my friends or my professional contacts, because it is just really housewifey.*" She left the website open as a reminder to herself to change the privacy settings once the interview was finished. Other boundaries between distinct online identities also revealed themselves to be too porous for comfort. Ann found her "real" and "hacker" personas mixed "*because I 'liked' the hacker dojo*" on Facebook, and Ava discovered an aggregation site that seemed to draw content from pages that she had 'liked' on Facebook (which was itself hidden from the SERP). This left her wondering if she should 'unlike' the pages within Facebook, so as to prevent this apparent information leak. Aggregation sites, which compile and mix results from different sources (including among others, telephone book listings, property records, court records, blogs, and social media profiles and status postings) were notable in the way they surprised and often disturbed the participants. They were perceived as interfering with the construction of multiple coherent identities, since they drew together content that was normally kept distinct [3], and as compromising privacy, since they dip into public data that was normally obscure or difficult to obtain, and so violate participants' expectations of privacy through obscurity [11,27].

Participants frequently remarked on how many aggregators they had seen, the mysterious sources of their information, and whether they would be apt to use one of them. A few participants had ignored the aggregator sites in the past, and so the interview was their first real encounter with their content. For Kim, it was PeekYou: "*Oh, and it's me, isn't it? What is it? I guess I could click on it. I've never seen it.*" Todd, who was fairly tolerant of the aggregators' information practices, explained his ambivalence: "*You'll see these kind of weird companies that seem to suck up information.... Actually I'm curious. 'Create a free account to see all 2 results.' Or sign in with one of my social network links. Facebook. Twitter. I have no desire to do that. ... location [redacted]. That's where I went to college. So I haven't lived there in almost 10 years. They just pulled*

some stuff from my website it looks like. Or maybe not from my website. ... So this doesn't make me want to give them my information and sign up as a useful resource ... People must."

Paradoxically, participants felt that the sites' inaccuracies and out-of-date information were also a problem. For example, Charlie referred to "*ancient*" information in Zoominfo and Ann noted, "*What am I doing in France? All my domains are still around.*" The sites routinely misidentified their subjects, mixing hard-won identities with those of doppelgangers. After Lynn clicked on a SERP link to 123movies.info, she was disturbed to see not only her picture and text from her IMDB page, but also a photo of two people she did not recognize; a 2007 newspaper page; a 1965 yearbook page; and a photo of a tour group. Unfamiliar content had been mixed with important identity surrogates: "*This must've been pulled from IMDB. I don't know. I have no idea where they would've gotten this photo. MySpace or Facebook before I made it secure. ... I have no idea who these people are. ... And the class of 1965—I'm definitely not in the class of 1965.*"

Finally, and most importantly, the aggregators seem to violate emerging social norms, which acknowledge that you should have control over certain elements of your online identity, and reciprocity principles (people are given information that they would not willingly supply to others). Participants were explicitly aware of these violations. With a sense of outrage, Lynn reported: "*So this one [search result] even has my dad's house from 10, 15 years ago, and my school. See? I just hate these sites. They just make me mad. They're just mining, mining, mining. Like, oh, there are my parents. Y'know. What's your father's middle name, that's a security question on bank accounts. It's so irritating.*" Even during the interview, she began writing email to say*, 'stop posting crap about me.'*

Strategies for finding unwanted content. Our findings illustrate some of the difficulties of controlling unwanted content. Participants were not always sure about the sources of leaks, or how to discover more without revealing more. They did, however, show a range of strategies that differ from typical information seeking in pursuing this content. These included diving deep (multiple pages) into the SERP, performing image searches, and trying to check up on content that may or may not have expired.

Information seeking research has long established that Internet searchers usually do not go beyond the first page or two of results [12]. By contrast, self-search appears to be one instance in which multiple SERP pages are viewed; this exception may well be a manifestation of the fact that self-search is usually performed surmising the eyes of others. Although our self-search exercise was artificial, participants seemed to be familiar with results that occurred quite far down in the rankings (if participants seemed engaged by self-search results, we allowed them to keep going through them until they stopped of their own accord).

In addition to going deep into the results, some participants also adapted their searches to focus on media types of particular concern, usually photos. Four participants called out image search as an important component of their self-search strategies. Mary perceived Google image search as an easy technique to reveal things that had gone awry and to quickly turn up content that was difficult to find with normal Web search. She told us, "*I love image search. Because this is easy. I don't have to dig. I can do an image search, and nope! That's not me. It's very efficient. I can see if I've been linked in any blogs or anything like that. Now I feel safe.*" Kim also performed image searches on herself, although as a blogger, she found this technique far less efficient than Mary did: "*The other thing that I do once in a while that kind of disturbs me more if I do a search for myself and then do images. That one, sometimes there's some strange stuff. I mean, it's fascinating that it knows [that it's me]. So it pulls up a lot of stuff that I put on my blog. So obviously Abraham Lincoln is not me.*"

Another aspect of auditing one's online presence may involve tailoring searches. This was especially the case for participants with common names and a minimal Internet presence, and for some, finding any content about themselves proved tricky. Jane said, "*I just don't come up [in the SERP] .. and then if I do [enters a tailored search term] .. if you put 'Jane Taylor University of [name redacted]' I came up on the Contact Us page for my department, which isn't really surprising*". Another of Jane's strategies was to enter the search term 'Jane Taylor Facebook', as she expected that if any further content would come up, it would be linked to the social networking site. Other participants searched within sites like Facebook and Flickr, or commented on other portals. Sophie noted, "*I wouldn't expect it to come up if I just typed in my name, but at school cos I was involved in drama a lot I wouldn't be surprised if something came up about plays and things like that, cos in school we have our own Facebook called [name redacted] .. but that was all private*".

In summary, when self-search plays a role in identity management, the SERP is often interpreted holistically as a form of profile. Because users do not have direct control over the SERP's appearance, strategies to influence it are usually indirect, through engaging with social media sites or investing in one's own website. The presence of doppelgangers can be interpreted in a variety of ways, from playful competitors to useful cover, and were in general received much more positively than other peripheral search results, such as aggregation sites, which seemed to pollute one's online identity, often by conflating online identities or by revealing irrelevant (but personal) details.

Self-search for discovery
A second motivation for self-search was to discover new content, often in connection with user-generated media. This content may be in the form of comments on or reviews of creative efforts; the search may also be performed to find

reuse or even plagiarism. Discovery searches might be performed fairly frequently if the participant has reason to believe that there might be interesting new content.

Reactions to creative efforts. Participants involved in creative efforts were on the lookout for reviews and listings of their output (although they might be wary of running into negative reviews). Kim reported digging deeply into the SERP via self-search (in addition to searching by her book's title) to find new reviews when her novel was first released: "*Especially when my book first came out [I searched for myself], to see what reviews were coming up on it. And what people are saying and stuff. It's been a year-and-a-half now so sales are much slower. And so there's not much action. Or change. And so I don't tend to do it as much.*" Similarly, Charlie said he kept an eye out for people discussing his band's music and performances "*...sometimes I've searched for like [band name redacted], or if I've played a gig recently maybe search on Twitter, see if there's anything…*" In fact, even those participants who were concerned with the stigma of self-search allowed themselves discovery searches. For example, Harry said: "*… the only time I've ever searched for myself online was when I searched for [band name redacted] online, so that was good to look for reviews and gig listings and stuff … it would be the only way I'd know that a particular website would be advertising our gig, so that was quite good …*"

Reuse of self-generated content. For participants who posted content they had created online, self-search had the additional motive of discovering and tracking reuse. If reuse verged on plagiarism, or was done without the desired permission, participants were sometimes angered, and motivated to track down the offender. Oliver recounted the reuse of his photos (a common practice—see [18]) with and without permission: "*Some of them- the photographs, mostly people have asked, but I've occasionally- for example, when Northern Rock [a UK bank] crashed, I took a picture of the queue, and that was ripped off and used without commission in blogs.*" He had also noted an instance of reuse for a set of PowerPoint slides, a practice which he usually attempted to thwart by disseminating PDF files: "*…I went to give a presentation. I can't remember if they asked or if they [just posted the slides]. Especially in PowerPoint format. I'm not so keen on that. ...If I go and give a talk, normally I would give them a PDF version ... But stuff just leaks, and sometimes that gets, you know, reused without the credit and so on.*"

Strategies for discovery. Because discovery is a standard information-seeking activity, it is perhaps better supported than examining the SERP as a profile. Two strategies stand out in this regard, alerts and image search. Although alerts may be associated with self-search (the Pew survey notes that this is an uncommon practice for maintaining one's digital footprint [16]), among our participants, alerts were generally employed in a discovery function. Image searches on the other hand served several different functions.

Alert tools in our study included Google Alerts, although some participants were content to use alerts for specific services, such as Twitter or Facebook, where they felt they were most likely to be discussed or tagged. Jane described her feelings towards a (now retired) Facebook feature, in which a selection of recent photos appeared at the top of the profile page: "*especially now the new Facebook's got like horrid pictures of you in your profile page, I feel like I need to keep going online and checking that there's not really awful pictures there showing in that top band…*". Other participants focused on references to their websites from other sites. Thomas described, "*… I will go and have a look and see how [my website] has been linked in [from other sites] and sometimes I will comment on sort of how it has been described.*"

In addition to the role we discussed earlier in weeding out undesirable content from the SERP, image search also helped participants track reuse. Of course, participants who were amateur photographers were anxious to check the use of their photos, but image search also proved to be an easy way of understanding how ideas proliferated and were adapted by others. For example, Oliver was interested to watch how some of his diagrams propagated within his own academic community: "*If you search on my thing, which is [topic redacted], say, and go to Google Images. ... That's mine. And then this is an adaptation of mine, so it's kind of interesting to see how it propagates – how people take one and then adapt it.*"

In summary, self-search is a means for discovering content about oneself, undertaken especially when the searcher expects to find reviews of or responses to content they have published online, or when they wish to understand how their content propagates through the web.

Self-search for re-finding

A third motivation for self-search is to return to content one has published or shared online, either as an intentional effort (e.g. a scholarly paper, a photo, or music) or as a smaller digital trace that refers to something else (e.g. a blog comment or product review). Charlie said, "*sometimes I'm searching my name so I can get to a paper, because maybe I don't have a copy.*" On the other end of the spectrum, Kim described searching for her own comments to recover facts she had found; these comments would be difficult to retrieve by other means. The searches also had the advantage of returning her writing in context: "*And I've noticed that if I post a lot on [a local blog], I actually want to go back and retrieve comments [that I've written]. Like I'll do research, and I'll say, dah-dah-dah. And later I'll think, what did I dig up in that research? I will actually go into Google and search for my name on [the local blog] and whatever the topic was and I'll find my previous post.*"

Re-finding familiar content is described in greater detail as an information-seeking activity in [6].

Self-search for entertainment

A fourth motivation we wish to highlight briefly is self-search for entertainment. This was described by a minority of participants, but we include it to emphasize that not all self-search is connected with serious endeavors. Indeed even earnest self-search might be concealed to some measure under the auspices of curiosity, and may feature aspects of playfulness as participants compete with their doppelgangers for SERP position and rediscover forgotten content. But further to these, some self-searches were undertaken explicitly for fun. Sophie commented on a practice that she and her friends had adopted: "*if you're just with a group of friends or something, or on the iPad, just lazing around, then quite often people just Google everyone in the room and see what comes up.*" Image search was a key part of this activity, possibly since images are easier to browse and are more entertaining to view as a group.

The archival value of self-search

A final aspect of self-search worth mentioning is the discovery of forgotten content. Participants did not necessarily describe searching with this goal in mind, so it is difficult to characterize as a motivation. However, it was notable that some participants found content that was truly unexpected. While in some cases this content was somewhat trivial ("*So I have two SoundCloud accounts because one I made when I forgot that I had the other one. [laughs]*" – Ann), others were more meaningful. For example, Charlie came across reviews of a track he had remixed a year ago but subsequently forgotten, and Lynn found a haiku she had written for a contest (and won), again less than a year before: "*This is the haiku contest I won this past Christmas! This I don't mind. I wrote a haiku about fruitcake, and I got an honorable mention.*"

DISCUSSION AND IMPLICATIONS

When we look across the range of motivations for self-search, focusing in particular on the SERP as a dynamic profile of the participant, we can pull out two dimensions that are salient to our goal of improving self-search: *controllability* and *expectedness*. By controllability, we mean the ability to intervene in and change what is out there, either directly or by changing one's online behavior; by expectedness, we are referring to whether the participant has seen the material before, and whether they know how it got there (its source and provenance). While we might tease out other dimensions, we are choosing to focus on these two because they bear on both technology and policy.

Controllable and expected content may arise through social media, or through Web publishing. For example, people may post public Instagram photos on their own accounts; they may create Facebook profiles that come up high on the SERP; or they may publish public comments on blog posts. The self-searcher expects to see these items, and can take them down or mark them private at will.

Next are controllable, but unexpected, items: for example, candid photos may be taken by a friend, put on Facebook, and tagged. Once the self-searcher is aware of this content, it can be untagged or similarly rendered less visible via the SERP. On the other hand, if the unexpected material is flattering, the self-searcher might find it desirable to push it further up in the SERP. However, in this category, we also find social media material the self-searcher expected to be private through obscurity, but which somehow has become visible (Facebook 'likes' are notorious in this regard).

Uncontrollable expected content stems from recent data practices. For example, some public records (arrests, phone listings, address databases) have become accessible online, and have found their way onto SERPs. In the US, services such as BlockShopper publish public real estate records such as home sales history and neighborhood information; Zoopla offers an equivalent service in the UK. Even content that self-searchers had a hand in creating may not fall under their direct control (e.g. media published on a collaborator's SoundCloud account), simply due to circumstances. This type of material may figure prominently if it is not outranked by other kinds of content.

Uncontrollable and unexpected content can be the result of old data being posted online. Forgotten content such as Usenet posts and articles in newspaper archives falls under this rubric, as do aggregator results (which may or may not be controllable through social means), current reviews, and other types of published material.

Framing an Audience-Directed SERP

As our framework shows, being in control of online content about oneself is a multifaceted problem that appears not to be completely solvable; even participants who had relatively sophisticated online and self-search practices were occasionally surprised by apparent information leaks. However, the controllability dimension does suggest some implications for design, which draw in particular on the idea of the SERP viewed holistically, as a kind of profile page. In particular, the idea of *building* a SERP that presents a desirable (or at least acceptable) face, in addition to attempting to *weed out* isolated pieces of content that in fact the user has little control over, is key to our recommendations here. We are aware of different types of sophisticated services which assist users in manipulating their SERPs (e.g. reputation.com) or in creating a coherent online identity (e.g. claimID.com), but we use our results to explore a few lightweight approaches.

First of all, we highlight the potential for control over the order of social media content. It seems mutually beneficial for search engine providers and end-users to have relevant social media content appear highest on a person's SERP. Thus, giving users some control over the order of search results that relate specifically to themselves (e.g. specifying the order in which Twitter and LinkedIn appear) is one possibility. The SERP may also be tailored for a specific

audience. If the person conducting a search is connected to the target through a shorter path on Twitter or LinkedIn, perhaps these sites could be prioritized; thus a friend would see different results than a prospective employer. Stretching this idea further, users might be permitted to perform a kind of "reverse personalization," tailoring the results of a search for themselves to meet the needs of different imagined searchers, such as family members vs. colleagues. Of course, this relies on the search engine's ability to accurately categorize searchers, but relationships that have been specified through social networks may allow users to dictate preferences here. For example, Facebook friends could also be presented with Pinterest or Instagram feeds in a SERP, whereas these could be hidden from people who are only connected via LinkedIn. Finally, the existence of relationships through social network sites may also act as a pointer to who is being searched for when doppelgangers are present, enabling friends to find selected content, while also supporting the general anonymity through obscurity that some users also valued.

Secondly, we highlight awareness of content that one is nominally in control of but that has the potential to 'leak'. Stutzman and Kramer-Duffield have explored the sources of Facebook users' expectations about these boundaries [24]. For example, when participants sign up for social network sites, they rarely expect their content to be visible to aggregator services, even though they have unwittingly exposed it. This becomes even more problematic because these sites frequently mix online identities that have been curated as separate. Allowing users to work backwards and identify privacy settings that underpin such leaks would help them understand the actual effects of their choices and would help them refine specific settings (e.g. by using real data exposure to say, "I think everybody should be able to see this, but not this"). Such episodes could be used as a provocation for sending messages to social network contacts who have posted undesirable content, or to reveal the sources of public data that aggregators draw upon. Thus, users would be aided in adjusting privacy options where control is permitted, and reaching a better understanding of the leaks' source where it is not. Of course, some aggregation services are unlikely to reveal their sources. However, presenting details in the SERP that allow inferences to be drawn could be useful; for example, specifying when a site obtained a photo could imply which social network site is responsible for the leak.

For content that is completely beyond one's control, being able to put some kind of "watch" on it would at least remove the need to continually dig deep into the search results to see if it persists or has been altered. In instances where content is deeply undesirable or compromises security, bids for removal that do not require the provision of further details seem essential. Finally, some of these public data sources of what is perceived to be private data become a matter of policy, not technology.

The Process of Self-Search

The above discussion brings to the fore how the results of self-search can be used. However, the process of self-search is also worth considering in design. While search is typically cast as something that produces results, rather than something with an experiential aspect, this is clearly not the case here (and indeed, recent research points to the potential for richer design around search in general [5]).

Our results suggest four ways to support the process of self-search: (1) assist in query formulation (to help people distinguish themselves from their doppelgangers and to find unexpected content); (2) show changes in the SERP over time (e.g. webpages' relative rankings); (3) present results organized around specific content; and (4) reveal the effects of personalization. In the first case, some participants' self-search strategies (e.g. 'Jane Taylor Facebook') simply turn up familiar content, but novel results can be important, as demonstrated by participants' enthusiasm when they found unexpected content they had created themselves. Secondly, changes in the SERP over time might help users identify the source of information leaks, draw comparisons with their doppelgangers, discover the progress (both desired and not) of various pages in the rankings, and know how long unwanted content has been present for. Our third suggestion capitalizes on users' interest in how specific pieces of content, such as mp3 files or photos, have been reused, commented upon, or have otherwise attracted attention. Finally, the effects of personalization are important to self-search, yet most users don't know how to control them. Not only do they want to turn personalization off, but they also want to ask questions like, "what would people professionally connected to me see?" Such features are similar to those used by businesses in managing their online presence, but our results suggest that everyday users may also find such content of interest.

The presence of doppelgangers represents an interesting instance whereby results that are largely irrelevant to a search may enrich the experience. There was an element of competition here that could be played upon, for example by showing how search rankings change over time. But we also saw genuine interest in who these people were and how well they were doing in their own lives. The fact that some of our participants appreciated the anonymity provided by their doppelgangers means that simply excluding them from a SERP would be a mistake. Interestingly, this may be an instance whereby if search results could be wholly accurate, for example by using separate 'entities' for doppelgangers, the shift may not be appreciated. This suggests that, on the one hand, users may value a SERP that contains doppelgangers and makes their relative rankings clear; on the other, being alerted to the presence of new content, and drawing together content that is about the user as some form of dashboard, which shows where it is hosted, if it has been reused or tagged, and even if it has disappeared, may be valuable.

CONCLUSION

Self-search is a practice that is usefully distinguished from other types of information seeking. Our results show that the activity typically stems from a handful of distinct motivations, and raises new concerns. We look in particular at the SERP as a form of dynamic profile, one that might make productive use of elements like doppelgangers, and might highlight unwanted porosity between personas and stand as a straw man that exposes privacy vulnerabilities. In so doing, we examine ways in which the SERP can be made more useful for the self-searcher, as well as to the self-searcher's imagined audience.

ACKNOWLEDGMENTS

REFERENCES

1. Beal, A. and Strauss, J. *Radically Transparent: Monitoring and Managing Reputations Online*. Wiley, Indianapolis, 2008.

2. Bich, V., Gerasoulis, A. and Glover, E. Systems and methods for predicting if a query is a name. US Patent, 11 October 2007. http://www.google.com/patents/US20070239735. Accessed 7 January 2014.

3. boyd, d. and Crawford, K. *Six provocations for big data*. Oxford Internet Institute, 21 September, 2011.

4. Claypoole, T. and Payton, T. *Protecting Your Internet Identity*. Rowman & Littlefield, Lanham, 2012.

5. Dörk, M., Carpendale, S. and Williamson, C. The information flaneur: A fresh look at information seeking. *Proc. CHI 2011*, ACM (2011), 1215-1224.

6. Dumais, S., Cutrell, E., Cadiz, J.J., Jancke, G., Sarin, R. and Robbins, D. Stuff I've Seen: A system for personal information retrieval and re-use. In *Proc. SIGIR 2003*, ACM (2003), 72-79.

7. Gehl, R. Ladders, samurai, and blue collars: Personal branding in Web 2.0. *First Monday 16*, 9 (2011).

8. Google Profiles. http://www.google.com/+/learnmore/profile. Accessed 7 January 2014.

9. Gross, B. Names of our lives. In *Proc. CSE 2009*, IEEE (2009), 747–752.

10. Henry, A. How to clean up your online presence and make a great first impression. *Lifehacker*, 28 November 2012. http://lifehacker.com/5963864. Accessed 7 January 2014.

11. Herzog, W. and Stutzman, F. The case for online obscurity, *California Law Review 101*, (2013), 1-35.

12. Jansen, B., Spink, A. and Saracevic, T. Real life, real users, and real needs: A study and analysis of user queries on the Web. *IP&M 36*, 2 (2000), 207-227.

13. Jones, R., Kuman, R., Pang, B. and Tomkins, A. Vanity fair: Privacy in querylog bundles. In *Proc. CIKM 2008*, ACM (2008), 853-862.

14. Jones, W. and Teevan, J., eds. *Personal Information Management*. UW Press, Washington, 2007.

15. Lindley, S., Marshall, C.C., Banks, R., Sellen, A. and Regan, T. Rethinking the web as a personal archive. In *Proc. WWW 2013*, (2013), 749-760.

16. Madden, M. and Smith, A. *Reputation Management and Social Media*. Pew Internet & American Life Project, 26 May 2010.

17. Marchionini, G. *Information Concepts: from Books to Cyberspace Identities*. Morgan & Claypool, 2011.

18. Marshall, C.C. and Shipman, F. The ownership and reuse of visual media. In *Proc. JCDL 2011*, ACM (2011), 157-166.

19. Mayfield, A. *Me and My Web Shadow: How to Manage Your Reputation Online*. A&C Black Publishers, 2010.

20. Nicolai, T., Kirchhoff, L., Bruns, A., Wilson, J. and Saunders, B. The self-Googling phenomenon: Investigating the performance of personalized information resources. *First Monday 14*, 12 (2009).

21. Odom, W., Zimmerman, J. and Forlizzi, J. 2011. Teenagers and their virtual possessions: design opportunities and issues. In *Proc. CHI 2011*, ACM (2011), 1491-1500.

22. Pash, A. Order free Google business cards for a limited time. Lifehacker. 1 May 2009. http://lifehacker.com/5236055/order-free-google-profile-business-cards-for-a-limited-time. Accessed 7 January 2014.

23. Soghoian, C. The problem of anonymous vanity searches. *ISJLP 3*, 2 (2007), 297-316.

24. Stutzman F. and Kramer-Duffield, J. Friends only: Examining a privacy-enhancing behavior in Facebook. In *Proc. CHI 2010*, ACM (2010), 1553-1562.

25. Trapani, G. Forget the business card. Just Google me. *HBR Blog Network*, 10 August 2009. http://blogs.hbr.org/trapani/2009/08/forget-the-business-cardjust-g.html. Accessed 7 January 2014.

26. Yao, T., Zhang, M., Liu, Y., Ma, S. and Ru, L. Empirical Study on Rare Query Characteristics. In *Proc. WI-IAT 2011*, Vol. 1. IEEE (2011), 7-14.

27. Zimmer, M. Privacy on planet Google: Using the theory of contextual integrity to clarify the privacy threats of Google's quest for the perfect search engine. *J. Bus. & Tech. Law 109*, 3 (2008), 109.