
Self-Presentation: Structured and semi-structured user profiles

Linda Le

University of Washington
School of Information
lindale13@gmail.com

Ivan Beschastnikh

University of Washington
Computer Science and Engineering
ivan@cs.washington.edu

David W. McDonald

University of Washington
School of Information
dwmc@uw.edu

Abstract

The study of self-presentation in online social networks has relied on data collected from structured profiles. However, structured profiles may present a biased view by influencing the type of information users express. This paper investigates an alternative form of examining self-presentation through the study of semi-structured profiles provided by Wikipedia. Overall, we seek to understand how reducing profile structure changes the nature of online presentation.

Introduction

User profiles are an important component in online communities serving social and functional purposes. Profiles are a form of self-presentation; they let a user create an online identity to express herself to fellow community members. A user profile often consists of a collection of personal information, individual traits, characteristics, and user interests. Many social networking sites use a template of predefined questions regarding the user's interests, preferences, and personal characteristics to guide the creation of a profile. We refer to this type of profile as *structured* -- profile content is structured by predefined questions or fields. However, profiles can also be *unstructured*, where the user is given a blank page, with little or no prompt as to what she should say about herself.

Copyright is held by the author/owner(s).

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.

ACM 978-1-60558-930-5/10/04.

Much of the prior research into online user profiles has defined fields. By looking at what fields have been completed, researchers have studied the relationship between profile completion and number of friends (Lampe, Ellison & Steinfeld, 2007); surveyed information disclosure behavior in online communities (Schrammel, 2009), and analyzed privacy techniques users perform to keep their identity safe (Young, 2009). An experimental study by Stecher & Counts (2008) considered how many fields were filled out in structured profiles and the perceived utility of each characteristic field. However, the nature of structured profiles may present a biased view of how people present themselves online. That is, given a specified list of possible attributes, it is likely that users would choose to present those attributes in their profile over others that are not easily included.

This potential bias in self-presentation motivates our research. In particular our current study considers how reducing the structure of online user profiles changes the nature of what users choose to present about themselves. We rely on the user pages of Wikipedia to conduct our study. Wikipedia is a popular online encyclopedia composed by volunteers. While much of the attention surrounding Wikipedia focuses on the encyclopedic content, Wikipedia has additional aspects that facilitate social coordination and social interaction that is a critical enabler of content creation.

Self-Presentation in Wikipedia: Userboxes

One important social facilitator allocated to each registered Wikipedia user is a *user page*. A user page consists of a blank web page that a user can customize with a textual description of herself or any number of

Wikipedia specific page decorations and tags. Wikipedia policy governing user pages states: *Wikipedia provides user pages to facilitate communication among participants in its project to build an encyclopedia. Generally, you should avoid substantial content on your user page that is unrelated to Wikipedia. Wikipedia is not a general hosting service, so your user page is not a personal website. Your user page is about you as a Wikipedian, and pages in your user space should be used as part of your efforts to contribute to the project.* While many user pages include Wikipedia “work” specific content, many user pages also include descriptions of personal information, interests, traits, and characteristics. In fact, the inclusion of personal attributes on user pages is popular enough that it has been simplified and encoded through Wikipedia’s template mechanism as userboxes.

Userboxes are small colored boxes displayed on a user page that are designed to communicate a single characteristic, interest, or trait of the user. The content of userboxes resembles text found in many structured profile fields. Userboxes are often placed on the page in a table that groups them in one place so they can be quickly scanned to get an overview of who the user is; much like any other online profile. Userboxes serve both social and functional purposes by highlighting skills and interests of the individual.

A userbox is placed on a user page by including the wikitext of a userbox template on the page. When the user page is rendered, the appropriate HTML of the userbox is transcluded into the resulting page (see Figure 1). Templates in Wikipedia can be parameterized and userbox templates use this feature to allow customization of userbox presentation.

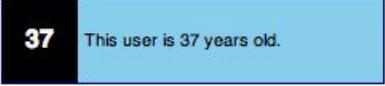
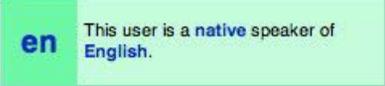
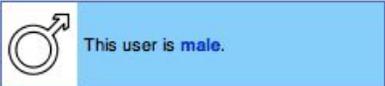
Userbox Image	Freq.
 <pre>{{Template:User_age 37}}</pre>	14347
 <pre>{{Template:User_en}}</pre>	12980
 <pre>{{User:UBX/male}}</pre>	5798
 <pre>{{User:The_Raven's_Apprentice/Userboxes/User_Firefox}}</pre>	5525
 <pre>{{User:Menasim/Userboxes/User_Google}}</pre>	4656

Figure 1. Top five userboxes with frequency and parametrized wikitext templates.

Userboxes represent a *semi-structured* profile creation technique. Userboxes exist somewhere between the structured user profiles of many social networking sites and a page with a free text description of the user.

Semi-Structured and Structured Profiles

Our analysis focuses on the specific text contained in each individual userbox. We parsed the English Wikipedia database dump from February 2008 to collect each application of all userbox templates to user pages. Our parser considers common userbox template forms and is conservative, allowing for few false positives but

possibly many false negatives. In total we found about 9,200 distinctive userboxes with a total frequency of about 662,000. Our study concentrates on the 280 most frequently applied userbox templates (Figure 1 shows the five most frequently used), each of these was used at least 450 times and in total our sample accounts for 42% of overall userbox usage.

Two coders independently viewed each userbox and categorized it as appropriate to fill one of the current Facebook fields. When coders disagreed on the appropriate field, the userbox was discussed until the coders reached agreement. In cases where the userbox did not match a field but still fit into the field groupings, these userboxes were coded under Facebook's field categories of Basic, Personal, Contact, or Work and Education information. Userboxes that did not seem appropriate for any of Facebook's profile fields were placed into an Other category.

In analyzing the distribution, we summed the usage frequency of each userbox as categorized in a Facebook field and divided it by the total usage of the sample.¹ This distribution differs from that of structured profiles. Table 1 presents a comparison of userbox distribution to four other research papers studying profile field completion.

From our sample of 280, 78.4% of the userboxes fit the paradigm of content found in a Facebook profile. The other 21.6% were either too ambiguous or irrelevant to the profile user space. In the distribution of userboxes

¹ Our analysis accounts for double counted sets of userboxes that map to the same Facebook category when used by the same user.

		Lampe 2007	Stetcher 2008	Young 2009	Schrammel 2009	Userboxes
Basic	Sex	93.8%	97%	M (75%) F (72%)	-	3.5%
	Birthday	83.8%	97%	M (95%) F (95%)	Fri (79%) Un (25.3%)	6.5%
	Interested in	-	66%	-	-	0.7%
	Religious Views	-	43%	-	-	0.2%
Personal	Activities	-	83%	-	-	9.3%
	Interests	77.7%	77%	M (70%) F (72%)	-	10.6%
	Favorite Music	78.2%	93%	-	-	5.6%
	Favorite TV Shows	46.5%	87%	-	-	0.5%
	Favorite Movies	80.1%	87%	-	-	0.4%
	Favorite Books	66.9%	83%	-	-	1.2%
	Favorite Quotations	73.8%	67%	-	-	0.2%
	About Me	59.8%	73%	-	-	37.1%
Contact	Emails	92.3%	-	M (85%) F (86%)	Fri (62.6%) Un (8.7%)	0.5%
	IM Screen Name	67.8%	-	M (20%) F (15%)	Fri (63.2%) Un (15.7%)	0.4%
Edu/Work	Concentration	89.5%	80%	-	-	0.3%
	High School	87.1%	83%	-	-	0.2%
	Employer	-	53%	-	-	0.2%

Table 1. Field completion comparison of userboxes (M) male (F) female (Fri) friends only (Un) unknown people (-) not filled.

relative to Facebook categories, 86.2% of the userboxes were classified as Personal Information, 9.5% as Basic Information, 0.89% as Contact Information, and 0.77% as Education and Work Information.

Our preliminary findings suggest that Wikipedians use semi-structured profiles to present a significantly different version of themselves when compared to users in communities with structured profiles. For example, by far the dominant category of userboxes is the equivalent of 'About Me' in Facebook, whereas in Facebook this category is not even in the top 5. Clearly, the purpose of these communities are different, however we believe the more open-ended nature of Wikipedia profiles has significant influence on what users disclose about themselves.

In the future we would like to consider personality characteristics of userboxes. In our experience we have found these to be highly revealing and nuanced. We would like to determine the scope of personality traits that userboxes may reveal and what this may signal to other editors and readers of Wikipedia.

References

- Lampe, C., Ellison, N., & Steinfield, C. (2007). A Familiar Face(book): Profile Elements as Signals in an Online Social Network. *SIGCHI, AMC Press*, 435-444.
- Schrammel, J., Koffel, C., & Tscheligi, M. (2009). How Much do you Tell? Information Disclosure Behaviour in Different Types of Online Communities. *CT*, 275-284.
- Stecher, K., & Counts, S. (2008). Thin Slices of Online Profile Attributes. *ICWSM, AAAI Press*, 127-135.
- Young, A., & Quan-Haase, A. (2009). Information Revelation and Internet Privacy Concerns on Social Network Sites: A Case Study of Facebook. *CT*, 265-273.