

---

# Findings of a User Study of Automatically Generated Personas

**Joni Salminen**

Qatar Computing Research Institute, Hamad Bin Khalifa University  
and

Turku School of Economics  
jsalminen@hbku.edu.qa

**Soon-Gyo Jung**

Qatar Computing Research Institute, Hamad Bin Khalifa University  
sjung@hbku.edu.qa

**Jisun An**

Qatar Computing Research Institute, Hamad Bin Khalifa University  
jisun.an@acm.org

**Haewoon Kwak**

Qatar Computing Research Institute, Hamad Bin Khalifa University  
haewoon@acm.org

**Bernard J. Jansen**

Qatar Computing Research Institute, Hamad Bin Khalifa University  
jjansen@acm.org

**Author Keywords**

Data-driven Personas, Persona Analytics, User Experience

**ACM Classification Keywords**

H.5.2. [Information Interfaces and Presentation] User Interfaces - Theory and methods, User-centered design

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

*CHI'18 Extended Abstracts, April 21–26, 2018, Montréal, QC, Canada*

© 2018 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-5621-3/18/04.

<https://doi.org/10.1145/3170427.3188470>

**Abstract**

We report findings and implications from a semi-naturalistic user study of a system for Automatic Persona Generation (APG) using large-scale audience data of an organization's social media channels conducted at the workplace of a major international corporation. Thirteen participants from a range of positions within the company engaged with the system in a use case scenario. We employed a variety of data collection methods, including mouse tracking and survey data, analyzing the data with a mixed method approach. Results show that having an interactive system may aid in keeping personas at the forefront while making customer-centric decisions and indicate that data-driven personas fulfill information needs of decision makers by mixing personas and numerical data. The findings have implications for the design of persona systems and the use of online analytics data to better understand users and customers.

**Introduction**

In on-going work, we have developed a system for automatic persona generation [1] using large-scale online social media and other online analytics data. Our premise is that automatically generating personas with actual user data and quantitative methods result in data-driven personas that are rapidly updated and correspond to real user behavior. As such, the APG approach addresses many of the prior critiques of personas not being based on actual user data, taking too long to create, and rather quickly staling as user



Figure 1a: Persona Photo.



Figure 1b: Persona Profile.

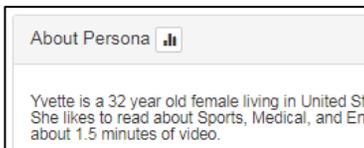


Figure 1c: Persona Description.

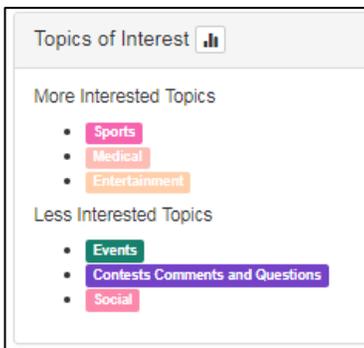


Figure 1d: Topics of Interests.

segments change [1]. These shortcomings have been widely discussed among HCI scholars [3].

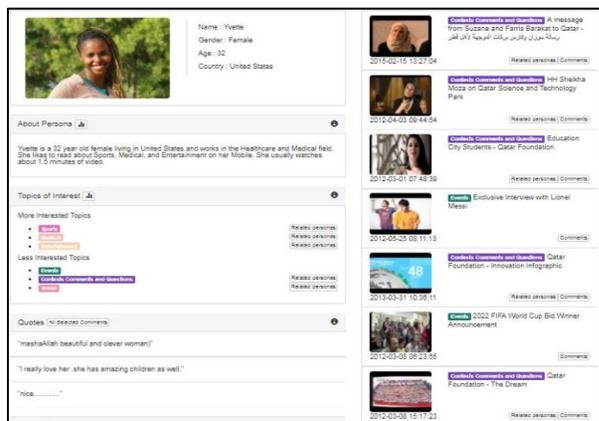


Figure 1: APG Persona Profile; see Figs. 1a-1d.

We have made substantial progress on system development with the Automatic Persona Generation (APG) system that is currently fully functional and processes tens of millions of user interactions with online content. The system is under beta deployment at several businesses and organizations. Being online, the system has a host of unique features that greatly expand the traditional format and manners of engagement with personas profiles, which have typically been a one or two-page static layout. APG offers a variety of presentation features. The APG system generates a typical profile for individual personas, as shown in Fig. 1, with excerpts in Figs. 1a - 1d. We have reported automatic persona generation methodology in previous research [1] and will refrain from repeating it here.

In preparation for a larger scale longitudinal user study, we organized and implemented a pilot study within the workplace of a major beta client of the APG system in a semi-natural environment with a diverse group of corporate employees. We collected an array of mixed data that sheds insights on potential use cases and system enhancements for APG, as well as on the implications of generating personas from online user data. We present our research, findings, and implications in the following sections. This research is important as prior work has shown that organizations that effectively leverage personas in their decision making see a positive return on investment [2].

### Research Objectives

The five APG system features that we were interested in evaluating are: (1) persona listing, (2) content with related personas, (3) current reach versus potential reach, (4) persona comparison, and (5) the chronology of personas. Each feature is explained in more detail in the following.

**Persona listing:** the user can generate a listing of 5 to 15 personas with options to select a particular social media platform, currently either YouTube or Facebook. Depending on the selections, a list of the appropriate number of personas is presented as a result listing of persona images (see Fig. 2 and Fig. 2a-2b). Once generated, the user can apply filters, such as location, to narrow the persona listing. Hovering over an image triggers a dialog box of the persona name and biographical data. Clicking on an image takes the user to the persona profile.



Figure 2a: Persona Profile Image.



Figure 2b: Content Selection Option.

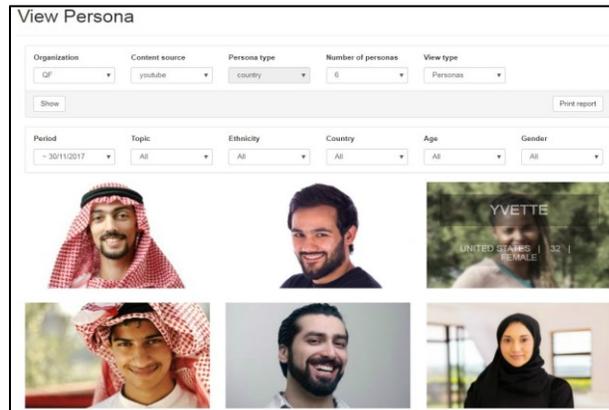


Figure 2: Persona Listing of 6 from a Range of 5 to 15. Hovering Over Persona Images Shows Persona Details (e.g., Yvette). See Figs. 2a and 2b.

**Content with related personas:** the user generates a list of the content (by selecting View type as 'Contents') in Fig. 3, based on how many personas are selected, and the results listing presents each piece of content with the associated personas that are interested in it. That is, instead of personas with related content, this feature shows the content and the associated personas, as shown in Fig. 3 and Fig. 3a.

**Current reach versus potential reach:** this feature permits the user to compare the current reach of the selected number of personas with the potential reach of the same set of personas, as shown in Fig. 4.



Figure 3a: Content With Associated Personas.

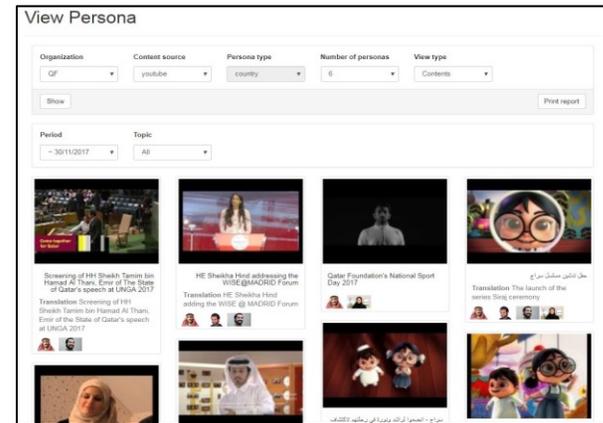


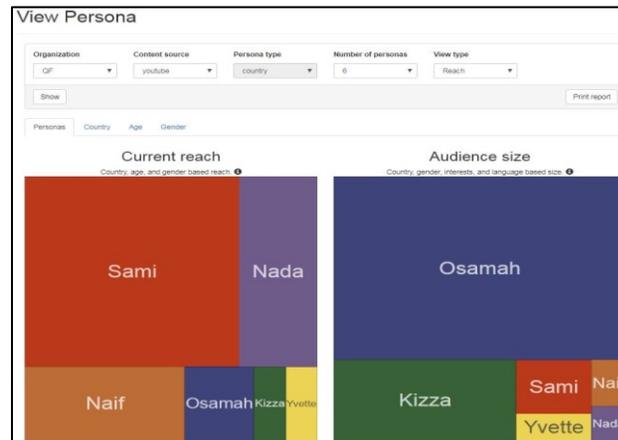
Figure 3: Based on the Number of Personas Selected, This Feature Displays All the Content and Personas that Content Is Associated With via Clickable Icons Below Each Piece of Content. See Fig. 3a.

Reach is the size of the user segment, which the APG calculates using the demographics and interest of the personas by leveraging the Facebook Marketing API from which we periodically collect interest-based demographic data with an automated Python script.

**Persona comparison:** this feature presents the listing of personas in a compact form with the basic information and top interesting content to facilitate comparison among the personas in the selected listing, as shown in Fig. 5.

<p><b>Persona Listing</b> enables the user to generate a listing of 5 to 15 personas by selecting a social media platform (data source)</p>
<p><b>Persona Content</b> shows a list of the content, based on how many personas are selected</p>
<p><b>Personas Reach</b> compares the current reach of the selected number of personas with the potential reach of same personas</p>
<p><b>Persona Comparison</b> presents the listing of personas in a compact form with the basic information and most interesting content</p>
<p><b>Persona Chronology</b> shows all the personas generated for the organization and their most and least interesting topics over time</p>

**Table 1:** Five Key Features of the APG System Evaluated in the User Study.

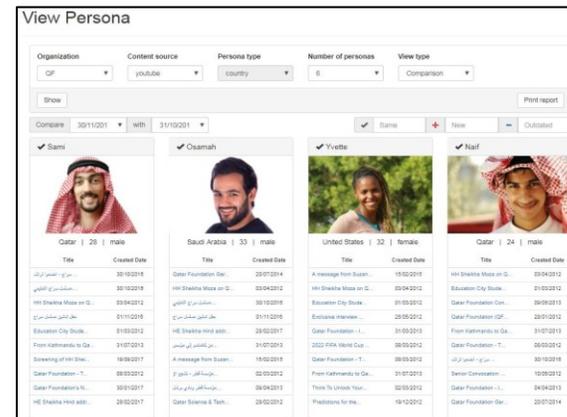


**Figure 4:** The Persona Reach Option Shows the Current Persona Portion Measured by Audience Reach Versus Potential Reach Given the Overall Audience Size.

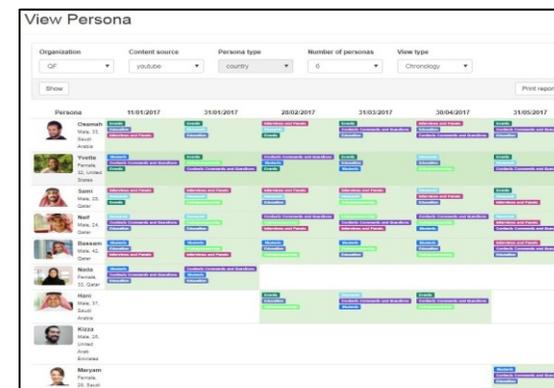
**Chronology of personas:** this feature shows all the personas generated for the organization and their most and least interested topics, presenting the changes of each persona over time, along with personas that enter and depart with each new data collection period, as shown in Fig. 6. Currently, the personas are generated on a monthly basis by re-running matrix factorization calculations [6]. The key aspects of the five APG system features evaluated are presented in Table 1.

**Methodology and Data Collection**

We conducted a semi-naturalistic user study of 90-minute duration within the workplace of a major worldwide corporation that has an extensive online presence on multiple social media channels. The organization is using personas for various aspects of their social media, customer relationship management, and online advertising efforts.



**Figure 5:** Persona Comparison Displays Each Persona With Associated Content.



**Figure 6:** The Persona Chronology Shows the History of Each Persona for Each Data Collection Period.

The three data collection methods are detailed in Table 2. There were 13 study participants, 4 females (30.8%) and 9 males (69.2%) from a variety of positions within the company as shown in Table 3.

<b>Quantitative Survey:</b> participants evaluated each of the five system features on a Likert scale of 5 from 'not very helpful' to 'very helpful' for task completion.
<b>Qualitative Survey:</b> participants provide, via a survey, qualitative responses to "Describe how the feature is helping to complete your task."
<b>System Logging:</b> participants use of the system and mouse movement was logged recording page view, mouse movement on the system pages, clicks, etc.

**Table 2:** The Three Data Collection Methods Employed in the User Study.

Position	No.
Social Media Officer	3
CRM Analyst	1
Digital Marketing	2
Data Analysis Officer	3
Manager	3
System Controller	1
Total	13

**Table 3:** Job Positions of the 13 Participants in APG User Study.

The study organization was as follows. The participants were given a brief introduction on personas to make sure they understand the concept, completed IRB approval, given a short overview of the APG system, presented the scenario, and then given approximately one hour to engage with the five features of the system within the context of the given scenario. The study was semi-naturalistic so that the participants were given a common task, but each participant worked on their own computer with minimal obstruction by researchers beyond answering questions and engaging in discussions concerning the APG system. The task was: *Your CEO has given you a task to plan a marketing campaign for your core customers that enhances their brand perception as a great company. Using the personas system, you need to analyze the current core customers and come up with the right message to achieve this goal. How will personas help you complete the task?* All participants stayed for the entire hour.

**Results**

*Quantitative Survey*

Each of the participants rated the five APG system features with results shown in Table 4. As shown, the participants gave each feature a reasonably high rating, with View Personas and View Reach being higher. This indicates that participants are satisfied with a detailed view of generated personas rather than its interaction (contents) or change (chronology or comparison).

*Qualitative Survey*

There were numerous qualitative comments from the 13 participants, including the access that the APG provided to relevant personas and based on current data. Due to space limitations, we focus on the most requested additional APG features.

APG Feature	Average	Max	Min	Std Dev
<b>View persona</b>	3.91	5	2	0.831
<b>View contents</b>	3.50	5	2	1.179
<b>View reach</b>	3.82	5	3	0.874
<b>View comparison</b>	3.60	5	3	0.843
<b>View chronology</b>	3.60	5	2	0.966

**Table 4:** Participant Ratings of the Five Features of the APG System Evaluated in the User Study.

- **Numbers:** Several participants requested more numbers. This was counter to our expectation that the value of personas is the lack of numbers. For example, "Aggregated interaction rate will be vital (*about chronology*)."
- **Additional visualization options:** There were several requests for more visualizations (e.g. pie charts and interactivity graph). We have already planned for these in the version 1.0 release.
- **Region-Level Personas:** In addition to country and city level personas, regional managers would like to have area personas (e.g., Europe, Asia, Americas). We are considering this feature.

*System Log Analysis*

All user interactions (i.e., mouse clicks) during the user study were logged, with some of the findings reported here. As shown in Table 5, there were 2,425 interactions, with an average of 186.54 interactions per participant.

Table 6 shows that the Persona Listing feature was the most interacted with (59.6%), with all five features

Feature	No.	%
Listing	860	59.6%
Content	264	18.3%
Reach	121	8.4%
Comparison	121	8.4%
Chronology	77	5.3%
Total	1443	100.0%

**Table 6:** Usage of the Five Key Features Ranked by the Number of Interactions.

No. Personas	Count	%
10	209	40.0%
5	131	25.1%
15	61	11.7%
9	43	8.2%
7	42	8.0%
6	22	4.2%
14	10	1.9%
11	4	0.8%
Total	522	100.0%

**Table 7:** The Most Popular Choices of the Number of Personas by the Participants During the Study.

being used multiple times. Although the Content feature was used frequently, it was the lowest rated by the participants.

As shown in Table 7, when the users were given the choice to select how many personas to view (between 5 and 15), 40% of the persona listings were for 10 personas, which may be indicative of user preferences, an aspect we will investigate in future research. There also seems to be a preference for multiples of 5, as 15 and 5 personas were also quite popular. The default choice given by the system is five personas.

Measure	No. of Interactions
Total	2425
Average	186.54
Max	536
Min	37
St. Dev.	130.3

**Table 5:** The Number of Total Interactions With the APG System and Associated Statistical Results.

As shown in Table 8, persona number 5 was the most commonly selected (32.6%) with personas 1 (21.1%) and 3 (11.9%) also popular. The most popular persona positions are at the first or last of the rows (see Fig. 2); suggesting an ordering effect. In addition, the first and the last personas in a row get the most views, indicating an observational pattern in the layout.

### Implications and Future Research

We will incorporate additional APG system features, specifically supporting numbers, prior to the launch of a full-scale user study. Findings concerning the user preferences for the number of personas wanted opens

future research into generating an optimal number of personas for an organization.

Persona No.	Count	%
5	170	32.6%
1	105	20.1%
3	62	11.9%
2	51	9.8%
Others	134	23.30%
Total	522	100.0%

**Table 8:** Persona Selection Frequencies During the Study.

Contrary to previous persona use highlighting the value of qualitative user attributes relative to numbers, our participants were keen to know the *numbers behind the data*. This leads us to assert that further investigation is needed to develop personas that combine both numerical and qualitative user attributes. This *persona analytics* leverages a combination of qualitative attributes (e.g., topical interests) and numerical data.

### References

- [1] S. Jung, J. An, H. Kwak, M. Ahmad, L. Nielsen, and B. J. Jansen, "Persona Generation from Aggregated Social Media Data," in *ACM Conference Extended Abstracts on Human Factors in Computing Systems 2017 (CHI2017)*, Denver, CO, 2017, pp. 1748-1755.
- [2] V. L. Drego and M. Dorsey, "The ROI Of Personas," Forrester Research 3 Aug. 2010.
- [3] C. N. Chapman and R. P. Milham, "The Personas' New Clothes: Methodological and Practical Arguments against a Popular Method," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 5, pp. 634-636, Oct. 2006.