3-2018

Automatic persona generation (APG): A rationale and demonstration

Soon-Gyo JUNG

Joni SALMINEN

Haewoon KWAK

Jisun AN
Singapore Management University, jisunan@smu.edu.sg

Bernard J JANSEN

Automatic Persona Generation (APG): A Rationale and Demonstration

Soon-gyo Jung
Qatar Computing Research Institute
Hamad Bin Khalifa University
sjung@hbku.edu.qa

Joni Salminen
Qatar Computing Research Institute
Hamad Bin Khalifa University
jsaminen@hbku.edu.qa

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

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

Bernard J. Jansen
Qatar Computing Research Institute
Hamad Bin Khalifa University
jjansen@acm.org

ABSTRACT

We present Automatic Persona Generation (APG), a methodology and system for quantitative persona generation using large amounts of online social media data. The system is operational, beta deployed with several client organizations in multiple industry verticals and ranging from small-to-medium sized enterprises to large multi-national corporations. Using a robust web framework and stable back-end database, APG is currently processing tens of millions of user interactions with thousands of online digital products on multiple social media platforms, such as Facebook and YouTube. APG identifies both distinct and impactful user segments and then creates persona descriptions by automatically adding pertinent features, such as names, photos, and personal attributes. We present the overall methodological approach, architecture development, and main system features. APG has a potential value for organizations distributing content via online platforms and is unique in its approach to persona generation. APG can be found online at https://persona.qcri.org.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI) → Empirical studies in HCI**

KEYWORDS

Personas; User Experience Research; User Analytics

ACM Reference format:
Jung, S.G., Salminen, J., Kwak, H., An, J., and Jansen, B. J. Automatic Persona Generation (APG): A Rationale and Demonstration. In *Proceedings of ACM SIGIR Conference on Human Information Interaction and Retrieval, New Brunswick, NJ, USA, 11-15 March 2018 (CHIIR'18)*, 4 pages. DOI: https://doi.org/10.1145/3176349.3176893

CHIIR'18, March 11-15, 2018, New Brunswick, NJ, USA.

ACM ISBN 978-1-4503-4925-3/18/03...$15.00
DOI: http://dx.doi. https://doi.org/10.1145/3176349.3176893

1 INTRODUCTION

The instrument of personas is used in a wide range of domains from software development to system design to marketing in order to describe and communicate about core users and customers. A persona is a fictional individual embodying the attributes of a key market segment that could be users, audience, or customers. In simple terms, a persona is a representation of a group of people with similar behaviors and demographics. One can think of a persona as a mental shortcut that explains the who, what, and why of a set of individuals. Personas can help define organizational strategy, develop new products, and improve customer operations. Personas are integrated into many design and product development processes [1, 2].

However, personas have traditionally been challenging to create [3, 4], as their creation is an expensive and slow process, involving ethnographic studies, surveys, or focus group interviews. Given that these methods are one-time data collection events, the created personas can quickly stale requiring another round of data collection. Also, without the opportunity to verify the data collection process, end users of personas lack confirmation of whether or not the personas are representative of their current target users. These limitations are some of the motivations for our efforts to automate persona creation with system development and underlying research that addresses the major shortcomings of manual persona creation.

APG leverages privacy-preserving aggregated data of user interactions with product content posted on major online social media and other analytics platforms. APG collects, processes, and decomposes this actual user data and then enriches the results systematically with descriptive attributes to produce data-driven persona profiles. APG can generate personas from millions of user interactions within a couple of days versus the months it takes to create personas via traditional approaches [5].

Here, we provide an overview of APG, including architecture and development. Then, we highlight some of the many features of APG, focusing on the core features. We then discuss commercial efforts underway, followed by a conclusion presenting a brief recap.

2 SYSTEM OVERVIEW

APG is built using a stable, robust, and scalable structure employing (a) the Flask web framework to support the front-end applications, services, and application programing interfaces (APIs), (b) PostgreSQL database for back-end data storage, processing, and (c) Python libraries including Pandas and scikit-learn for data analysis.

For configuration and data collection, APG accesses the targeted online social media platforms, for example, Facebook or YouTube, via the analytics API provided to the account holders. Typical user profile data from these platforms are demographic variables of gender, age, country location, and which site the user comes from, although at an aggregated level. Via the social media platform's API, the APG system also collects the detailed interactions of users with the online content. This detailed level of data is accessible only to managers of a particular social media channel.

To automatically generate personas from this data, our methodology [6] requires a sequential approach, consisting of:

- identifying the distinct user interaction patterns from the data set,
- linking these distinct user interaction patterns to the set of user demographic groups,
- identifying impactful user demographic groups from the data set,
- creating shell personas via demographic attributes, and
- enriching these shell personas to create rich persona descriptions.

As the data from online platforms are aggregated, APG must disaggregate it. To do this, we develop a matrix representing users' interaction with the online content, such as videos. We denote by **V** the g×c matrix of g user groups (G_1, G_2, ..., G_g) and c contents (C_1, C_2 ,..., C_c). The element of the matrix **V**, V_{ij}, is any statistic that represents the interaction of user group G_i for content C_j. With this matrix as the basis, APG can identify first distinct user behavior patterns (which can be patterns of associations between users and contents) and then the impactful user segments from this set of distinct user patterns.

Once APG has the matrix **V**, the next step is discovering the underlying latent factors, i.e. the user interaction patterns, that become the basis of the personas. APG uses non-negative matrix factorization (NMF) [7] that can find multiple behavioral patterns even from a single group. As a behavioral pattern can be associated with multiple demographic groupings, APG selects the demographic group with the largest coefficient.

The result of NMF is a set of shell personas that we turn into rich personas by adding personal attributes. To generate a name for a persona, we build a dictionary of names by collecting popular names by gender and year from the 181 countries. Through <age group, gender, and country> of a representative group, APG can automatically assign an age, gender, and country appropriate name to a persona. To assign a photo to a persona, APG stores nearly 4,000 commercial stock photos of models for different ethnicities, genders, and ages, to which the copyrights were purchased. Through <age group, gender, ethnicity, country> of a representative user segment, APG can assign an appropriate photo to a persona. APG can filter users based on topical interests by leveraging the collection of content consumed by that persona.

The final result of this attribute incorporation is rich insightful persona profiles automatically generated from aggregated, privacy preserving social media data, with the entire process outlined in Fig. 1.

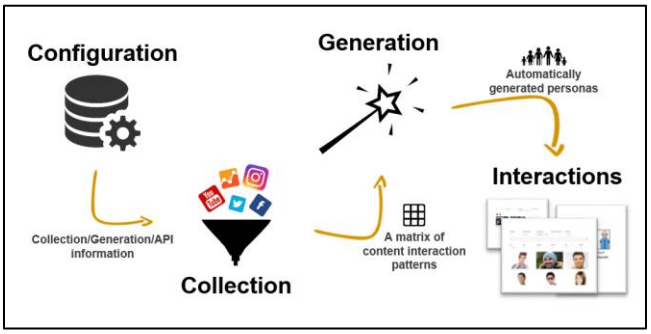

Figure 1: APG flowchart from server configuration to data collection, to persona generation, to users accessing the system.

We now present some of the main features of APG to demonstrate the capabilities of the system, with others shown during the live demonstration.

3 SYSTEM FEATURES

3.1 APG Interface and Persona Listing

3.1.1 APG Interface. As shown in Fig. 2, a user who wishes to generate personas can select the number to generate, with the system currently set for a minimum of 5 and a maximum of 15. The user can also select the desired platform and is provided options for viewing the personas. The user can generate the desired number of personas via the Show button.

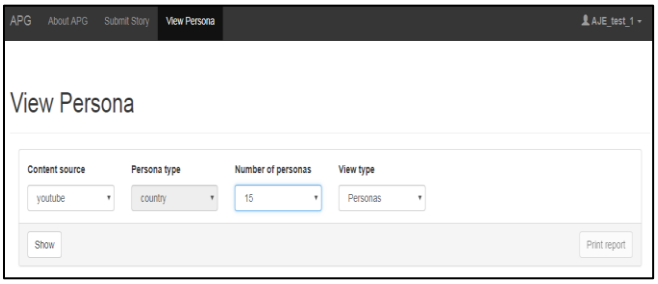

Figure 2: APG interface with options for selecting the content source platform (e.g., Facebook, YouTube), the number of personas to generate from the data (5 to 15), and the view type (various formats of viewing the relationship between content and personas). The options in the screen capture are set to YouTube and 15 personas.

3.1.2 APG Persona Listing. The listing of the personas, combined with the mouseovers, afford the user the ability to get an overview of the audience segments in terms of demographic information, as shown in Fig. 3. By hovering over one of the

persona images, the basic persona attributes are displayed, in this case '*Andrew, United Kingdom, 31, Male*'. The user can also change content collections, change the number of personas, or apply different filterings, such as data set, topics of interest, ethnicity, country, age, or gender.

3.2 Persona Profile

Clicking on one of the images in the persona listing will display the corresponding persona profile, with one persona profile displayed in Fig. 4. Each of the displayed sections of the persona is discussed below.

3.2.1 Entire persona set. Along the top of the persona profile is a series of preview pictures permitting navigation to another of the persona profiles.

3.2.2 Image and base information. The persona profile contains an appropriate image along with the base information (i.e., name, age, country, and gender) of the persona.

3.2.3 About Persona. This section is a short snippet in paragraph form that recaps the base information, combined with topics of interest, the device used for the interaction with the online platform, and level of engagement with content (see Fig. 4). Clicking on the graph icon shown in the title bar displays the proportions of career, marriage, parenting, and educational status (see Fig. 5).

3.2.4 Topics of Interest. As part of the matrix decomposition, we can identify the specific pieces of content that a persona interacts with. APG topically classifies this content, displaying the most and least interesting topics for the persona (see Fig. 4), which is an on-going research efforts [8].

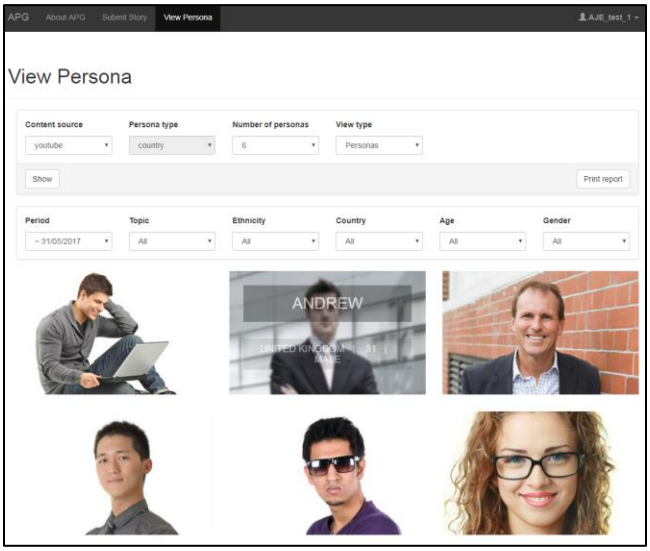

Figure 3: APG persona listings, with mouse-over data displayed for one persona. As APG archives all data collections, the user can change the data set. The user can also apply filters of Topic, Ethnicity, Country, Age, and Gender.

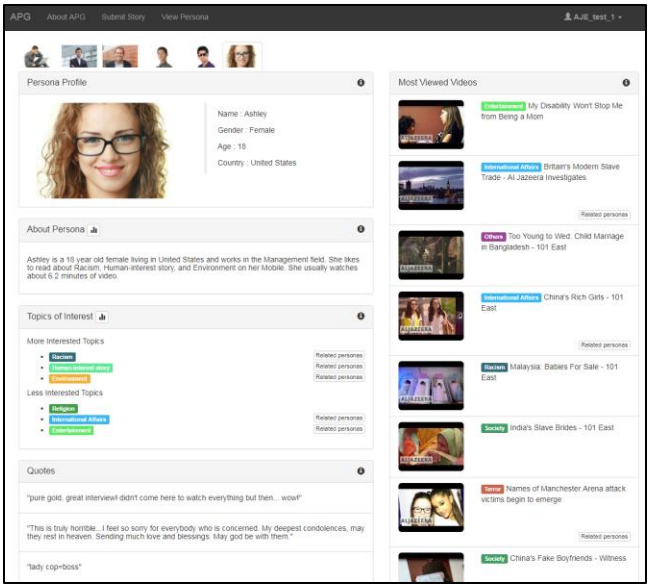

Figure 4: Persona profile for Ashley. The top ribbon presents the entire set of personas, in this case 6. The image is gender and age appropriate. There is an About section outlining device used and interaction time. Most liked and least liked topics are displayed, followed by social media comments left by those who are like Ashley. Along the right rail are top videos interacted with.

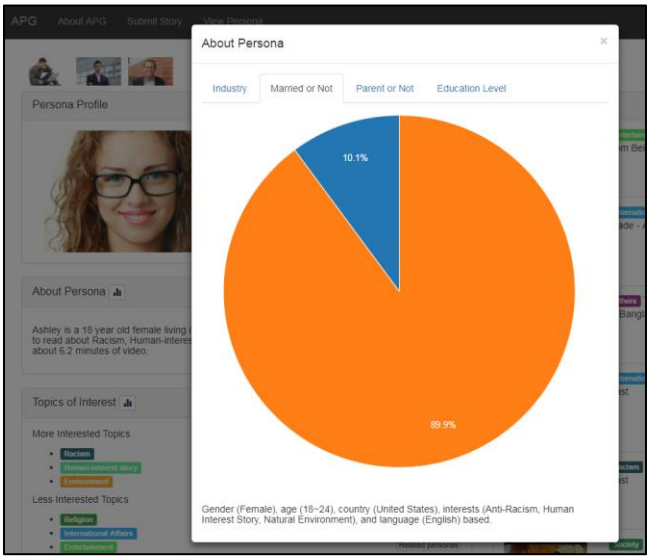

Figure 5: For the About section, clicking on the graph icon displays breakdown on career, marriage, parenting, and education of the persona. In this case, marital status.

3.2.5 Quotes. To provide contextual and personal insights of the persona, social media comments aligned with the topics of interest are displayed in the persona profile (see Fig. 4).

3.2.6 Most Viewed Videos. The persona profile displays the top content most interacted with by this persona. Each content piece is linked to the actual online content (see Fig. 4).

3.2.7 Audience Size. The persona profile also displays the potential reach of the persona, which is the size of the user segment based on demographical attributes and topical interests. The potential reach numbers are generated via a series of Targeting Audiences API calls to the Facebook Marketing (see Fig. 6).

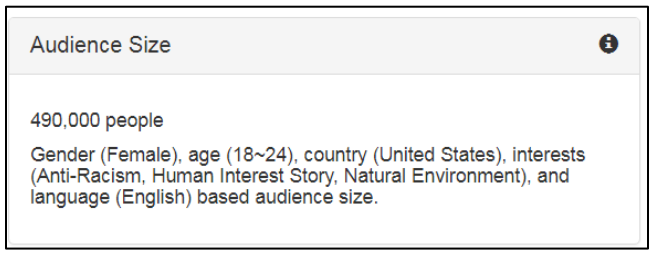

Figure 6: The Audience Size section of the persona profiles displays the number of people that match the persona's attributes, both demographically and behaviorally. This feature helps the end user of APG determine the potential impact of the user segment represented by this persona.

There are a multitude of other features and capabilities that will be presented during the live demonstration, including comparing personas between periods, comparing current user segment size for personas against potential audience size, and displaying personas related to each item of content [9].

4 RESULTS AND DISCUSSION

4.1 Comparison with Other Systems

As far as we know, APG is the first system to use large amounts of online behavioral user data for quantitative and automatic persona generation. Earlier efforts with data-driven personas [10] are more limited, as discussed in [5].

4.2 Commercialization Efforts

Currently, APG is in beta use with three major multi-national corporations, one major non-profit organization, and two SMEs in the retail vertical. Efforts are underway to expand the list of beta clients. The APG project has received funding for development, which will take the system to a Technology Readiness Level 8 (TRL-8). The APG has one evaluation license in place and it participated in an Entrepreneur in Residence (EIR) program, with start-up funding in progress.

5 CONCLUSIONS

In this research, we demonstrate that APG can automatically create personas from large scale, quantitative, aggregated user data from major online social media platforms. The resulting personas are based on real data reflecting real user behavior and can be updated with ease. APG is flexible for application in a wide range of contexts.

REFERENCES

[1] Elina Eriksson, Henrik Artman, and Anna Swartling, "The Secret Life of a Persona: When the Personal Becomes Private," presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France, 2013.

[2] Erin Friess, "Personas and Decision Making in the Design Process: An Ethnographic Case Study," presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, Texas, USA, 2012.

[3] K. Ronkko, "An Empirical Study Demonstrating How Different Design Constraints, Project Organization and Contexts Limited the Utility of Personas," in Proceedings of the 38th Annual Hawaii International Conference on System Sciences, 2005, pp. 220a-220a.

[4] Christopher N. Chapman and Russell P. Milham, "The Personas' New Clothes: Methodological and Practical Arguments against a Popular Method," in Human Factors and Ergonomics Society Annual Meeting, San Francisco, CA, 2006, pp. 634-636.

[5] Jung, S., An, J., Kwak, H. Ahmad, M., Nielsen, L., and Jansen, B. J. (2017) Persona Generation from Aggregated Social Media Data. ACM Conference on Human Factors in Computing Systems 2017 (CHI2017). P. 1748-1755.

[6] An, J., Kwak, H., and Jansen, B. J., (2017) Personas for Content Creators via Decomposed Aggregate Audience Statistics. Advances in Social Network Analysis and Mining (ASONAM 2017). Sydney, Australia, p. 632-635. Australia, 31 Jul – 3 Aug.

[7] Daniel D. Lee and Sebastian H. Seung, "Learning the Parts of Objects by Non-Negative Matrix Factorization," Nature, vol. 401, pp. 788-791, 1999.

[8] Zarrinkalam, F., Kahani, M., Bagheri, E. (2018) Mining user interests over active topics on social networks, Information Processing & Management, 54(2) 339-357.

[9] Salminen, J., Nielsen, L., An, J., Jung, S.G., Kwak, H., and Jansen, B. J. (2018) Is More Better?": Impact of Multiple Photos on Perception of Persona Profiles. ACM CHI Conference on Human Factors in Computing Systems (CHI2018), Montréal, Canada, 21-26 April.

[10] McGinn and N. Kotamraju, "Data-driven persona development," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2008, pp. 1521–1524.