

# Creating Personas to Reuse on Diversified Projects

Andrey Araujo Masiero<sup>1,2</sup>(✉) and Plinio Thomaz Aquino Jr.<sup>1</sup>

<sup>1</sup> Centro Universitário da FEI - Fundação Educacional Inaciana Pe. Sabóia de Medeiros, São Paulo, Brazil

{amasiero,plinio.aquino}@fei.edu.br

<sup>2</sup> Universidade Metodista de São Paulo, São Paulo, Brazil

**Abstract.** This paper presents an automatized creation process for Personas user modeling focus on minimize stereotyping and to increase Persona's reuse on many different projects. This creation process has focus on similarity and automation which, are some main issues of variation from project to project. We discuss this process applying it on two different projects. First is a medical web system (HCI-M) and the second one is a human-robot interaction project with Sony AIBO pet robot (HRI-P). Results show that the process makes possible to minimize the stereotyping and also we reuse Personas from project HCI-M to help us on planning phase of project HRI-P which, turns it practicable.

**Keywords:** QSIM · Clustering · User modeling · Personas

## 1 Introduction

Worries about users during product development are more frequent on design and software teams nowadays [10]. To become possible to develop products focus on users is necessary, first, to comprehend them. For understanding users is required to identify their needs, motivations, objectives, skills, and among others features which, will help us to define the target audience profile [1, 4, 10].

Jung discuss on his work [9] that a person can adopt different personalities according to a certain scenario. This capability Jung calls as Personas. Cooper describes Personas as hypothetical archetypes once it proves that a Persona really represents real users of a launching product [3].

However, user modeling with Personas has as objective to increase communication between project teams. Exchange information is favored due to Persona is like a fictitious character who has name, biographic description and even a picture to illustrate it. In that way, designers and developers can focus on user's features only and saving efforts to create user models using thoughts about how they think that final users are going to be. This is important because avoid bias user models from project team [1].

When designer creates some product and pay attention on needs and behavior of a certain Persona, he will be attending a biggest number of real product users.

Thus, many works present methods to create Personas [7,12,15]. Nevertheless, biggest part of these works are manual and can lead this process with bias or stereotyped according to specialist experience. One way to minimize the bias on Personas creation process is to use clustering algorithms and then to perform statistical data analysis for finding relevant information and most similar users between each other. This paper presents a whole process for creating Personas using a clustering algorithm as a support tool and how to reuse the Personas created on diversified projects.

Clustering algorithm brings some benefits for reuse process and it is main focus of this paper. Some benefits which, can be mentioned here is the automation process and high speed during data mining from project HCI-M to project HRI-P. Another important thing to consider is that on some project the difference between one project to another is the similarity between them. It influences directly the number of personas created on the project.

This paper goes first with the explanation of the whole process for creating Personas and each step of the process explains particularities of two real projects that we applied it. First project was a web system developed for a hospital (Project HCI-M) and the second one is applied for human-robot interaction proposals (Project HRI-P). After that, we discuss some ideas presented during the description of process. In the end, we present conclusions and future works about creation and reuse of Personas.

## 2 Creating Personas

Many projects present methods for creating Personas [7,12,15]. However, the biggest part of these projects present a manual process that can lead to create bias Personas or stereotyped ones, in other words, according to specialist experience. One way to minimize the bias on creation process is the use of clustering algorithms and after this to perform statistical data analysis to find most relevant information and subjects with biggest similarity between each other. That way, a process for creating Personas with six basics steps presented on Fig. 1.

### 2.1 Step 1 - User Information Collection

Many methods are used to collect data from users. Some examples are questionnaires, system event logs, observation, focus groups, and brainstorms, among others [14–16]. Questionnaires are an easy way of declared data collection, using online tools. Users can answer in a comfortable place and pay more attention during the process. However, questionnaire creation process is not easy to execute. Specialist needs to be careful for elaborating questions and answer options to not generate doubts. Another way to collect user data information is to use event log collection. This kind of data collector helps specialist to look through user behavior during system's use and it helps to follow the evolution of user profile over time. In that way, information about user profile are reused on other projects. These two manner to collect user information are the focus on this paper, and it will be detailed as following.

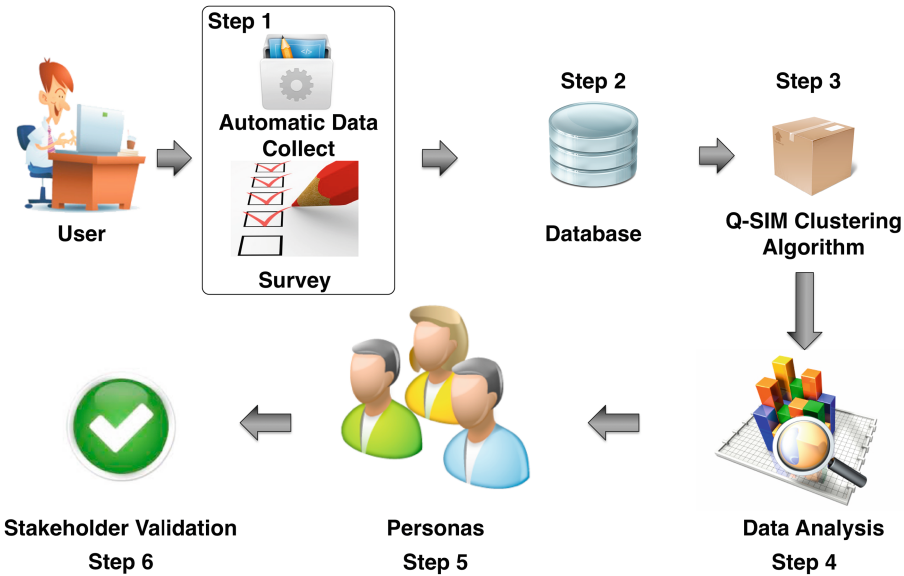


Fig. 1. Personas Creation Process

**Questionnaires.** With questionnaires it is easy to collect information declared by subjects using an online tool. Thus, subjects can answer in places that they think it is better to make this task and they pay more attention on it. Beyond that, it is easy to diversify the population of target audience. This concept can be applied at two different moments of the project. First one is on conception idea, before project begins. The second one is after build the project which, can be used to improve final product or even during developing time to fix designer issues.

To build questionnaires isn't an easy task and it requires attention to elaborate questions and answers once each of those items cannot be ambiguous. Furthermore, it is needed to identify the population who will answer the questionnaire to maximize results.

At project HRI-P we used questionnaires pre-defined by technique Big-Five [6]. As a project on human-robot interaction (HRI) we are interesting on psychological behavior of subject. After some related works studies we realize that Big-Five produces the result needed to create Personas with interaction characteristics. It will help us to develop HRI components for our robots.

**System Event Logs.** The automatic data collection through system's log helps to trace user behavior during the use. Beyond that, it is easy to measure the computational skills of users. An important thing to work with automatic data collection is to determine what kind of system the information will be collected, i.e., desktop system, web system, mobile system, among other ones. It will help to determine what component can be used.

D'Angelo [5] presents a list of variables that could help to capture user skills, scenario complexity and so on for web systems. Burzacca and Paternó [2] present variables applied to mobile systems. Even though work focus on usability tests, these variables can identify users profile. Furthermore, this information is closer to real profile of users than questionnaire answer. The capture automation decrease the subjectivity of user behavioral information and user preferences and it turns possible to follow user computational experience evolution.

Decision of how to collect information depends on project. At project HCI-M we decided to adapt the component presented by D'Angelo [5] using seven variables focus on user skill. It was a web system and we need to trace the evolution of learning of the users. The chosen variables are: (I) Interval of time to fill the fields on a form; (II) Typing speed; (III) Percentage of "backspace" key press; (IV) Amount of errors in form filling; (V) Amount of recurrent errors in form filling; (VI) Usage of double clicks when a single click is expected; and (VII) Unexpected click (or clicking on a not clickable component).

## 2.2 Step 2 - Preparing Information

Algorithms always perform your process based on numerical information. Even if the problem presents categorical or textual information. Categorical or textual information are transformed, in some way, on numerical information during the process to be an input for algorithms. Because of this, it is preferable that such information is stored or captured directly on numerical format. On way to perform it is through human interpretation. Human will translate textual values into numerical codes making possible to keep the fidelity on analyzed data. When the algorithm makes this translation, instead of a specialist, some mistakes can occur on a semantic classification or syntactic of attributes values significance.

With translation executed and it implemented on data collect mechanism, it is necessary to define how it will be stored. First option is the use of a Comma-Separated Values (CSV) file. This file separate each attribute by comma storing them at columns. A second option to store information is to use a database system. To store this kind of information is preferable to use only one table without any normal form, because it will make easy for the read processing of the algorithm.

There are many other kinds of store information like XML and JSON file. For the projects studied here we use CSV file and database system. At project HCI-M we choose database due to integration with the web system was easier than use another way to store this information.

As we use Google Forms for questionnaires on project HRI-P, we decided to use CSV file due to is how the tool store information already. It make easy to adapt QSIM algorithm to receive this kind of entrance for processing.

## 2.3 Step 3 - Performing QSIM

At this moment data collected can be processed by clustering algorithm. The algorithm used to perform is QSIM [13]. A clustering algorithm which, finds the

number of existing classes based on data similarity. By using QSIM on grouping process is interesting to diversify Q value (minimum similarity) to observe group behavior. At both projects were executed this variation of Q value. During the experiments, we noted that some similarity values reproduce a minimal group element exchange. Because of it, we determined a similarity classification to turn communication easier between specialists then to use numerical values. Table 1 presents the classification determined.

**Table 1.** Similarity Intervals

Intervals	Classification
$Q > 0.69$	High Similarity
$0.49 < Q < 0.7$	Moderate Similarity
$Q < 0.5$	Low Similarity

QSIM presents some advantages in grouping process, mainly on user profile information. First one is the quality for keeping similarity inside group. At this point QSIM was the best algorithm [13]. It doesn't need of preview information about how many groups exists, as the classical algorithm k-means [8]. It is necessary to inform only the desire similarity that QSIM finds exactly how many groups exist for this Q value. Results of grouping compared to k-means are similar, however when exists a dense information QSIM makes softer boundaries than k-means [13]. All in all, QSIM spends a higher computational time of processing than k-means if it has a big number of elements on his Related Sets. At this point, k-means has a better performance than QSIM. A version of QSIM is implement in Java Programming Language and available on-line through the link <http://amasiero.github.io/qsim/>.

With creation of groups complete it is important to perform two analysis procedures to obtain a better use of Personas. First step is verify what group has the biggest density. Those groups can help specialist to determine the most significant Personas of the project. Second step is verify the existence of groups that represent the same Persona. In that case, those groups have to become one avoiding duplicity. The second step explanation has more sense after the execution of step 4 of whole process, due to attribute values of each group are defined and we can identify if there are two or more groups with the same attribute values.

## 2.4 Step 4 - Data Analysis

After step 3, groups are determined and the number of generated groups represents how many Personas exist on the project. However, the information that will compose Personas is not ready yet. For each attribute it should finds the measure of central tendency to determine the value for that group. The most common measures are mean, median and mode [12]. Although it is important to

attempt for a rule present by Masiero et al. [11], where the mean is only validate for attributes that has no bias problems. It can be checked through Eq. 1.

$$CV < 0.3 \quad (1)$$

where CV is data coefficient of variation. If condition of Eq. 1 is true then data mean can be used as measure of central tendency, else it is recommended to use median or mode. Through this procedure is possible stipulate a common value that will compose each attribute of group. At this moment, it is good to translate the categorical or textual variable for its original state. Finishing it, Personas should be create on presentation format for the best team communication.

## 2.5 Step 5 - Personas Creation

To finally create Personas in your final state methods from interaction scenarios and problem scenarios are necessary [1]. With these methods, we can add description for all users needs, skills, motivations and objectives that was quantified on earlier steps. Now with Personas ready to use, it is necessary to validate them with project's stakeholders.

## 2.6 Step 6 - Validating Personas Significance

All Personas created during this entering process are presented to project's stakeholders. They will validate if presented Personas are corresponding to the target audience of the project. At this validation process, it could occurs some small adjusts to increase the Persona's description quality. After that, the set of Personas is presented to all team, developers and designers, so they can always focus on these profiles during the project's life-cycle.

## 3 Results and Discussions

This process was applied on two different projects. First one is a medical web system and second one is a human-robot interaction project. As presented at Sect. 2, first step of the process is to determine what variables will be capture by data collect mechanism (manual or automatic one). For the first project, we adapted 7 from 28 variables presented by D'Angelo [5], as discuss on Sect. 2.1. The focus on these variables is user computational skills.

After variables definition and implementation of collect component at the web system, data was collected during test with users. It was recorded 200 records during the tests. Then some similarities variations, homogeneous group was noted with Q value equals to 0.6 in other words 60% of similarity between the subjects. It generate of five Personas. Applying step 4 of the process, it obtained results present on Table 2.

During the process of step 4 we validate that there is no bias for anyone of the variables. Thus, we applied for each variable the data mean to obtain common


**Table 2.** Common information obtained through step 4 of the process.

# Persona	Typing speed (touch/sec)	Interval to fill the fields on a form (sec)	% Backspace	Amount of errors in form filling	Amount of recurrent errors in form filling	Usage of double clicks when a single click is expected
1	1.21	4.17	8.33%	16	2	0
2	1.31	5.49	1.47%	3	1	1
3	1.46	6.41	48.52%	5	1	0
4	0.92	5.94	75%	7	1	0
5	1.11	11.46	23.22%	2	1	0

value of the group, once the rule of Eq. 1 has been attempted, presented on cells of Table 2. The sequence of common values generation for each group, it was necessary to create Personas formatted for team presentation. Table 3 presents one of five Personas obtained during the process.

Dr. John (Persona of Table 3) was created based on information of row 2 from Table 2. It was the only Persona with a value different of zero for attribute “Usage of double clicks when a single click is expected”. It surprised the team because he has satisfy values for the others attributes which, indicates high computational skills. Analyzing all the set of information about this Persona, we realize that

**Table 3.** Persona 2: Dr. John

Picture:	
Name:	Dr. John
Description:	Doctor, 43 yld, he is responsible for hospital’s innovation department. He is a technologic enthusiast and he likes to spend his time in creation of automatize systems for his researches. He looks for technology updates which can bring improvements to the team workflow. He is worried about information security and control access to its information. He spends lots of time developing simple applications with tools like Access and e-mail to keep contact with people. His errors rates are lower than other users.

despite high computational skills people that compose it has experience only on desktop systems which, is necessary to interact with double click frequently.

Following the validation with stakeholders' projects was done and as expected some users were recognized among Personas created. In that way, it is possible to affirm that the application of the process was successful.

The main objective of second project is to identify the Personas behavior to serve as a guideline for developing new social robots mechanism. Although, Personas created for project HRI-P are behavioral and we use then to create the reception robots interaction for the hospital of project HCI-M. These robots will substitute the team of public service support which, gives out passwords and information to the public. The data collection was made with pre and post questionnaires created focus on information of Personas created on the first project. The answers were quantified from process of identification of people behavior profile called Big Five (see Sect. 2.1). This process quantifies user answers into a numerical degree of intensity for each Big Five attribute. Thus, it is intuitive to identify behavior profile of the user. Table 4 presents the common values for Big Five attributes after step 4 analysis process.

**Table 4.** Common information obtained through step 4 of the process for project HRI-P.

#	Age	Gender	Extroversion	Agreeableness	Conscientiousness	Neuroticism	Openness
1	7	Female	5.0	4.5	5.0	4.5	6.0
2	11	Male	5.0	4.5	4.0	4.5	4.5
3	18	Male	4.5	5.0	5.5	4.0	5.5
4	23	Female	5.0	5.0	5.5	5.0	5.0
5	41	Male	5.0	4.5	6.0	3.5	6.5

The values presented on Table 4 help us to create the description of Personas. How it contains psychological information, the description can be detailed with behavioral data, like openness for new experience as row 5. This information will help us to identify better information for interact with this kind of Persona. To create Personas based on the kind of Table 4 helps to minimize Persona's stereotyping once team project has real data to support description creation and they do not need to think how is target audience of the project.

This human-robot interaction project used a Q value equals to 0.8 which means 80% of similarity desired. High similarity can best split groups with small samples or data record. In that case, validation process occurs a little bit different, because the team is your own client. After creation of Personas were analyzed videos of each test and so it was possible to realize that, some subjects are similar to some Personas, as earlier project.

In that way, that the proposed process by this work supports an automatize creation of Personas for any kind of project and that Personas created through this process can be reused in another projects decreasing the cost and analyses



time of users at a first step. Also, Personas create in future projects can help to improve older projects like happen on project HCI-M and project HRI-P. Beyond that, the process tends to minimize the Persona's stereotyping which, before are only based on information acquired by team experience.

## 4 Conclusion and Future Works

The process presented at this work allows that Personas can be created in an automatized way and it minimize stereotyping of Personas. This is an important point because it formalize that user-centered projects effectively attempt objectives and needs of target audience.

Another important point that may be highlighted here is that QSIM algorithm makes it easy due to it finds the number of existent groups in database keeping the similarity quality inside each group, proved on [13]. The proposed cycle implementation in many projects of the same company or even the same market segment allow knowing the users of new products or service before to start the project. All of it through the user modeling based on Personas.

It is also possible to follow Personas evolution all time long. It can determine a cycle of life as the user they represent. Some preliminary tests show that is possible to realize some intersection between Personas and projects which, makes reuse of Personas and also to create a Personas repository practicable. However, more tests are necessary to complete this task fully. This extra works are developing and we will publish it soon.

**Acknowledgment.** To CAPES PROSUP Scholarship and to FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) for financial support.

## References

1. Aquino Junior, P.T., Filgueiras, L.V.L.: A expressão da diversidade de usuários no projeto de interação com padrões e personas. In: Proceedings of the VIII Brazilian Symposium on Human Factors in Computing Systems. IHC 2008, Sociedade Brasileira de Computação, pp. pp. 1–10, Porto Alegre, Brazil (2008). <http://dl.acm.org/citation.cfm?id=1497470.1497472>
2. Burzacca, P., Paternò, F.: Remote usability evaluation of mobile web applications. In: Kurosu, M. (ed.) HCII/HCI 2013, Part I. LNCS, vol. 8004, pp. 241–248. Springer, Heidelberg (2013)
3. Cooper, A., Reimann, R., Cronin, D.: About face 3: the essentials of interaction design. Wiley, India (2007)
4. Cooper, A.: The Inmates are Running the Asylum. Macmillan Publishing Co., Indianapolis, IN, USA (1999)
5. D'Angelo, F.d.M.: Identificação automática de perfis de grupos de usuários de interfaces web (2012). biblioteca Digital de Teses e Dissertações da FEI. Disponível em: <http://tede.feui.edu.br/tede/tde-busca/arquivo.php?codArquivo=231/>
6. Gosling, S.D., Rentfrow, P.J., Swann, W.B.: A very brief measure of the big-five personality domains. *J. Res. Pers.* **37**, 504–528 (2003)

7. Guimarães, D.B., Carvalho, C.R.M., Furtado, E.S.: Panorama, oportunidades e recomendações para o contexto brasileiro de interação humano-computador e design centrado no usuário a partir do uso de personas. In: Proceedings of the 10th Brazilian Symposium on on Human Factors in Computing Systems and the 5th Latin American Conference on Human-Computer Interaction. IHC+CLIHC 2011, pp. 167–176. Brazilian Computer Society, Porto Alegre, Brazil (2011). <http://dl.acm.org/citation.cfm?id=2254436.2254467>
8. Jain, A.K.: Data clustering: 50 years beyond k-means. *Pattern Recogn. Lett.* **31**(8), 651–666 (2010)
9. Jung, C.: *The archetypes and the collective unconscious*, vol. 2. Bollingen (1975)
10. Masiero, A.A.: *Algoritmo de agrupamento por similaridade aplicado a criação de personas* (2013). biblioteca Digital de Teses e Dissertações da FEI
11. Masiero, A.A., de Carvalho Destro, R., Curioni, O.A., Junior, P.T.A.: Automapersona: a process to extract knowledge automatic for improving personas. In: Stephanidis, C. (ed.) *HCI International 2013 - Posters' Extended Abstracts*, pp. 61–64. Springer, Heidelberg (2013)
12. Masiero, A.A., Leite, M.G., Filgueiras, L.V.L., Aquino, Junior, P.T.: Multidirectional knowledge extraction process for creating behavioral personas. In: *Proceedings of the 10th Brazilian Symposium on on Human Factors in Computing Systems and the 5th Latin American Conference on Human-Computer Interaction, IHC+CLIHC 2011*, pp. 91–99. Brazilian Computer Society, Porto Alegre, Brazil (2011). <http://dl.acm.org/citation.cfm?id=2254436.2254454>
13. Masiero, A.A., Tonidandel, F., Aquino Junior, P.T.: Similar or not similar: this is a parameter question. In: Yamamoto, S. (ed.) *HCI 2013, Part I. LNCS*, vol. 8016, pp. 484–493. Springer, Heidelberg (2013)
14. Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., Carey, T.: *Human-computer interaction*. Addison-Wesley Longman Ltd., Boston, MA (1994)
15. Pruitt, J., Adlin, T.: *The Persona Lifecycle: Keeping People in Mind Throughout Product Design*. Morgan Kaufmann Publishers, San Francisco (2005)
16. Rogers, Y., Sharp, H., Preece, J.: *Interaction design: beyond human-computer interaction*. Wiley & Sons, New York (2011)