

What's in a Persona? A Preliminary Taxonomy from Persona Use in Requirements Engineering

Devi Karolita^{1,2}^a, John Grundy¹^b, Tanjila Kanij¹^c, Humphrey Obie¹^d and Jennifer McIntosh³^e

¹*Department of Software Systems and Cybersecurity, Faculty of Information Technology, Monash University, Melbourne, Australia*

²*Department Informatics Engineering, Faculty of Engineering, Palangka Raya University, Palangka Raya, Indonesia*

³*Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Melbourne, Australia*

Keywords: Persona, Requirements Engineering, Curation, Taxonomy, Analysis.

Abstract: Personas have been widely used during requirements engineering-related tasks. However, the presentation, composition, level of details and other characteristics varies greatly by domain of use. To better understand these, we formed a curated set of nearly 100 personas from 41 academic papers and analysed their similarities and differences. We then used our analysis to formulate a preliminary taxonomy of personas used for Requirements Engineering-related tasks. We describe our key findings from our analysis with examples, our preliminary taxonomy, and discuss ways the taxonomy can be used and further improved.

1 INTRODUCTION


A 'persona' is a fictitious character that summarises the major traits of the actual end users of a proposed product (Kolski and Warin, 2018). First introduced for use in Software Engineering (SE) by Alan Cooper (Cooper, 1999), personas have gradually become a resource to better understand end users, especially during Requirements Engineering (RE) related activities. Many activities in RE aim to elicit, document and analyse end users requirements, as well as to validate these captured requirements (Sommerville, 2016). The main purpose of these activities is to ensure end users' needs for the proposed software product are well understood and eventually met.


In RE related tasks, having personas as a representation of target software end users enables requirements engineers to identify user requirements (Ho and Lin, 2019) (Cleland-Huang et al., 2013) (LeRouge and Ma, 2010) (Sim and Brouse, 2015) and predict how the end users might possibly interact with the proposed software product (Bowles, 2007). Personas can be used as the basis to define the proposed product


requirements (Sim and Brouse, 2014) which then support functional and non-functional requirements specification (Nunes Rodrigues et al., 2018). Furthermore, personas can help requirements engineers discover redundant requirements (Sim and Brouse, 2014) and convey potential issues from the specified requirements (Aoyama, 2005) (Lopez-Lorca et al., 2014) (Abd Malik and Azuddin, 2013). Personas are not a replacement for end users during RE, but compliment use of focus groups, surveys and interviews. Personas allow a wide range of different target end user characteristics to be reasoned about, especially when it is difficult to directly meet with many real end users.


There is no standard format to present personas used in RE. Normally, personas contain context specific depictions of target software end users, including their way of thinking, behaviour, goals, and motivations. A study argued that persona attributes can be grouped into three categories: identical attributes, aggregate attributes, and cosmetic attributes (Alvertis et al., 2016). Identical attributes are persona characteristics that will remain the same for each user group, while the aggregate attributes are the aggregation of the user attributes (e.g., comfort with technology of a particular age group), and cosmetic attributes identify the persona (e.g., name, photograph).


A few previous studies investigated the information contained in personas and dissected persona information into sections or layers. Nielsen et al. anal-

^a <https://orcid.org/0000-0001-6908-9785>

^b <https://orcid.org/0000-0003-4928-7076>

^c <https://orcid.org/0000-0002-5293-1718>

^d <https://orcid.org/0000-0002-6322-2984>

^e <https://orcid.org/0000-0002-6655-0940>

ysed existing persona templates, grouped the information contained in those templates into key categories, and compared the existing templates to 47 collected personas (Nielsen et al., 2015). Salminen et al. also investigated persona contents in order to analyse information presented in quantitative personas (Salminen et al., 2020). However, these studies do not specifically focus on the use of personas in the RE process. Moreover, the scope of these studies were limited to a specific geographic location and to personas resulting from quantitative approaches. Persona descriptions have been divided into context-free dimensions and context-dependent dimensions (Antle, 2008). A study introduced a concept of basic persona that can be quantitatively generated from a large scale data (Wöckl et al., 2012). Furthermore, persona can be decomposed into two layers: basic layer and external layer. The aim of the decomposition of persona descriptions is to make personas more reusable (Marcengo et al., 2009). Nevertheless, a more contextual layer of persona is required to enable persona being relevant for use in different contexts and domains. The missing layer needs to be equipped with a practical recommendation to help requirements engineers to have some ideas on what to be included in the persona descriptions.

To address these limitations, in this work we searched for a range of persona usages in RE-related activities. We curated 98 collected personas from 41 relevant academic publications, which we refer to as our curated *Persona Corpus*. During the curation process, we analysed what key information is found in these personas. As the result of our analysis, we identified 12 key domains in which personas in our *Persona Corpus* were used. We identified three key persona dimensions: the way a persona can be narrated, its formatting, and the length and detail of persona descriptions. We also identified five major demographic information included in persona descriptions: name, choice of photograph, gender, age, and tagline. We identified key human factors presented in each persona, and grouped them into persona facets, and we mapped the use of these facets to each identified domain. The result of our mapping was used as the basis to develop preliminary *RE-based persona taxonomy*. By acknowledging the fact that persona is a context-specific tool, we also formulate a domain-based facets recommendation that can be used to create better personas to be utilised in a specific domain. The key contributions of this work include:

- We curated a set of 98 personas from 41 academic publications describing use of personas for various RE-related tasks;

- We analysed these personas from several perspectives to understand their similarities and differences, including domain of use, RE task, presentation, level of detail, and key factors;
- We developed a preliminary taxonomy of personas aimed to provide practical support RE-related activities; and
- We propose a set of further research directions to apply and improve our taxonomy.

The rest of this paper is organised as follows. Section 2 presents our motivation for this study, and Section 3 presents key related work. Section 4 details our research methods, and Section 5 presents the findings resulting from our persona curation and analysis. Section 6 presents a preliminary persona taxonomy, Section 7 summarises our findings, study limitations, and presents opportunities for future work, and Section 8 concludes our research.

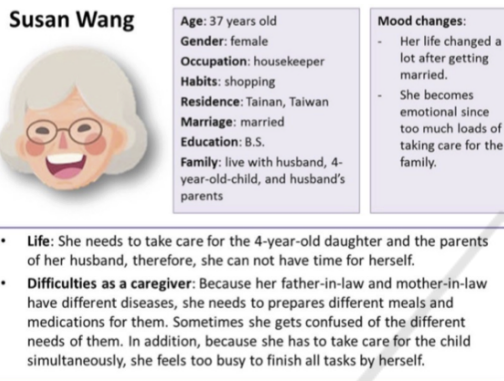
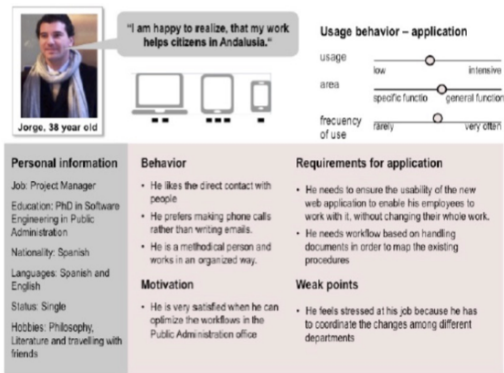
2 MOTIVATION

As part of our work in investigating the use of personas in Requirements Engineering (RE), we reviewed a large number of papers describing personas used in RE-related activities. After reviewing these personas, we noticed that many of these personas showed noticeable differences. Consider those in Figure 1 (our full persona list can be found in our online Appendix¹). This shows three very different personas used for different RE domains and tasks. All of these personas were used in requirements elicitation and analysis phase.

The persona on the top was used in Software Development domain to introduce a concept of *Context-based Persona Stories* (Sedeno et al., 2017). The persona was presented using bullet points and formatted according a particular structure. The persona in the centre was presented in a semi-structured format, and was used to identify the needs of older adults (i.e., medical services, meal preparation, and daily needs) in order to design assisted living technologies (Ho and Lin, 2019). Meanwhile, the persona on the bottom was presented in a narrative form without any particular structure for its format. This persona was used in the Education domain to develop an educational mobile app (Askarbekuly et al., 2021).

We noticed a number of differences in the way such personas were presented, including the way the personas were narrated, the use of visual representations (i.e., a photograph of a real person, a cartoon

¹<https://doi.org/10.5281/zenodo.7312341>



Jamal is male and 25 years old. He comes from a historically Muslim culture, with traditionally strong family ties. Currently, he is staying away from the family and relatives due to having studies and job. He is an active user of educational mobile apps, such as Duolingo, watches YouTube habitually. Jamal wants to keep in touch with his family and friends, and feels that it's important. However, he has various reasons preventing him from doing so such as being busy, forgetful, and having other priorities such as work and studies.

Figure 1: Bullet points persona (top) (Sedeno et al., 2017); Semi-structured persona (centre) (Ho and Lin, 2019) ((c) Springer, reused with permission); and Brief persona (bottom) (Askarbekuly et al., 2021) ((c) Springer, reused with permission).

picture), the format used to present the personas, the different human factors presented in the persona descriptions, and the level of detail in different personas.

These differences in the personas used to support RE-related work motivated us to collect more personas and carry out an analysis of their similarities and differences. We wanted to understand how persona contents, representation, level of detail, facts, and other information may vary, including the problems that were encountered when developing personas and how these were evaluated and addressed. To do this, we collected a large number of personas used for RE-related tasks from a wide range of publications. This collection of curated personas forms a *Persona Corpus* that we analysed and make available for other researchers to use in their own work. To carry out this work, we ask the following key research questions:

RQ1. What are the different domains that personas have been used to aid requirements engineering? Using our Persona Corpus, we want to identify the domains where personas have been used in to date.

RQ2. What can be found in personas used in each domain? This research question focuses on discovering what persona facts exist, their similarities and differences across different domains. We wanted to identify the different persona facets and group these facets based in their similarities.

RQ3. How can we build a preliminary persona taxonomy to aid requirements engineers? Based on the identified persona facets and domains of use from RQ1 and RQ2, we wanted to formulate a preliminary persona taxonomy and domain-based persona facets recommendation for RE related-tasks. The taxonomy aims to help persona creators and users in generating reusable personas. Our domain-based persona facets recommendation can help customise generic personas for use in specific domains.

3 RELATED WORK

Nielsen et al. analysed personas used in industry setting (Nielsen et al., 2015). A literature study was used to analyse 12 existing persona templates and grouped the information presented in the templates into five major categories: background information, design related information, business and marketing related information, graphics, and miscellaneous. The researchers then compared the existing templates to persona descriptions of 47 personas used in 13 Danish industries and organisations. The results show that a Danish persona style is different from recommendations given by the existing templates analysed in the study, particularly in business and marketing. Salminen et al. conducted a study to create a template for data-driven personas (Salminen et al., 2020). The study focused on analysing data-driven user personas. The researchers extracted information from 31 personas created using quantitative techniques. This study categorises quantitative personas richness into three level: simple, moderate, and high. The study also shows the differences of the information contained in quantitative personas and the personas created using mixed methods. The study argues that quantitative methods resulting in a chart-like presentation which display quantitative data. On the other hand, personas resulting from mixed methods have a more contextual and narrative-like descriptions.

The two studies above do not specifically focus on personas used in the RE related tasks. In addition,

tion, those studies focus in personas used in specific geographic location and personas generated by a particular approach. This provides an opportunity for further investigation on persona contents which can contribute important information for researchers and requirements engineers, particularly with respect to human-centred RE.

Some previous works investigated how personas can be decomposed to aid reusability. The idea is to streamline the process of persona creation. The division can be performed by separating persona attributes that are static (or have less likelihood to change) that can be reused in different contexts (Antle, 2008) (Moser et al., 2011). A study that proposed Child-Persona technique argued that a persona description be divided into context-free dimensions and a context-dependent one (Antle, 2008). Context-free dimensions of persona consist of data concluded from theoretical understanding, while a context-dependent dimension consists of data specific to the project. A concept of *basic persona* was proposed with the purpose to streamline persona development process and enhance reusability (Wöckl et al., 2012). Using data retrieved from a survey on European older adults, the researchers generated 30 personas representing the elderly. The study discovered some major clusters variables of persona description, including *self-perceived general health*, *self-reported limitations with Activities of Daily Living*, *cognitive function*, *make use of home care services*, *economic situation*, and *social activities*. In addition, the study also proposed additional variables to elaborate each cluster to give more detail information. The resultant personas were claimed to be applicable in different projects and were extendable to specific contexts. Marcengo et al. decomposed persona into two major sections: the basic persona and the external layer (Marcengo et al., 2009). The basic persona serves as the base of a persona that consists of information that have less likelihood to change through different contexts. Therefore, this layer of persona is can be reusable for different contexts and domains. The external layer of a persona consists of context-specific information which enables the persona to be more relevant to a particular project or context.

Two of these studies ((Antle, 2008) and (Wöckl et al., 2012)) were designed for a particular group of age which resulting to a limited applicability. While the persona layering framework was argued being able to allow persona creators to develop reusable personas (Marcengo et al., 2009), however it has to be equipped with a practical recommendation that can be used to formulate a more contextual external layer of personas.

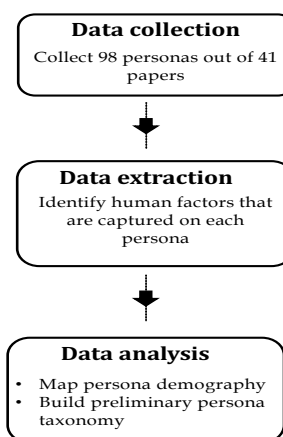


Figure 2: Research method.

4 METHOD

In order to achieve the aim of this research, we: (i) curated a set of personas collected from 41 publications; (ii) extracted key human factors captured in each persona; (iii) analysed the personas to understand key persona demography; and (iv) built a preliminary persona taxonomy. Figure 2 summarises the process undertaken. The following subsections detail our research process.

4.1 Data Collection

We performed academic literature database searching to collect our data. The search string used in the search can be seen in Table 1. The search was also limited to publications published between January 2000 and December 2021. The search was then conducted over six databases: ACM Digital Library (ACM), SpringerLink, IEEE Xplore (IEEE), Engineering Village, Wiley Online Library, and Taylor & Francis Online.

There were a total of 833 publications returned from all databases. We filtered out the returned publications by reading the title and abstract. Duplicated publications were also removed, resulting in 248 publications for final selection.

For our final selection, we focused only on publications that provided concrete persona examples. Consequently, we found 41 publications from which we collected 98 persona examples.

4.2 Data Extraction and Synthesis

From 41 selected publications (referred to as PC01 to PC41), we collected 98 personas (referred to as Per01

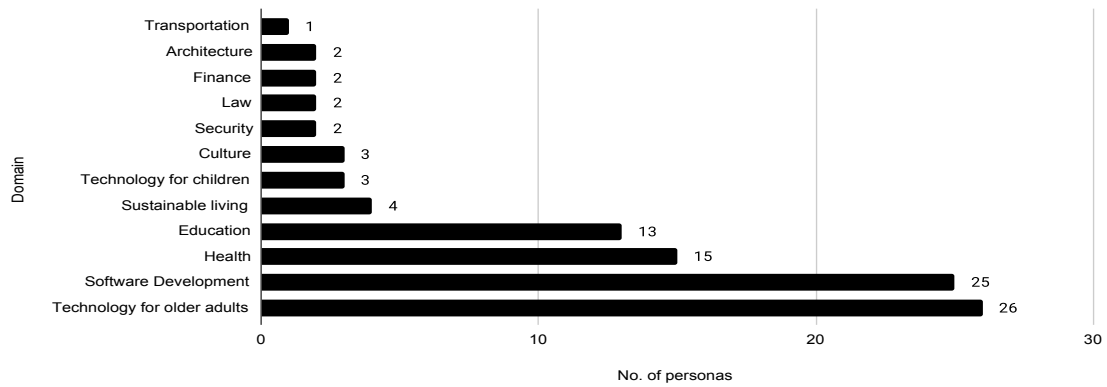


Figure 3: Different Domains Personas Used For.

Table 1: Search string.

personas	AND	“Requirements Engineering” OR “Requirements Engineering Process” OR “Requirements Elicitation” OR “Requirements Specification” OR “Requirements Analysis” OR “Requirements Gathering” OR “Requirements Identification” and “Requirements Validation”
----------	-----	--

to Per98); listed in our online Appendix ². These personas were grouped by the domain they were used in. Our next step was to identify how the personas were presented e.g. text, table, graphic, etc, and how they were constructed e.g. from the focus group, interview, from existing persona, etc. We then carefully read each persona and extracted the human factors from each persona, including both those factors that are explicitly and implicitly described in the persona. To avoid confusion during human factors identification, we used the Merriam-Webster dictionary to define some of the terms used during the extraction process.

5 CURATED PERSONAS

5.1 Domains of Use

Forty-one publications were included in our review from which we collected 98 concrete personas into our Persona Corpus. We identified 12 domains where the personas were used in and grouped the Persona Corpus by the domain. These domains are: technology for older adults, software development, health (physical and mental), education, sustainable living, culture, technology for children, architecture, finance, law, security, and transportation. Figure 3 summarises the total number of domains personas applied to.

²<https://doi.org/10.5281/zenodo.7312341>

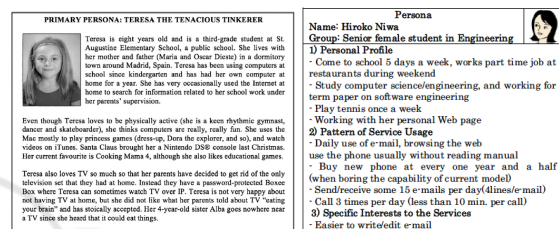


Figure 4: Narrative persona (left) (Acuna et al., 2012) ((c) Elsevier, reused with permission) and Bullet points persona (right) (Aoyama, 2005) ((c) IEEE, reused with permission).

5.2 Persona Dimensions

We focused on reviewing only text-based personas i.e. those using text to describe factors that make up the person. The main reason is that most of the collected personas in our Persona Corpus are text-based. We discovered multiple ways, or dimensions, that text-based personas can be presented in. Based on our Persona Corpus analysis, we identified three key dimensions of text-based persona representation: **persona narration**, **persona format**, and **persona length**. An overview of these key text-based persona dimensions can be seen in Table 2.

We term **persona narration** as the way persona descriptions can be narrated. Text-based personas can be described either in a **narrative** fashion or in a straightforward manner using **bullet points**. Examples are shown in Figure 4. Narrative personas are typically written in a story-like flow, narrating the characteristics of the personas from their general background information (e.g., name, age, personality) to context-specific aspects (e.g., interaction with technology, life achievement, social interaction). Good examples of narrative personas are [Per01], [Per25], and [Per36].

On the other hand, bullet-point based personas are more straightforward compared to narrative personas in terms of presenting the key persona attributes

Table 2: Overview of text-based personas in Persona Corpus.

Persona narration	Narrative approach	Per01, Per02, Per03, Per04, Per05, Per06, Per07, Per08, Per09, Per12, Per13, Per14, Per15, Per16, Per17, Per18, Per19, Per25, Per26, Per27, Per28, Per29, Per30, Per31, Per32, Per33, Per34, Per35, Per36, Per37, Per38, Per39, Per40, Per41, Per42, Per43, Per44, Per47, Per48, Per49, Per50, Per51, Per52, Per55, Per56, Per57, Per58, Per59, Per60, Per61, Per62, Per63, Per64, Per65, Per66, Per67, Per68, Per69, Per70, Per71, Per72, Per73, Per74, Per75, Per76, Per77, Per78, Per79, Per80, Per81, Per82, Per83, Per84, Per85, Per86, Per87, Per88, Per89, Per90, Per91, Per92, Per93, Per94, Per95, Per96, Per97, Per98	87
	Bullet points	Per10, Per11, Per20, Per21, Per22, Per23, Per24, Per45, Per46, Per53, Per54	11
Persona format	Unstructured	Per01, Per02, Per15, Per16, Per17, Per18, Per19, Per20, Per21, Per22, Per23, Per25, Per26, Per27, Per28, Per29, Per30, Per31, Per32, Per36, Per37, Per38, Per49, Per50, Per51, Per52, Per53, Per54, Per70, Per75, Per76, Per77, Per78, Per79, Per80, Per81, Per82, Per83, Per84, Per85, Per86, Per87, Per88, Per89, Per90, Per92, Per93	47
	Semi-structured	Per03, Per04, Per05, Per06, Per07, Per08, Per09, Per10, Per11, Per12, Per13, Per14, Per33, Per34, Per35, Per39, Per40, Per41, Per42, Per43, Per44, Per47, Per48, Per55, Per56, Per63, Per64, Per71, Per72, Per73, Per74, Per91, Per94, Per95, Per96, Per97, Per98	37
	Structured	Per24, Per45, Per46, Per57, Per58, Per59, Per60, Per61, Per62, Per65, Per66, Per67, Per68, Per69	14
Persona length	Normal	Per01, Per02, Per03, Per04, Per05, Per06, Per07, Per08, Per09, Per10, Per11, Per12, Per13, Per14, Per15, Per16, Per17, Per18, Per19, Per21, Per24, Per25, Per26, Per28, Per29, Per33, Per34, Per35, Per39, Per40, Per41, Per42, Per43, Per45, Per46, Per52, Per55, Per56, Per57, Per58, Per59, Per60, Per61, Per62, Per63, Per64, Per65, Per66, Per67, Per71, Per72, Per91, Per97, Per98	54
	Brief	Per20, Per22, Per23, Per27, Per30, Per31, Per32, Per36, Per37, Per38, Per44, Per47, Per48, Per49, Per50, Per51, Per53, Per54, Per68, Per69, Per70, Per73, Per74, Per75, Per76, Per77, Per78, Per79, Per80, Per81, Per82, Per83, Per84, Per85, Per86, Per87, Per88, Per89, Per90, Per92, Per93, Per94, Per95, Per96	44

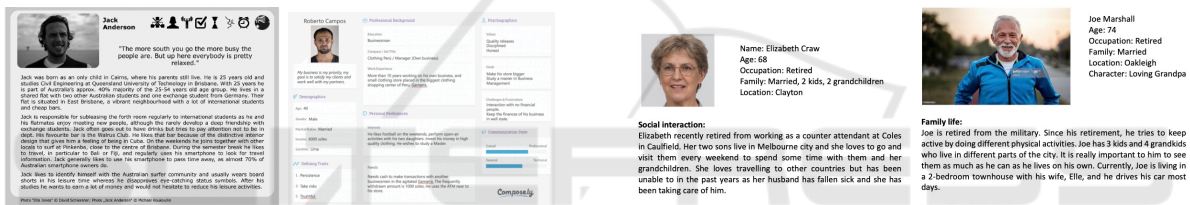


Figure 5: Unstructured persona (left) (Lachner et al., 2015) ((c) Springer, reused with permission), and Structured persona (right) (Aguirre et al., 2021) ((c) Springer, reused with permission).

(i.e., demographic information, general background, context-specific information). Good examples are presented in [Per10], [Per11], and [Per45]. Interestingly, we found one persona ([Per52]) presented using both approaches. Demographic information and general personal stories of the persona were presented in a narrative manner, while context-specific persona stories was presented using bullet points.

Another dimension we identified is persona description format. We found that there are three main ways used to format text-based persona descriptions: **unstructured**, **semi-structured**, and **structured**. Narrating a persona in an unstructured manner means that the persona is described without any binding structure (and/or order) to present the persona attributes. Some good examples of this type of persona are [Per26], [Per30], [Per89]. Two examples are shown in Figure 5.

Text-based personas can be narrated in a semi-structured fashion. The persona attributes are grouped based on their similarities; such as demographic-



Figure 6: Semi-structured persona examples.

related attributes (e.g., name, age, gender, marital status), skills, social interaction, and computer experience. There is no binding rule in grouping the persona attributes, as we observed that even personas used in the same project can have different groupings (see Figure 6). As examples, [Per39] and [Per40] were used in the same case study and had different sections in their persona descriptions. [Per39] had *ICT usage*, *Relation to grandchild(ren)*, *Goals*, *Frustration and pain point*, and *Primary usage reasons*. Whereas, [Per40] had sections for *Social contacts*, *Interest and experience in new communication technologies*, and *Requirements*.

For structured personas, the descriptions are narrated by following a particular format defined by the persona creators (see Figure 5). In [Per65] and [Per66], the personas had sections presenting demographics, defining traits, professional background,



Figure 7: Examples of normal-length persona (left) and brief persona (right).

personal preferences, psychographics, and communication style.

The last dimension we defined for our curated text-based personas is **persona length**. From our Persona Corpus review, we found that text-based personas can be narrated either in **'normal length'** (more than ten sentences) or **'briefly narrated'** (ten sentences or less); as can be seen in Figure 7. Normal-length personas can be seen in [Per17], [Per64], and [Per97]. Some examples of brief personas are [Per47], [Per70], and [Per96]. Interestingly, there are case studies that incorporated both normal-length and brief personas, as mentioned in [PC8], [PC11], [PC24], and [PC31].

After reviewing our Persona Corpus on how the personas are narrated, we found that the majority of text-based personas in our Corpus are presented in narrative fashion (87 personas), while only 11 personas are presented using bullet points. We also found almost equal number of unstructured and semi-structured personas (47 and 37 personas respectively), whereas there are 14 structured personas. We also discovered that there were 54 normal length personas and 44 briefly written ones.

We found there were three methods to create the personas: qualitative, quantitative, and mixed methods. There are 56 personas that were qualitatively created, 26 personas that were created using mixed methods, and 7 personas were created in a quantitative manner. The description of **qualitative personas** are mostly presented in a *narrative form* (51 personas), described in *no particular structure* (39 personas), and *briefly written* (37 personas). As for the personas created using **mixed methods**, 25 personas were narrated in a *narrative* manner, 16 personas were formatted in a *semi-structured* fashion, and 21 personas have *normal length*. Moreover, from the personas that were **quantitatively generated**, there are 5 personas that were described *narratively*. Three personas were formatted in either *unstructured* or *semi-structured manner*, and 5 personas have *normal length*.

5.3 Demographic Information in Personas

We identified key demographic information presented in persona descriptions. These include name, visual representation, gender, age, and tagline. Table 3 shows the frequency of occurrence of the information for each domain that our curated personas are used for.

Name is the most common human factor in persona descriptions and is normally used in personas in ten domains. Seventy-one personas have only first name and 21 personas have full name. Only six personas (used in domain *education* and *sustainable living*) in our Persona Corpus do not have a name. Instead, they are labeled with a code to distinguish them one from another.

There are 82 personas in our Corpus that include an **age** in their description, especially personas used in age-related domain (*technology for older adults* and *technology for children*). People aged **25-64 years** are the most presented population in our Persona Corpus (36 personas), followed by people aged 65 years and over (29 personas). There are some personas used in four domains that do not mention any age in their descriptions. Interestingly, personas used in the *security* domain do not have information about age.

Based on our analysis, we found that either **photograph** (real person) or **picture** (cartoon-like image) are used to provide a visual representation of a persona. In our Corpus, there are 53 personas that included photograph and ten personas used picture. All personas used in domain of *sustainable living*, *law*, and *transportation* do not provide any visual representation. In total, there are 35 personas that do not include photograph or picture.

In regard to **gender** representation, 54 personas do not include any gender-related information, particularly personas used in domain of *technology for children*, *security*, *law*, and *transportation*. For personas that presented gender information, there are 25 personas representing **female** population and 19 personas representing **male** population. No non-binary personas were found. We also found some personas do not explicitly mention a gender. Persona gender is usually presented using gender-related terms, such as "72-year-old woman".

We also observed that some personas included a **tagline** that summarises persona characteristics in one-line statements. This approach is argued can enhance the memorability of the persona (Björndal et al.,). In total, there are 40 personas that included a one-line statement in their description. Based on our review, *software development* is a domain in which

Table 3: General human factors in our Persona Corpus.

Domain	Name			Visual representation			Gender			Age					Quote/One-liner		
	Fullname	Firstname	NA	Photograph	Picture	NA	Male	Female	NA	0-14 years	15-24 years	25-64 years	65 years and over	NA	One-liner (describing statement)	Personal statement	NA
Technology for older adults	4	22	0	9	2	15	5	9	12	0	0	3	23	0	3	0	23
Software Development	5	20	0	23	2	0	7	7	11	1	4	16	1	3	10	9	6
Health	2	13	0	7	4	4	1	3	11	0	3	6	3	3	4	1	11
Education	2	8	3	2	2	9	2	1	10	2	0	3	0	8	3	2	8
Sustainable living	1	0	3	0	0	4	2	2	0	1	1	1	1	0	0	0	4
Technology for children	0	3	0	3	0	0	0	0	3	3	0	0	0	0	0	0	3
Culture	2	1	0	3	0	0	1	0	2	0	1	2	0	0	0	2	1
Security	0	2	0	2	0	0	0	0	2	0	0	0	0	2	2	0	0
Law	0	2	0	0	0	2	0	0	2	0	0	2	0	0	0	0	2
Finance	2	0	0	2	0	0	1	1	0	0	1	1	0	0	0	2	0
Architecture	2	0	0	2	0	0	0	2	0	0	0	2	0	0	0	2	0
Transportation	1	0	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0
Total	21	71	6	53	10	35	19	25	54	7	10	36	29	16	22	19	58

the most personas included tagline (19 personas).

Twenty-two personas used a third-person perspective (referred to as **one-liner (describing statement)**) to succinctly describe the persona, such as 'Sedentary old person', 'Passive and stingy', and 'Insurance seeker'. In addition, there are 19 personas that concisely described persona characteristics using a first-person manner (referred to as **personal statement**). Some of the examples are "The main thing is that I arrive punctually at the destination", "Between work and college, I always need cash to pay a thousand thing", and "I only drive the car if I have to".

6 Preliminary Persona Taxonomy

6.1 Persona Human Factors

We wanted to develop a preliminary persona taxonomy that can divide persona descriptions into two layers: (1) generic information; and (2) context-specific information.

Firstly, we identified a number of human factors reflected in persona descriptions from our Persona Corpus analysis. We then grouped these identified human factors into persona facets on the basis of their similarities. Those facets then were divided into a more general set of groups based on a preliminary taxonomy of human aspects introduced by Grundy et al. (Grundy et al., 2022). Our grouping can be seen in Table 4.

We then categorised the identified persona attributes into three human aspect groups: (1) Personal characteristics; (2) Skill, experiential or environmental-influenced characteristics; and (3) Group or multiple human characteristics.

Human facets that fall into **personal characteristics** group are *demographic information* (age, name, gender), *personal attributes* (attitude, behaviour, personality, preference, interest, hobby), *physical well-being* (health challenge, health status,

body measurement), and *mental well-being* (mental challenge, emotional feeling). In **skill, experiential or environmental-influenced characteristics** group, we included *personal story* (activity, achievement, memorable incident, life experience, life value), *interaction with technology* (ICT usage, ICT literacy, adaptation to technology, possessions of gadgets, wearable device usage), *skill level* (skill, health literacy), *education* (education, learning experience), *environmental-influenced characteristics* (spoken language), *human values* (life value, family tradition, religious belief), and *socio-economic status* (financial situation). Under **group or multiple human characteristics** group, we put *work status* (occupation, income), *family environment* (living arrangement, family structure, parental intervention), *geographic location* (current location), *collaboration and communication style* (work experience, social interaction, complain experience), and *culture* (cultural suitability, culture).

6.2 Persona Facets to Domain

Secondly, we mapped the human facets identified in each domain in which our curated personas were used in. Table 5 summarises the human facets mapping of our Persona Corpus. We divided these facets into two main layers: (1) **internal layer**; and (2) **external layer**. The internal layer of persona consists of a general background information of persona which falls into personal characteristic group. On the other hand, the external layer consists of context-specific information depending on the context and (or) the domain the personas are used in. We found that most of the personas included **motivation, goal, and concern/frustration/pain point** in their descriptions. There are a few domains that did not include all of these three attributes in the description (*security, law, finance, architecture, and transportation*). We acknowledge that this results from only a small number of collected personas used in these domains.

Table 4: Human factors categorisation in Persona Corpus.

Human factors	Human facets	Human aspect groups
age, name, and gender	Demographic information	Personal characteristics
attitude, behaviour, personality, preference, interest, and hobby	Personal attributes	
health challenge, health status, and body measurement	Physical well-being	
mental challenge, emotional feeling	Mental well-being	
activity, achievement, memorable incident, life experience, and life value	Personal story	Skill, experiential or environmental-influenced characteristics
ICT usage, ICT literacy, adaptation to technology, possessions of gadgets, and wearable device usage	Interaction with technology	
skill, health literacy	Skill level	
education, learning experience	Education	
spoken language	Environmental-influenced characteristics	
life value, family tradition, religious belief	Human values	
financial situation	Socio-economic status	Group or multiple human characteristics
occupation, income	Work status	
living arrangement, family structure, parent intervention	Family environment	
current location	Geographic location	
work experience, social interaction, complain experience	Collaboration and communication style	
culture suitability, culture	Culture	

Table 5: The mapping of identified human facets in Persona Corpus.

Domain	Motivation	Goal	Concern/frustration/pain point	INTERNAL LAYER				EXTERNAL LAYER											
				Personal Characteristics				Skill, experiential or environmental-influenced characteristics				Group or multiple human characteristics							
				Demographic information	Personal attributes	Physical well-being	Mental well-being	Personal story	Interaction with technology	Skill level	Education	Environmental-influenced characteristics	Human values	Socio-economic status	Work status	Family environment	Geographic location	Collaboration and communication style	Culture
Technology for older adults	✓	✓	✓	✓	✓			✓	✓	✓					✓	✓	✓		
Software development	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓			✓	✓	✓	✓		✓
Physical health	✓	✓	✓	✓	✓	✓		✓	✓	✓				✓	✓	✓	✓	✓	
Mental health	✓	✓	✓	✓	✓		✓	✓			✓			✓	✓			✓	
Education	✓	✓	✓	✓	✓			✓	✓	✓	✓				✓	✓	✓	✓	✓
Sustainable living	✓	✓	✓	✓	✓			✓	✓						✓	✓	✓	✓	
Technology for children	✓	✓	✓	✓	✓			✓	✓		✓				✓	✓	✓	✓	
Culture	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	
Security	✓	✓		✓	✓		✓	✓		✓					✓	✓		✓	
Law	✓			✓											✓	✓		✓	
Finance		✓		✓	✓								✓		✓	✓	✓	✓	
Architecture		✓		✓	✓			✓	✓						✓	✓	✓	✓	
Transportation		✓	✓	✓	✓			✓							✓	✓	✓	✓	

In addition to these three persona attributes, we also discovered that there are some facets on each human aspects group that were included in personas used in all identified domains. Under *personal characteristics* aspect of persona, **demographic information** and **personal attributes** are the facets with a

high occurrence. Whereas for the *skill, experiential or environmental-influence characteristics* aspect, we identified that most of personas in our Corpus included **personal story** and **interaction with technology** facets. We also identified that under *group or multiple human characteristics* facets, most of the

Table 6: Preliminary persona taxonomy.

INTERNAL LAYER	
Personal characteristics	Demographic information
	Personal attributes
EXTERNAL LAYER	
Motivation	
Goal	
Concern/frustration/pain point	
Skill/experiential/environmental-influenced characteristics	Personal story
	Interaction with technology
Group or multiple human characteristics	Work status
	Family environment
	Geographic location
	Collaboration and communication style

personas included **work status**, **family environment**, **geographic location**, and **collaboration and communication style**.

6.3 Persona Taxonomy

Based on this analysis, we developed a preliminary persona taxonomy that can be used for requirements engineering in different domains and contexts. This is outlined in Table 6. Based on our Persona Corpus analysis, the *Demographic information* and *Personal attributes* in **Internal layer** of persona consist of human factors we mentioned in Table 4 for personas used in different domain and context.

However, for the **External layer** of persona there are some considerations need to be taken. First, the set of human factors for each human aspect may differ from domain to domain, not to mention from context to context. As an example, *interaction with technology* in domain **technology for older adults** contains information about how the elderly adapt with technology as an addition to technology usage-related information.

Secondly, the value assigned to each human factor may also be different across domains. For instance, in domain of **education**, human factors under *interaction with technology* facet depict how personas use technology to support their learning and teaching activity, while in Physical health domain, the human factors portray the use of technology to help persona to maintain their physical well-being.

Each domain that we identified from our Persona Corpus analysis requires some different customisation in terms of the persona attributes that need to be included in a persona description for that domain. Therefore, we recommend persona facets for each domain to address the requirements which can be seen in Table 7.

Based on this recommendation, we can see the alignment of persona facets recommendations for cer-

tain domains. For an example, in the **physical health** domain, we recommend that persona descriptions should include key facets as follow: *physical well-being* (e.g., health challenge, health status), *skill level* (e.g., health literacy), and *socio-economic status* (i.e., financial situation). We also recommend that personas used in the **technology for older adults** domain should include *skill level* facet that shows technology literacy of the persona.

7 DISCUSSION

In this study, we conducted a review analysis of 98 personas from 41 publications which is referred to as Persona Corpus. In this analysis, we discovered three key text-based persona dimensions, which are: persona narration, persona format, and persona length. Moreover, we identified the domains in which the personas were used and discovered emerging persona facets in each domain. The result served as a basis to develop preliminary persona taxonomy, a recommendation to generate and/or validate personas.

We also investigated how persona dimensions impacted the richness of personas. For this study, we used the number of identified human factors in each persona in our Persona Corpus to determine the persona richness.

We categorised personas in our Corpus based on the domain they were used in. From each domain, we analysed the emerging human factors, which we then grouped into persona facets. By conducting content analysis, we separated facets that can be reused in different domains from domain-specific facets. Based on the analysis, we developed a preliminary taxonomy for personas that can be used to generate personas for use in cross-domain (see Figure 6). However, we acknowledge that generic personas may not be relevant to be used in a particular domain. Therefore, we also formulated a recommendation to customise a generic persona to cater to the requirements of each domain, as shown in Table 7.

There are some **major benefits** of our results, especially for persona creators. First, the dichotomy of personal layers (internal and external layer) can enhance **persona reusability**. By separating facets that can be used in different contexts and domains, persona users can reuse some parts of existing personas for other projects. This can **streamline the process to generate personas**. Second, domain-based persona facets recommendation can add more dimensions to reusable personas and make those personas more **relevant to the contexts** they are used in. Moreover, the recommendation can give persona creators ideas

Table 7: Recommendation to customise domain-based persona facets.

	Human aspects	Persona facets	Human factors	Domain	
INTERNAL LAYER	Personal characteristics	Physical well-being	health challenge	Software development, Physical health	
			health status	Physical health	
		Mental well-being	body measurement	Physical health	
			mental health	Mental health	
EXTERNAL LAYER	Skill/experiential/environmental-influenced characteristics	Skill level	emotional feeling	Security	
			health literacy	Technology for older adults	
		Education	skill	Physical health	
				Software development	
			education	Education	
				Culture	
		Environmental-influenced characteristics	learning experience	Security	
				Software development	
		Human values	spoken language	Mental health	
				Education	
		Socio-economic status	life value	Education	
				Technology for children	
		Group or multiple human characteristics	Culture	religious belief, family tradition	Culture
				financial situation	Education
culture suitability	Software development				
culture	Education				

on what facets should be included in the persona description, which, once more, can make the persona creation process less time-consuming. Third, both the taxonomy and domain-based persona facets recommendation can be useful for persona **early persona validation process** to check the richness of personas in terms of the facets present in resultant personas.

However, we must acknowledge some threats to the validity of our results. First, we collected a small number of personas used in several domains, which are *sustainable living, culture, technology for children, architecture, finance, law, security, and transportation*. This resulting in less comprehensive analysis on personas used in the aforementioned domain. Therefore, further data collection is needed, particularly for personas in those domains. Second, our proposed persona taxonomy and domain-based persona facets recommendation provide a preliminary taxonomy for personas for use in RE-related tasks. An empirical evaluation is needed to evaluate and refine both the taxonomy and recommendations for our RE-related task persona taxonomy.

Several areas of future research are needed. This study can be extended by analysing other types of persona representation (e.g., model-based persona, visual-based persona). Expand the investigation to a more detailed research, including exploring the results of the findings with practitioners' experiences when using personas which we are currently research-

ing. Elaborate the domain, e.g., education can be elaborated to early childhood education, primary education, and many more. The results of our study can be used as a foundation to develop a persona-based knowledge graph, which further can be extended to a persona-based knowledge graph tool. This tool will be useful to recommend persona facets and facet values need to be presented in personas.

8 CONCLUSION

Personas described in the academic literature as being used for various Requirements Engineering related tasks and domains vary considerably. These variations include their human aspects captured, format, narrative style, level of detail, usage of images, and so on. We curated a total of 98 personas described in 41 academic studies relating to RE-related usage of the personas. From analysis of these personas, we developed a preliminary taxonomy of different facets, representations, human aspects, domains of use. Our taxonomy can guide those engaged in RE-related tasks in formulating and choosing appropriate personas. Our 98 curated personas can serve as a reusable corpus of personas for diverse domains.

ACKNOWLEDGEMENTS

Karolita is supported by Australia Awards Scholarship and Monash Departmental Top-Up Scholarship for her Ph.D. study at Monash University, Australia.

Grundy, Kanij, and McIntosh are supported by the Australian Research Council (ARC) Laureate Fellowship project FL190100035. McIntosh is also funded by a National Health and Medical Research Council (NHMRC) Synergy Grant (APP2010268) and NHMRC Participation in Cancer Screening Programs Grant (APP2014703).

REFERENCES

- Abd Malik, S. and Azuddin, M. (2013). Mobile technology for older people: Use of personas. In *2013 Int. Conf. on Research and Innovation in Information Systems (ICRIIS)*. IEEE.
- Acuna, S. T., Castro, J. W., and Juristo, N. (2012). A hci technique for improving requirements elicitation. *Information and Software Technology*, 54(12).
- Aguirre, J., Falconi, F., Serrano, R., Moquillaza, A., and Paz, F. (2021). Improving the withdrawal functionality on atm using a ucd framework. a case study. In *Design, User Experience, and Usability: Design for Diversity, Well-being, and Social Development*, Lecture Notes in Computer Science.
- Alvertis, I., Papaspyros, D., Koussouris, S., Mouzakitis, S., and Askounis, D. (2016). Using crowdsourced and anonymized personas in the requirements elicitation and software development phases of software engineering. In *2016 11th Int. Conf. on Availability, Reliability and Security (ARES)*. IEEE.
- Antle, A. N. (2008). Child-based personas: Need, ability and experience. *Cognition, Technology and Work*, 10(2).
- Aoyama, M. (2005). Persona-and-scenario based requirements engineering for software embedded in digital consumer products. In *13th IEEE Int. Conf. on Requirements Engineering (RE'05)*. IEEE.
- Askarbekuly, N., Solovyov, A., Lukyanchikova, E., Pimenov, D., and Mazzara, M. (2021). Building an educational product: Constructive alignment and requirements engineering. In *Advances in Artificial Intelligence, Software and Systems Engineering*, Lecture Notes in Networks and Systems.
- Björndal, P., Rissanen, M. J., and Murphy, S. Lessons learned from using personas and scenarios for requirements specification of next-generation industrial robots. In *Design, User Experience, and Usability. Theory, Methods, Tools and Practice*, Lecture Notes in Computer Science.
- Bowles, J. (2007). The personification of reliability, safety, and security. In *2007 Annual Reliability and Maintainability Symposium*.
- Cleland-Huang, J., Czauderna, A., and Keenan, E. (2013). A persona-based approach for exploring architecturally significant requirements in agile projects. In *Int. Working Conf. on Requirements Engineering: Foundation for Software Quality*. Springer.
- Cooper, A. (1999). *The inmates are running the asylum*. Software-Ergonomie '99.
- Grundy, J., Mueller, I., Madugalla, A., Khalajzadeh, H., Obie, H. O., McIntosh, J., and Kanij, T. (2022). Addressing the influence of end user human aspects on software engineering. In *Communications in Computer and Information Science*, volume 1556 of *Communications in Computer and Information Science*.
- Ho, S.-H. and Lin, C. J. (2019). The requirement analysis for developing the assisted living technology for the elderly. In *Int. Conf. Cognitive Cities*.
- Kolski, C. and Warin, B. (2018). *From Persona to Living Persona, Preliminary Data from a Pilot Study in HCI Education*. Int. Conf. Learning and Collaboration Technologies.
- Lachner, F., von Saucken, C., 'Floyd' Mueller, F., and Lindemann, U. (2015). Cross-cultural user experience design helping product designers to consider cultural differences. In *Cross-Cultural Design Methods, Practice and Impact*, Lecture Notes in Computer Science.
- LeRouge, C. and Ma, J. (2010). User profiles and personas in consumer health technologies. In *2010 43rd Hawaii Int. Conf. on System Sciences*.
- Lopez-Lorca, A. A., Miller, T., Pedell, S., Mendoza, A., Keirnan, A., and Sterling, L. (2014). One size doesn't fit all: diversifying" the user" using personas and emotional scenarios. In *6th Int. Workshop on Social Software Engineering*.
- Marcengo, A., Guercio, E., and Rapp, A. (2009). Personas layering: A cost effective model for service design in medium-long term telco research projects. In *Human Centered Design*, Lecture Notes in Computer Science.
- Moser, C., Fuchsberger, V., and Tscheligi, M. (2011). Using probes to create child personas for games. In *8th Int. Conf. on advances in computer entertainment technology, ACE '11*.
- Nielsen, L., Hansen, K. S., Stage, J., and Billestrup, J. (2015). A template for design personas: Analysis of 47 persona descriptions from danish industries and organizations. *Int. journal of sociotechnology and knowledge development*, 7(1).
- Nunes Rodrigues, G., Joel Tavares, C., Watanabe, N., Alves, C., and Ali, R. (2018). A persona-based modelling for contextual requirements. In *Int. Working Conf. on Requirements Engineering: Foundation for Software Quality*. Springer.
- Salminen, J., Guan, K., Nielsen, L., Jung, S.-g., and Jansen, B. J. (2020). A template for data-driven personas: Analyzing 31 quantitatively oriented persona profiles. In *Human Interface and the Management of Information. Designing Information*, Lecture Notes in Computer Science.
- Sedeno, J., Schon, E. M., Torrecilla-Salinas, C., Thomaschewski, J., Escalona, M. J., and Mejias, M.

- (2017). Modelling agile requirements using context-based persona stories. 13th Int. Conf. on Web Information Systems and Technologies.
- Sim, W. W. and Brouse, P. (2015). Developing ontologies and persona to support and enhance requirements engineering activities – a case study. *Procedia computer science*, 44.
- Sim, W. W. and Brouse, P. S. (2014). Empowering requirements engineering activities with personas. volume 28 of *Procedia Computer Science*.
- Sommerville, I. (2016). *Software Engineering, Global Edition*.
- Wöckl, B., Yildizoglu, U., Buber, I., Diaz, B., Kruijff, E., and Tscheligi, M. (2012). Basic senior personas: a representative design tool covering the spectrum of european older adults. 14th ACM SIGACCESS Conf. on Computers and Accessibility.

