Article

Data-Assisted Persona Construction Using Social Media Data

Dimitris Spiliotopoulos [1],* , Dionisis Margaris [2] and Costas Vassilakis [1]

[1] Department of Informatics and Telecommunications, University of the Peloponnese, 22100 Tripolis, Greece; costas@uop.gr

[2] Department of Informatics and Telecommunications, University of Athens, 15784 Athens, Greece; margaris@di.uoa.gr

* Correspondence: dspiliot@uop.gr; Tel.: +30-2710-372203

check for updates

Abstract: User experience design and subsequent usability evaluation can benefit from knowledge about user interaction, types, deployment settings and situations. Most of the time, the user type and generic requirements are given or can be obtained and used to model interaction during the design phase. The deployment settings and situations can be collected through the needfinding phase, either via user feedback or via the automatic analysis of existing data. Personas may be defined using the aforementioned information through user research analysis or data analysis. This work utilizes an approach to activate an accurate persona definition early in the design cycle, using topic detection to semantically enrich the data that are used to derive the persona details. This work uses Twitter data from a music event to extract information that can be used to assist persona creation. A user study in persona construction compares the topic modelling metadata to a traditional user collected data analysis for persona construction. The results show that the topic information-driven constructed personas are perceived as having better clarity, completeness and credibility. Additionally, the human users feel more attracted and similar to such personas. This work may be used to model personas and recommend suitable ones to designers of other products, such as advertisers, game designers and moviegoers.

Keywords: persona; user experience; user interface design; topic modelling; usability; personalization; cultural events

1. Introduction

Personas are constructs that represent user archetypes and have been used extensively in various stages of human-computer interaction design. Personas have traits that subsets of potential users exhibit [1]. The traits can be mapped to user requirements, usually not in a one-to-one manner but rather as abstracted characteristics that the user requirements can be derived or fine-tuned to. Personas may be useful in various stages of a design. For the requirements or needfinding phase, designers use them to formulate the requirements, and explain them to the design team by applying them to the personas. For the design phase, prototyping may be applied to personas for testing. For the evaluation, the product is applied to the personas in order to evaluate whether the expected goals are met and identify whether the needs were met and to what extent. For the deployment stage, personas may be applied accordingly, depending on the actual persona and product, from market share penetration evaluation to research objectives fulfilment to ethical and inclusiveness verification [2].

Personas can be integrated to use case models, resulting in revised or adapted use case models that describe more compact sets of use cases [3]. They may also be created and utilized as part of stories and scenarios [4] for enriching scenario or case-driven interaction design processes. On the other hand, other studies include recommendations toward separating scenarios and persona descriptions [5].

Designing personas well is an important aspect, since only a small number of personas may be created for them to be useful. Therefore, personas have to be concise enough to accurately express the design requirements, as well as conceptually abstract enough to provide the required coverage of the user types and requirements [6]. One of the major shortcomings of personas is that they can hardly account for change, especially fast change. Even the most well-constructed personas may become partially obsolete or inaccurate after a period of time, resulting in the need for additional effort, time and expense in order to repair inconsistencies and lost credibility [7].

Salminen et al. did an extensive study to evaluate how persona creation and utilization are affected by statistical online analytics using big data [8]. The critical points that were reported are the main challenges that this paper addresses and are as follows:

1. Creating personas is a costly and lengthy process,
2. Personas may be biased by their creators,
3. Personas may be non-verifiable and untrustworthy when they depend on the information used for their creation,
4. Personas may become inaccurate over time.

There are several approaches to data-driven persona construction, systems and methodologies for quantitatively generating personas using large amounts of online social media data [9]. However, data-generated personas were found to suffer from similar shortcomings to the traditional manually constructed personas. Coherency and consistency is an inherent problem and a challenge for persona designers when constructing and utilizing the elements of information in a meaningful and usable form [10].

Automatically generated personas address challenge 1 and 2 by utilizing web data to automatically create a number of personas, requiring minimal human involvement. However, they cannot address trustworthiness and inaccuracy over time. Designer-generated personas are costly and take time to create. Additionally, they may be biased by their creators. However, they generally address challenges 3 and 4 better than automatically created personas, since designers may pick sources and specific data that seem trustworthy, as well as selecting representative information that they deem to be as futureproof as possible. From the above information, it would be interesting to put an approach to the test that may address all challenges to a certain degree.

This work examines how traditional persona design may be assisted by persona metadata derived from fast-changing big data. Building on the identified shortcomings of the manual and the data-generated persona construction and their individual advantages, this paper proposes a hybrid approach that is simple enough to apply, yet contextual and analytical so as to provide useful insights.

The structure of the paper is outlined as follows: Section 2 presents the related work. Section 3 presents the motivation and rationale behind this work. Section 4 details the experimental setup and method. Section 5 presents the results of the user study experiments in persona construction, while Section 6 presents the evaluation of the cultural event persona construction. Section 7 discusses the paper's results and outlines future work.

2. Related Work

Personas are not constant. One of the major points of critique in the use of personas is that they become irrelevant or non-applicable very soon. Therefore, the designer needs to account for variations and update the personas. The change can be significant, even for six-month or yearly periods. An advantage of data-driven persona construction is that continuous time-stamped data may be used to account for persona variation over time [11]. The challenge for data-driven persona construction is to monitor and identify how change happens over time: the veracity, velocity and volume. Findings show that topical interests, as reflected by personas constructed using data from online sources, change by an average of over 20%, while only a third of the personas in those cases experience topical consistency [12]. This shows the necessity for a constant update of the personas in

order to reflect the changes in topical interests. The frequency of the employed routine data analysis to achieve the updates reflects upon the design lifecycle [13].

Automatic Persona Generation is the implementation of a methodology for quantitatively generating data-driven personas from online social media data [14]. Personas may be generated automatically, in real time, using very rich social media data that include timestamps such as YouTube, eliminating most of the labor associated with persona construction [15]. On the other hand, personas that are built from user data may be incomplete or incomprehensible, and therefore unusable as they are, requiring the designer to barge in and fill in the blanks [16]. Therefore, the interpretation of the persona characteristics is designer-dependent and sometimes designer-biased.

Tapping into social web information is a challenging task, mainly due to reasons that are associated with the processing and analysis of social web data [17,18]. On the other hand, the personas themselves are created for various tasks; there are personas that are required for marketing, for social research and for educational designs, amongst others. There are personas that can cover all possible cases (elastic) or personas that are only useful for narrow or very specific cases. Christoforakos et al. examined marketing stakeholder personas for prototyping [19], while Schoch et al. created personas to understand social barriers and used them for prototyping a web app [20]. Ozkan et al. showed the importance of how designers or product owners, in their case the university faculty, regard the disconnection between them and users, in their case the students, using personas as a design technique for revamping a university school curriculum [21].

Personas may be influenced by their designers or by researchers who are making assertions about the expectations of other users. The designer team, as well as the example data and their size, form a dynamic mix that unknowingly assigns bias to the personas [22,23]. Salminen et al. examined data-generated personas under the assumption that bias may be affected by the age and gender of the persona as well as by the number of generated personas [24]. The study found that a small number of personas increased the bias, which would be a valid hypothesis since the bias-inducing parameters would be exaggerated in a small set of personas. In their study, female personas were found to be underrepresented for small persona sets. Therefore, algorithmic bias is present and a manual validation by experts is necessary.

The vast number of data may lead to personas that either summarize user requirements or contain overly precise information, making them have an insignificant impact [25]. Demographics are a textbook example for broad data that require proper taming so that they either do not result in unnecessary large number of personas or they are not spread thin and consumed by other persona attributes. A study using YouTube data from videos utilizing the full demographic classification showed that 2772 demographic-based personas would be generated using the existing demographic groupings for gender, age and origin [26]. An et al., (2018) used aggregated data to define customer behavior segments and created personas based on the demographics from those segments [27].

Co-creating personas benefits users so as to have them engage in accessible design, and it achieves a broader inclusion of demographics in the co-created personas [28]. Extending the use of personas outside user archetype modelling, personas can be used for roleplay simulation with real users for collaborative design [29]. In a recent study, interaction design across cultures could be aided using child-generated personas [30]. The study found that children could be more expressive, providing details based on enthusiasm, which in turn provided behavioral and activity-based thematic scenarios.

Depending on the generated number of personas, unsupervised learning methods, such as clustering or topic modelling, may be used to cluster the personas based on their attributes, thereby providing a means to go from raw data to understandable semantics. The actual attributes and the way they are presented to designers affects their perception of the personas. Salminen et al. determined that using actual numbers to describe attributes had a positive effect on the perception of the persona usefulness by users such as analysts but a significantly negative effect on the perception of the persona completeness by both analysts and market experts [31].

Transparency in data-driven generated personas is achieved by providing the sources of the information on the personas. Transparency affects credibility (decrease), completeness (increase) and clarity (increase). The persona gender also affects the perceived completeness of the persona by the user, but this was evident only for female personas [32].

Incomplete personas that may not contain certain types of information constitute an attempt by researchers to eliminate factors that induce bias and uncertainty. For example, "thin" personas that do not contain personalization samples such as a name and picture but retain demographic (gender, origin, age) and behavioral attributes are used for automated categorization methods, such as clustering to reduce the numbers of generated personas [33]. That way, persona sets are described by their clustered core objective information and avoid the causes of subjectiveness that the personal attributes would induce.

Another way to fine-tune or reduce persona sets is by traditional large scale online surveys and a quantitative analysis of the questionnaire information so as to revise the persona sets or even create additional personas that were not generated by the data-driven methods [34]. Xu and Lee identified persona types for online shopping communities using large scale surveys [35]. They analyzed the data in terms of social connections and characteristics, such as reading and posting behavior, which led, via clustering, to a limited number of personas as categories of users. Those "very thin" personas were described by their main representative social behavior characteristic and an accompanying descriptive sentence. An additional aspect that designers can keep in check is perceived likability. Studies show that, similar to designer bias, users and designers are affected by visual properties. To keep this effect from happening, pictures (stock, generated or otherwise, e.g., sketched) may be refrained from being used so that the acceptance of the persona by the users will not be affected by the likeability of the persona picture [36].

3. Motivation

Data-driven persona generation may utilize multiple analysis techniques as well as traditional methods for accurate persona construction. Kim et al. used a trend analysis as well as face-to-face interviews and online surveys to extract cybersecurity-attributed user characteristics [37]. They used this hybrid technique to compare the data from the three sources (trend analysis, face-to-face interviews and online surveys) in order to formulate the personas. Such a post-analysis was quite difficult, since the data collection from the sources was performed in parallel and the data were not cross-fed during or after the collection process. Therefore, the datasets had different granularity and coarseness values, as well as no automatic connections, which the users were then tasked to understand and correlate.

All the aforementioned identified issues related to the collection of data, analysis of the persona attributes, construction and use of personas result in problems that end users, designers, marketeers and researchers ultimately face [38,39]. Matthews et al. identified the main issues with personas with regard to users, finding them misleading and distracting as well as abstract and impersonal [40]. The authors argue that perhaps a more prudent approach to persona formulation would be to avoid persona attributes that mislead and distract the users. Furthermore, they deem this aspect as being more important than striving to create engaging personas.

From the above information, it is argued that automatically data-generated personas cannot fully replace the designers delving into the data and the insights and intuitions that they gain for the design requirements. Several studies also identify specific persona shortcomings that trigger mistrust, causing the designers to refrain from adopting them fully for their design approaches. Achieving a balance between data-generated personas and human intuition is the motivation of this work.

The hypothesis is that, based on the related literature, data-assisted persona construction may yield more accurate personas. The approach of this work is that, instead of collecting human knowledge from questionnaires and interviews and combining or fusing the knowledge with the data-generated personas, the designers can be assisted on a higher level with data-processed information related to persona construction. The information is stripped from any data or aspects that affect human

decisions on a sentimental or likeness level, thereby shielding the designers from knowingly or unknowingly induced bias. This way, the design process is supported by the data analysis, while at the same time allowing the designers to utilize higher-level data knowledge in their traditional persona construction approach.

In the following subsections, we present the experimental approach to big data assisted persona construction, examining the effectiveness of an elaborate data analysis for the created personas, and comparing it with the traditional and frequently used data collection and analysis by human designers.

4. Experimental Setup and Method

In the following paragraphs, we elaborate on (i) the experimental design, (ii) the data-driven persona metadata-assisted designer user study and (iii) the evaluation of the persona designs using standard metrics.

4.1. Data

The data were collected from Twitter for the well-known live music event @rockamring from January to February 2020 (inclusive). The collection crawled the most frequent relevant hashtags, such as #rar2020 and #rockamring. The former was used for the topic modelling, and both were used for the pictures, videos and links. We used a pipelined process to clean up and validate the data [41]. Out of the posts collected, 1811 were used for topic modelling using the Latent Dirichlet Allocation (LDA) approach [42,43].

The topics were modelled as interesting based on LDA clusters, the tie strength of the context words from a sentiment analysis, and quantitative social sharing information from their associated Twitter posts [44]. User information, including gender, demographics and name/photo were excluded to avoid biasing the designers for or against specific information. Figure 1 shows the topics of interest as extracted for the aforementioned period.
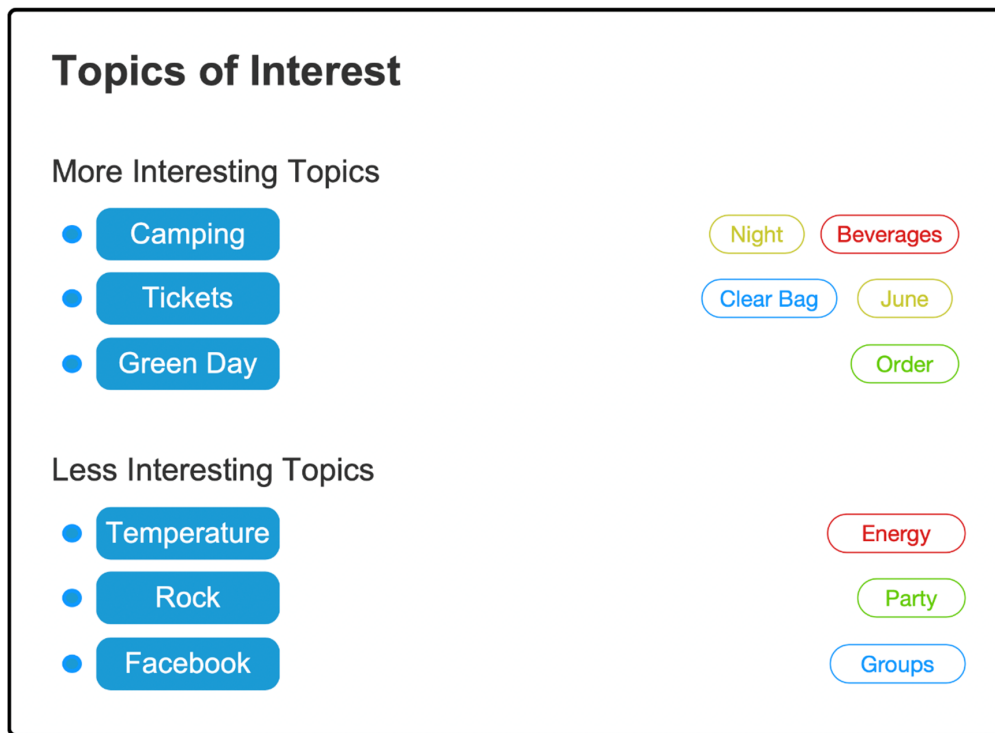

Figure 1. Topics presented to the designer participants. The main topics are on the left and related keywords are on the right. The keyword colors represent the identified sentiment (green: positive, yellow: neutral, red: negative, blue: no sentiment).

4.2. Participants

The user study and evaluation participants were recruited through the University forum and social networks. Thirty English speaking participants were selected, 57% male and 43% female. The average age of the participants was 22 years. All were undergraduate and graduate students, while 70% of them reported having taken an HCI-related course and all had previously participated in human studies. All reported familiarity with the use of social networks for obtaining information. All participants attended an informal lecture on personas and user design. Examples of personas and their use were provided during that session. After their familiarization, they were explained about the study specifics and given the tasks (Table 1).

Table 1. Persona construction and evaluation task breakdown.

Group	Task	Sources
A	create personas using topic modelling from twitter data	topic information
B	create personas using any information deemed useful	Twitter, analytics tools
A + B	evaluate personas of two random participants, one from each group	persona perception scale, usability evaluation questionnaire

The participants were randomly placed into two groups of 15 people each. The task was to construct thin personas, so photos, biographical information, personal status, quotes, work and background text were optional. The reason for this was to eliminate or minimize the potential bias for the persona peer evaluation. The participants were given a minimum of one hour and a maximum of two hours to construct personas for the selected music event. They were told to use an online translation app for the non-English content, knowing that the specific event had a large amount of German language content. The study facilitators recorded details of interest on paper during the sessions.

5. Persona Construction

All participants from both groups created personas in the allotted timeframe. Since the participant experience in creating personas varied, the time spent to finish the tasks could not reflect on the data usefulness for either group. A total of 159 personas were constructed by all participants. Table 2 shows the breakdown of the number of personas created per group and participant gender.

Table 2. Number of personas constructed per participant group (task) and gender.

Group	All Participants	Male Participants	Female Participants
A	4.4 (std: 1.18)	4.38 (std: 1.19)	4.43 (std: 1.27)
B	6.2 (std: 1.37)	6.67 (std: 1.00)	5.50 (std: 1.64)
A + B	5.3 (std: 1.56)	5.59 (std: 1.58)	4.92 (std: 1.50)

Group A participants constructed 66 personas in total, while Group B participants constructed 93, which was 41% more. It is also evident that the female and male participants of Group A constructed about the same average number of personas, while in Group B, male participants constructed more than one additional persona than their female group partners did.

The task of the Group B participants was more demanding, since they were not provided with the topic analysis information and had to explore the data on their own. In order to do so, they utilized several online Twitter analytics tools, such as the Tweet Sentiment Visualizer (https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/), which enables insight into sentiment,

topics, timeline, connections, maps and timelines (Figure 2), and the Floom (https://floom.app/) app, which enables a quick look into the keywords mentioned in Twitter streams (Figure 3).

Figure 2. Looking into the data for persona construction: sentiment (**top**) and topics (**bottom**). The users select and view the sentiment for the entities of interest and the cluster information for a deeper view of how topics are semantically identified.

Figure 3. Standard keyword cloud for Twitter streams. The designer may use the keywords to search social media and the web for user comments that they can utilize for the persona construction.

The Group B participants also used Twitter and Facebook as a source. They utilized the information that was automatically generated to quickly access and verify the content and select representative posts and threads for useful information.

6. User Evaluation

The participants were asked to evaluate randomly assigned personas created by their peers. Each respondent evaluated two personas, for a total of 318 persona evaluations. The personas were assigned randomly, one from each group, for every evaluator. This work utilized the Persona Perception Scale, developed by Salminen et al. [45], for a user evaluation of social-media data-generated personas. This scale fully applies to our peer evaluation tasks, since it accounts for multiple aspects of interest. The exploratory nature of this work encouraged the use of a full scale for the facilitators to be able to observe possible subtle differences between the two groups.

Group A participants used our topic modelling tool to get the information for the personas (Figure 4). The topic analysis also provided trending information for both the topics and the keywords. Additionally, we utilized a customized model for the SentiStrength sentiment analysis tool, which included emotions [43]. The designers utilized the information for the present detailed topics, which included sentiment and trending information.
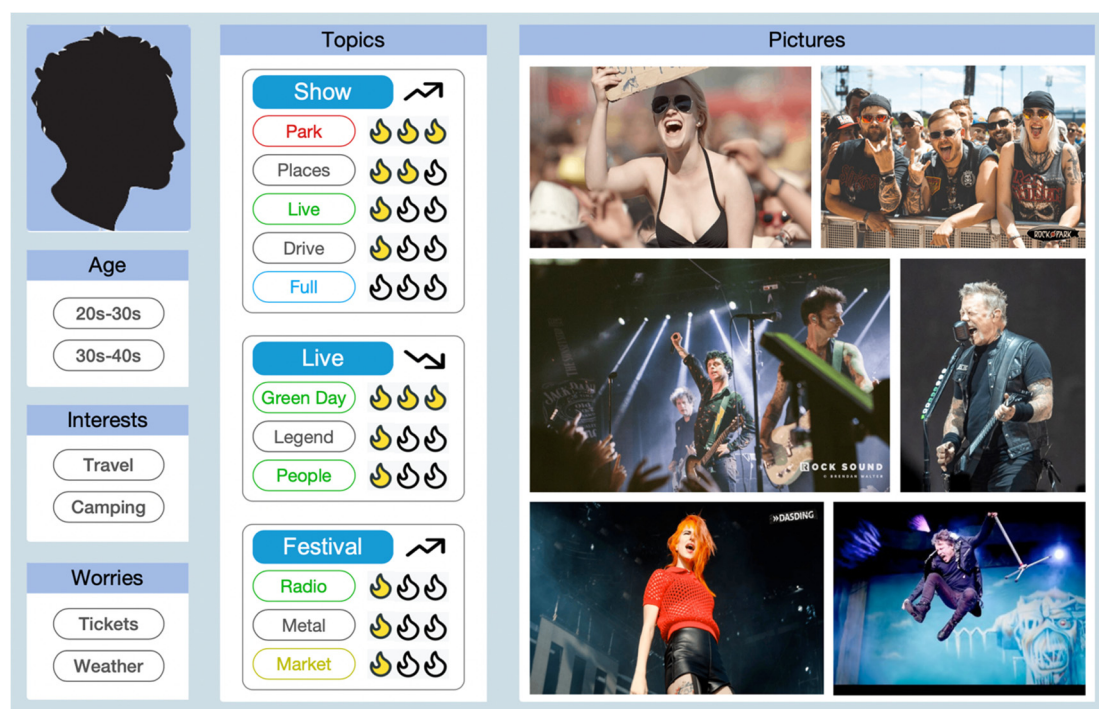

Figure 4. User-constructed topic modelling-assisted thin persona.

Group B participants used the output of the analysis tools as a guide for the validation and selection of the most prominent information to use for their persona construction (Figure 5). All sections of the persona were selectable, and the data options were editable. Observing the designers, they edited the age groups based on social media data, selected the number of keywords and validated their sentiment using at least two sources of information. Picture selection was a necessary process, since variations of the same pictures would originally lead to a few pictures overwhelming the selection. The users reported that the Group B personas exhibited a much greater variety of pictures.
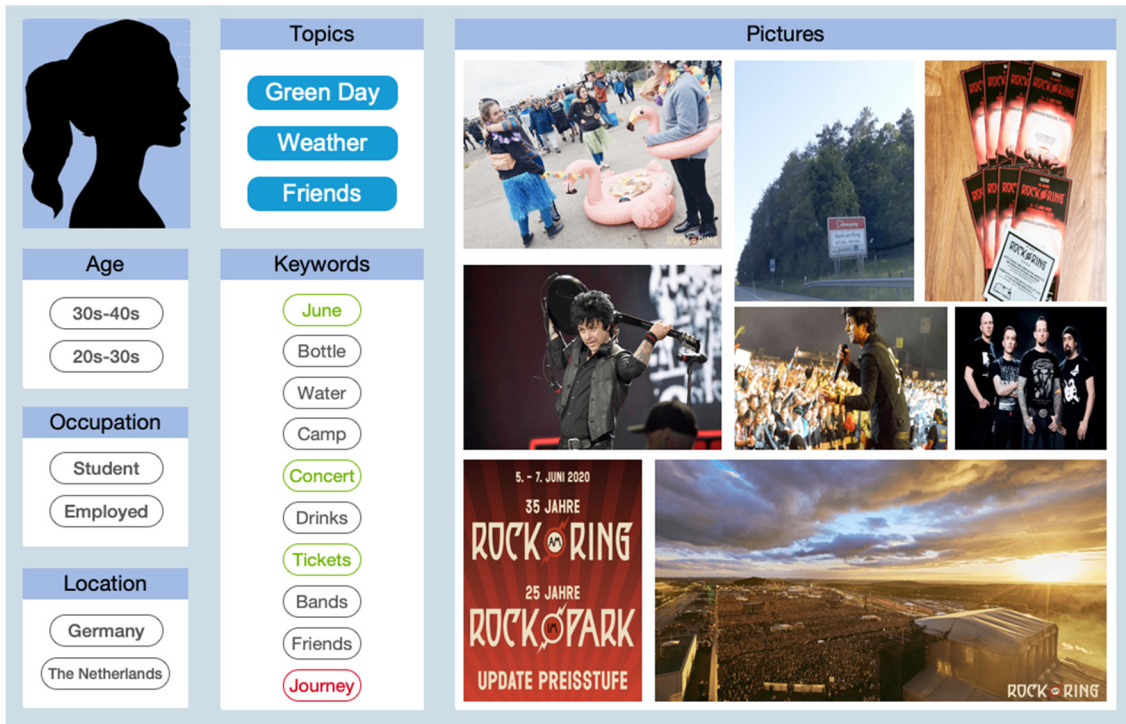
Figure 5. User-constructed meta-data-assisted thin persona.

Figure 6 depicts the average user evaluation responses per perceived persona aspect on a standardized Likert scale of 1–5.
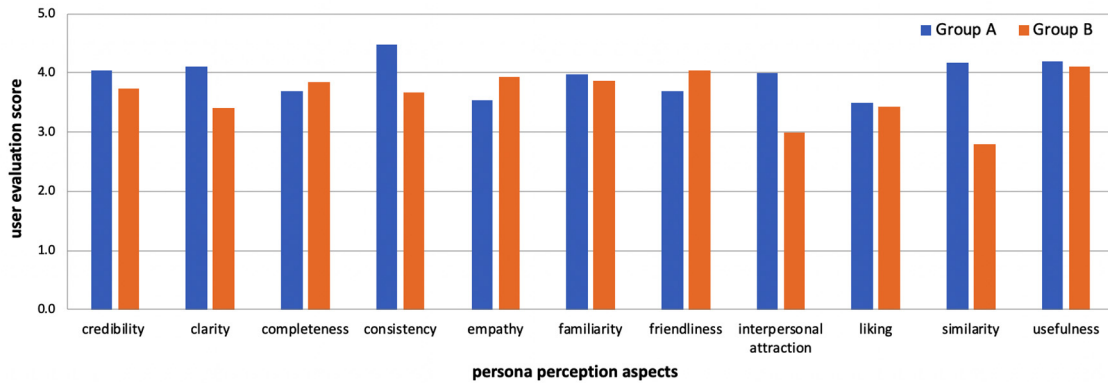

Figure 6. Persona perception user study results.

The Group A designers who utilized the advanced topic information scored higher on credibility, clarity and consistency, while Group B scored higher in completeness and empathy. This can be explained by the fact that Group A participants utilized already curated and streamlined topic information from the same source, while Group B participants accessed multiple sources that were also different per participant. The use of the same source information led to clearer and to more consistent results that were also perceived as more credible. On the other hand, the use of multiple sources led to a higher number of constructed personas, some of which contained more personal (or less general) information bits that were perceived as being more complete as a whole, as well as more sympathetic by the evaluators.

The two groups scored similarly for familiarity and liking, showing that the topic modelling accurately reflected on familiar terms and descriptions, and both approaches resulted in personas

that were liked by the respondents. Personas constructed by Group B were perceived as marginally friendlier, which was also the result of the utilization of more personal information bits.

The major findings of this study were the responses on the interpersonal attraction (the level of attractiveness of the personas by the participants) and the similarity (the level of perceived similarity of the participants to the personas). The personas constructed by Group A participants scored much higher in the evaluation for both aforementioned aspects. This was not expected, nor was it hypothesized before the study. One possible explanation was that the clarity and consistency of the personas constructed using the topic modelling knowledge resulted in the users feeling more attracted and similar to the personas. Another possible explanation, mentioned by the evaluators during the post-study discussion, was that topic modelling clustered the data to more abstract notions, thereby flattening possible extreme or outlier data that could lead to unattractive personas. Even if the number of such personas would be very low, it might still affect their perceived attractiveness and feeling of similarity for the evaluators.

The participants also self-reported their acceptance and confidence for the personas they created. The rationale behind this metric is derived from the user experience evaluation, where the designs are evaluated by the end users and the designers use that feedback to self-reflect on their designs. In our case, the participants evaluated other personas but also their own on the bases of their acceptance to use their personas themselves and their confidence about their response. Figure 7 shows that the participants of Group B reported a much higher acceptance with a similarly high confidence. The participants of Group A reported a high acceptance of their own designs, which was, however, lower on average than that of the other group, with a very high confidence. Based on the literature, this is an expected result, and it is justified from the fact that the Group B participants were fully responsible for the data collection, analysis and persona design. Thus, they were confident that they did their best to design personas that they would use themselves. On the other hand, the participants of Group A used the already analyzed data to the best of their abilities, and they were confident that they produced very good results. However, they could not be sure that the data they had in their hands provided the maximum coverage of the requirements.
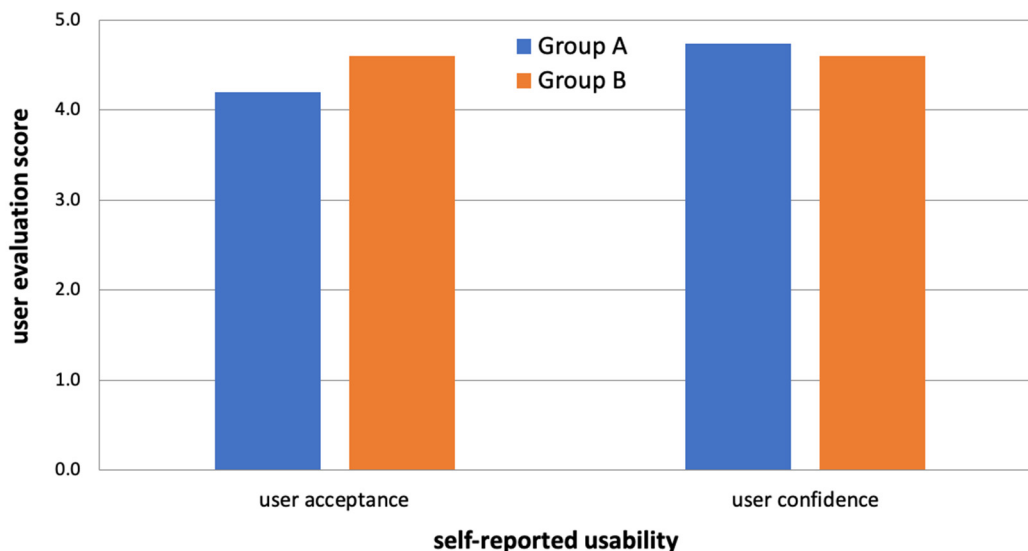

Figure 7. Participants' self-reported acceptance and confidence about their constructed personas.

7. Conclusions and Future Work

This paper presented a human study that aimed to examine the effect of big data utilization on persona construction. It followed the rationale, derived from earlier works, that automatically

data-generated personas cannot fully replace the designer's immersion in actual data in terms of persona creation.

The findings showed that deep analysis and the use of data analytics, such as topic modelling, can lead to personas that are perceived as clear, consistent and complete. Furthermore, this persona design is perceived as very appealing to users and the personas as something that the users would feel quite similar to. This approach requires much less effort than traditional human-directed data analysis and may be especially helpful for limited scope personas, such as music events, thematic museums (e.g., war museum), as well as educational or medical applications.

Based on the findings of this work, an optimal approach to persona construction using big data analytics could be a combination of the two approaches that were examined, or even a triple combination of data-generated personas using data analytics and manual analysis for refinement. This work has basic limitations with regard to the target of the persona construction, namely a music event. This is a limited scope domain that was selected to demonstrate how data analytics may reveal aspects not easily discovered by designers, but how it may also allow for an extensive human study to be made possible by limiting the data and the scope of the experiments.

To evaluate the findings from this work against purely automatic data-generated personas, a comparative evaluation, including personas automatically generated using approaches such as the one by Salminen et al. [46], would be required. However, this was not applicable for our experiments because of the focus shift and the additional effort that this would require by the participants, as well as the complexity that the endeavor of such a non-standard evaluation between the three (automatic data-generated, topic modelling, user analysis) approaches would introduce. Moreover, there are already existing works that compare fully automatically generated personas with traditional ones, and which have yielded results that have been discussed in this paper [47].

This work has limitations that are bound by the tools, the data and the users. The tools and their use is a matter of personal expertise by the designers. The data that are used also comprise a designer choice (in this case, it was a cultural event), as do the sources. The tools were selected for their ease of use, since the users were familiar with them. As a process, the persona construction would not have been affected by using different or additional tools; however, the content and user decisions could have been. For example, tweaking the LDA parameters or having additional data for the analysis would possibly yield different results, and the users would have to work with those as their choices. However, an automatic persona construction would also have been affected by such parameters. The same core data sources have been used for all user groups in order to retain comparative fidelity.

To monitor changes, topic change information may be displayed, such as trending topics, a timeline view and clustering with regard to sentiment. This would allow designers to edit or amend their personas to account for major cases. Specific situations, such as the recent global COVID-19 pandemic, may lead to specific considerations regarding design thinking (for quarantine or online user experiences), introducing new potential users and methods for content delivery that requires a fast adaptation of user design. The persona design and update would be key for a user design's rapid adaptation to new situations and emerging requirements, by utilizing the information change from the main differences of the personas.

For future work, we are planning to include an analysis of textual information from existing social network users for the automatic adaptation of existing personas with regard to their content description for a fully-fledged persona construction [48,49]. Additionally, multiple sources, such as Facebook, could be utilized for automatic enrichment, since users reported that they found interesting information and expected it to be a valid resource for cultural event-based user content.

Author Contributions: Conceived, designed and performed the experiments; analyzed the data and wrote the paper, D.S., D.M. and C.V. All authors have read and agreed to the published version of the manuscript.

References

1. Guo, A.; Ma, J. Archetype-Based Modeling of Persona for Comprehensive Personality Computing from Personal Big Data. *Sensors* **2018**, *18*, 684. [CrossRef] [PubMed]
2. Miaskiewicz, T.; Kozar, K.A. Personas and user-centered design: How can personas benefit product design processes? *Des. Stud.* **2011**, *32*, 417–430. [CrossRef]
3. Dittmar, A.; Forbrig, P. Integrating Personas and Use Case Models. In Proceedings of the Human-Computer Interaction—INTERACT 2019, Paphos, Cyprus, 2–6 September 2019; Lamas, D., Loizides, F., Nacke, L., Petrie, H., Winckler, M., Zaphiris, P., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 666–686.
4. Nielsen, L. Personas in Use. In *Personas—User Focused Design*; Nielsen, L., Ed.; Springer: London, UK, 2019; pp. 83–115, ISBN 978-1-4471-7427-1.
5. Nielsen, L. Making Your Personas Live. In *Personas—User Focused Design*; Nielsen, L., Ed.; Springer: London, UK, 2019; pp. 161–170, ISBN 978-1-4471-7427-1.
6. Pruitt, J.; Grudin, J. Personas: Practice and Theory. In Proceedings of the 2003 Conference on Designing for User Experiences—DUX '03, San Francisco, CA, USA, 6–7 June 2003; ACM Press: New York, NY, USA, 2003; pp. 1–15.
7. Salminen, J.; Jansen, B.J.; An, J.; Kwak, H.; Jung, S. Are personas done? Evaluating their usefulness in the age of digital analytics. *Pers. Stud.* **2018**, *4*, 47–65. [CrossRef]
8. Jansen, B.J.; Salminen, J.O.; Jung, S.-G. Data-Driven Personas for Enhanced User Understanding: Combining Empathy with Rationality for Better Insights to Analytics. *Data Inf. Manag.* **2020**, *4*, 1–17. [CrossRef]
9. Jung, S.G.; Salminen, J.; An, J.; Kwak, H.; Jansen, B.J. Automatically Conceptualizing Social Media Analytics Data via Personas. In Proceedings of the International AAAI Conference on Web and Social Media, Stanford, CA, USA, 15 June 2018.
10. Salminen, J.; Sengün, S.; Jung, S.; Jansen, B.J. Design Issues in Automatically Generated Persona Profiles: A Qualitative Analysis from 38 Think-Aloud Transcripts. In Proceedings of the 2019 Conference on Human Information Interaction and Retrieval, Glasgow, Scotland, UK, 10–14 March 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 225–229.
11. Jung, S.; Salminen, J.; Jansen, B.J. Personas Changing Over Time: Analyzing Variations of Data-Driven Personas During a Two-Year Period. In Proceedings of the Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland, UK, 4–9 May 2019; Association for Computing Machinery: New York, NY, USA, 2019.
12. Jansen, B.J.; Jung, S.; Salminen, J. Capturing the change in topical interests of personas over time. *Proc. Assoc. Inf. Sci. Technol.* **2019**, *56*, 127–136. [CrossRef]
13. Kouroupetroglou, G.; Spiliotopoulos, D. Usability methodologies for real-life voice user interfaces. *Int. J. Inf. Technol. Web Eng.* **2009**, *4*, 78–94. [CrossRef]
14. Jung, S.-G.; Salminen, J.; Jansen, B.J. Giving Faces to Data: Creating Data-Driven Personas from Personified Big Data. In Proceedings of the 25th International Conference on Intelligent User Interfaces Companion, Cagliari, Italy, 17–20 March 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 132–133.
15. An, J.; Cho, H.; Kwak, H.; Hassen, M.Z.; Jansen, B.J. Towards Automatic Persona Generation Using Social Media. In Proceedings of the 2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), Vienna, Austria, 22–24 August 2016; pp. 206–211.
16. Chang, Y.; Lim, Y.; Stolterman, E. Personas: From theory to practices. In Proceedings of the 5th Nordic conference on Human-computer interaction building bridges—NordiCHI '08, Lund, Sweden, 20–22 October 2008; ACM Press: New York, NY, USA, 2008; pp. 439–442.
17. Schefbech, G.; Spiliotopoulos, D.; Risse, T. The Recent Challenge in Web Archiving: Archiving the Social Web. In Proceedings of the International Council on Archives Congress, Brisbane, Australia, 20–25 August 2012; pp. 1–5.
18. Aivazoglou, M.; Roussos, A.O.; Margaris, D.; Vassilakis, C.; Ioannidis, S.; Polakis, J.; Spiliotopoulos, D. A fine-grained social network recommender system. *Soc. Netw. Anal. Min.* **2020**, *10*, 8. [CrossRef]

19. Christoforakos, L.; Tretter, S.; Diefenbach, S.; Bibi, S.-A.; Fröhner, M.; Kohler, K.; Madden, D.; Marx, T.; Pfeiffer, T.; Pfeiffer-Leßmann, N.; et al. Potential and Challenges of Prototyping in Product Development and Innovation. *i-com* **2019**, *18*, 179–187. [CrossRef]

20. Schoch, E.; Choi, A.M.L.A.; Lee, H.; Connor, S.; Rose, E.J. The Food Locker: An Innovative, User-Centered Approach to Address Food Insecurity on Campus. In Proceedings of the 37th ACM International Conference on the Design of Communication, Portland, OR, USA, 4–6 October 2019; Association for Computing Machinery: New York, NY, USA, 2019.

21. Ozkan, D.S.; Reeping, D.; McNair, L.D.; Martin, T.L.; Harrison, S.; Lester, L.; Knapp, B.; Wisnioski, M.; Patrick, A.; Baum, L. Using Personas as Curricular Design Tools: Engaging the Boundaries of Engineering Culture. In Proceedings of the 2019 IEEE Frontiers in Education Conference (FIE), Covington, KY, USA, 16–19 October 2019; pp. 1–7.

22. Niskanen, K.; Bosch, M.; Wils, K. Scientific Personas in Theory and Practice—Ways of Creating Scientific, Scholarly, and Artistic Identities. *Pers. Stud.* **2018**, *4*, 1–5. [CrossRef]

23. Bosch, M. Looking at Laboratory Life, Writing a Scientific Persona: Marianne van Herwerden's Travel Letters from the United States, 1920. *L'Homme* **2018**, *29*, 15–34. [CrossRef]

24. Salminen, J.; Jung, S.-G.; Jansen, B.J. Detecting Demographic Bias in Automatically Generated Personas. In Proceedings of the Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland, UK, 4–9 May 2019; Association for Computing Machinery: New York, NY, USA, 2019.

25. Salminen, J.; Jung, S.; Jansen, B.J. The Future of Data-driven Personas: A Marriage of Online Analytics Numbers and Human Attributes. In Proceedings of the 21st International Conference on Enterprise Information Systems—Volume 1: ICEIS, Heraklion, Crete, Greece, 3–5 May 2019; SciTePress: Setúbal Municipality, Portugal, 2019; pp. 608–615.

26. Jung, S.-G.; An, J.; Kwak, H.; Ahmad, M.; Nielsen, L.; Jansen, B.J. Persona Generation from Aggregated Social Media Data. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 1748–1755.

27. An, J.; Kwak, H.; Jung, S.; Salminen, J.; Jansen, B.J. Customer segmentation using online platforms: Isolating behavioral and demographic segments for persona creation via aggregated user data. *Soc. Netw. Anal. Min.* **2018**, *8*, 54. [CrossRef]

28. Neate, T.; Bourazeri, A.; Roper, A.; Stumpf, S.; Wilson, S. Co-Created Personas: Engaging and Empowering Users with Diverse Needs Within the Design Process. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland, UK, 4–9 May 2019; Association for Computing Machinery: New York, NY, USA, 2019.

29. Li, B.; Segonds, F.; Mateev, C.; Lou, R.; Merienne, F. Design in context of use: An experiment with a multi-view and multi-representation system for collaborative design. *Comput. Ind.* **2018**, *103*, 28–37. [CrossRef]

30. Sim, G.; Shrivastava, A.; Horton, M.; Agarwal, S.; Haasini, P.S.; Kondeti, C.S.; McKnight, L. Child-Generated Personas to Aid Design Across Cultures. In Proceedings of the Human-Computer Interaction—INTERACT 2019, Paphos, Cyprus, 2–6 September 2019; Lamas, D., Loizides, F., Nacke, L., Petrie, H., Winckler, M., Zaphiris, P., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 112–131.

31. Salminen, J.; Liu, Y.-H.; Engün, S.; Santos, J.M.; Jung, S.; Jansen, B.J. The Effect of Numerical and Textual Information on Visual Engagement and Perceptions of AI-Driven Persona Interfaces. In Proceedings of the 25th International Conference on Intelligent User Interfaces, Cagliari, Italy, 17–20 March 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 357–368.

32. Salminen, J.; Santos, J.M.; Jung, S.-G.; Eslami, M.; Jansen, B.J. Persona Transparency: Analyzing the Impact of Explanations on Perceptions of Data-Driven Personas. *Int. J. Hum. Comput. Interact.* **2019**, *36*, 788–800.

33. Jansen, B.J.; Jung, S.; Salminen, J. Creating Manageable Persona Sets from Large User Populations. In Proceedings of the Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland, UK, 4–9 May 2019; Association for Computing Machinery: New York, NY, USA, 2019.

34. McGinn, J.; Kotamraju, N. Data-Driven Persona Development. In Proceedings of the Twenty-Sixth Annual CHI Conference on Human Factors in Computing Systems—CHI '08, Florence, Italy, 5–10 April 2008; ACM Press: New York, NY, USA, 2008; pp. 1521–1524.

35. Xu, Y.; Lee, M.J. Identifying Personas in Online Shopping Communities. *Multimodal Technol. Interact.* **2020**, *4*, 1–19.

36. Salminen, J.; Jung, S.-G.; Santos, J.M.; Jansen, B.J. Does a Smile Matter if the Person Is Not Real? The Effect of a Smile and Stock Photos on Persona Perceptions. *Int. J. Hum. Comput. Interact.* **2020**, *36*, 568–590.

37. Kim, E.; Yoon, J.; Kwon, J.; Liaw, T.; Agogino, A.M. From Innocent Irene to Parental Patrick: Framing User Characteristics and Personas to Design for Cybersecurity. *Proc. Des. Soc. Int. Conf. Eng. Des.* **2019**, *1*, 1773–1782. [CrossRef]

38. Margaris, D.; Kobusinska, A.; Spiliotopoulos, D.; Vassilakis, C. An Adaptive Social Network-Aware Collaborative Filtering Algorithm for Improved Rating Prediction Accuracy. *IEEE Access* **2020**, *8*, 68301–68310. [CrossRef]

39. Kizgin, H.; Dey, B.L.; Dwivedi, Y.K.; Hughes, L.; Jamal, A.; Jones, P.; Kronemann, B.; Laroche, M.; Peñaloza, L.; Richard, M.-O.; et al. The impact of social media on consumer acculturation: Current challenges, opportunities, and an agenda for research and practice. *Int. J. Inf. Manag.* **2020**, *51*, 102026. [CrossRef]

40. Matthews, T.; Judge, T.; Whittaker, S. How Do Designers and User Experience Professionals Actually Perceive and Use Personas? In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Glasgow, Scotland, UK, 4–9 May 2019; Association for Computing Machinery: New York, NY, USA, 2012; pp. 1219–1228.

41. Risse, T.; Demidova, E.; Dietze, S.; Peters, W.; Papailiou, N.; Doka, K.; Stavrakas, Y.; Plachouras, V.; Senellart, P.; Carpentier, F.; et al. The ARCOMEM Architecture for Social- and Semantic-Driven Web Archiving. *Futur. Internet* **2014**, *6*, 688–716. [CrossRef]

42. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.

43. Antonakaki, D.; Spiliotopoulos, D.; Samaras, C.V.; Pratikakis, P.; Ioannidis, S.; Fragopoulou, P. Social media analysis during political turbulence. *PLoS ONE* **2017**, *12*, e0186836. [CrossRef] [PubMed]

44. Chorley, M.J.; Colombo, G.B.; Allen, S.M.; Whitaker, R.M. Human content filtering in Twitter: The influence of metadata. *Int. J. Hum. Comput. Stud.* **2015**, *74*, 32–40. [CrossRef]

45. Salminen, J.; Kwak, H.; Santos, J.M.; Jung, S.-G.; An, J.; Jansen, B.J. Persona Perception Scale: Developing and Validating an Instrument for Human-Like Representations of Data. In Proceedings of the Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; Association for Computing Machinery: New York, NY, USA, 2018.

46. Salminen, J.; Sengun, S.; Kwak, H.; Jansen, B.; An, J.; Jung, S.-G.; Vieweg, S.; Harrell, D.F. Generating Cultural Personas from Social Data: A Perspective of Middle Eastern Users. In Proceedings of the 2017 5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), Prague, Czech Republic, 21–23 August 2017; IEEE: New York, NY, USA, 2017; pp. 120–125.

47. Salminen, J.; Jansen, B.J.; An, J.; Kwak, H.; Jung, S.-G. Automatic Persona Generation for Online Content Creators: Conceptual Rationale and a Research Agenda. In *Personas—User Focused Design*; Nielsen, L., Ed.; Springer: London, UK, 2019; pp. 135–160, ISBN 978-1-4471-7427-1.

48. Margaris, D.; Vassilakis, C.; Spiliotopoulos, D. Handling uncertainty in social media textual information for improving venue recommendation formulation quality in social networks. *Soc. Netw. Anal. Min.* **2019**, *9*, 64. [CrossRef]

49. Margaris, D.; Vassilakis, C.; Spiliotopoulos, D. What makes a review a reliable rating in recommender systems? *Inf. Process. Manag.* **2020**, *57*, 102304. [CrossRef]