My Type, My Style: Digital Library Personas Based on Web 2.0 User Behavior

Authors: Gu Liping

Date: 2016-08-12T00:00:00+00:00

Abstract

To design better interactivity for digital libraries and information services in the Web 2.0 era, creating and using personas is a cutting-edge trend, and it is built upon understanding user behavior. Therefore, we investigated the social software experiences of 497 doctoral and master' s students from the Chinese Academy of Sciences, and established 4 virtual personas based on the survey results through chi-square analysis, factor analysis, and cluster analysis.

Full Text

Preamble

Title: User Does What User Want: Digital Library Persona Based on Web 2.0 Users' Behaviors

Authors: Gu Liping
Library of City University of Hong Kong, Hong Kong
Library of Chinese Academy of Sciences, Beijing 100190

Abstract: In order to design better interaction for digital libraries and information services in the Web 2.0 era, creating and using personas is an emerging trend, which is built upon understanding user behaviors. Therefore, we investigated the social software experiences of 497 doctoral and master' s students at the Chinese Academy of Sciences, and through chi-square analysis, factor analysis, and cluster analysis of the survey results, established four virtual personas.

Keywords: Web 2.0; group; community; user behavior; persona

1. Introduction

The perspective of "social physics" suggests that while individual behavior may not be accurately predictable, when individuals are part of a group, the group' s structure and mutual constraints create collective consciousness and behavior. By studying these structures and constraints, human behavior can often be predicted [1-2]. In most cases, classification by gender, major, and age can effectively determine what characteristics and behaviors certain people might have [3], but is this really true? For digital libraries, if we cannot identify true Web 2.0 groups, we cannot understand user characteristics, behaviors, interaction habits, and predict their needs from the group' s structure and mutual constraints. This means some Web 2.0 services cannot be developed on a reliable basis of user understanding.

This study introduces and empirically validates a novel user research method that combines user surveys with software design—Personas—into the digital library domain.

2. Literature Review and Core Concepts

Alan Cooper defines Persona as "a precise description of users and what users want to accomplish"[4]. In Chinese translations, Persona is rendered as "人物角色" (character role) [4-5]. Current literature on creating and using Personas in the digital library field includes "Using Personas to Understand the Needs and Goals of Institutional Repository Users" published in the D-Lib September/October 2008 issue [6], and "Digital Library Interaction Interface Design Based on Web 2.0 Personas" published in this journal in September of the same year [3]. The former builds user models to guide system construction based on qualitative research, while the latter establishes user models to guide information services and identifies expandable research directions for personas based on quantitative research. This paper continues the latter' s research approach, creating and using personas from a quantitative perspective.

3. Research Design

3.1 Operational Definitions and Research Steps

Based on clear conceptual definitions, the specific operational definitions and research steps are as follows:

Statistical description of data to describe the distribution of observed variables;

Analysis of variance to test whether social attributes (physiological characteristics such as gender and age, and social roles such as education level and major) correlate with user behavior;

Factor analysis to extract factors suitable for clustering;

Cluster analysis to classify users and form persona categories;

Identification of persona characteristics and behaviors;

Finally, depiction of the personas.

3.2 Research Methods and Tools

Based on student data provided by the co-organizer (Graduate School of Chinese Academy of Sciences), a stratified random sampling questionnaire survey was conducted. With support from the National Science Library of Chinese Academy of Sciences, questionnaires were distributed and collected from June 9, 2008 to July 20, 2008. A total of 3,000 copies were mailed out, with 501 valid returns and 4 invalid questionnaires.

SPSS version 12.0 was used for data analysis, employing analysis of variance and factor analysis.

4. Web 2.0 Usage Frequency of Survey Respondents

Usage frequencies are shown in .

5. ANOVA Results

Contingency tables were used for chi-square tests, also known as analysis of variance. The process involved:

Establishing null and alternative hypotheses; H0: assuming two variables are independent; H1: two variables are associated.

Consulting the chi-square distribution table, taking the significance level = 0.05, observing the critical value corresponding to degrees of freedom df in the distribution table, then comparing whether the variance value exceeds the critical value.

If the former exceeds the latter, the H0 null hypothesis is rejected and H1 is established; otherwise, the opposite holds.

Using four variables—gender, age, education level, and major—corresponding to seven usage frequency variables including IM, a total of 28 ANOVAs were conducted. The 28 contingency tables are omitted in this section (but can be extrapolated from and), with only the test results published in .

Major showed correlation with six other variables, but if the "other" category in major classification is removed, the correlation in ANOVA becomes non-significant, indicating that research objects primarily focused on natural sciences may differ from studies at other comprehensive universities. Among other correlation tests, only gender correlated with P2P, and age correlated with Blog. Such few correlations are not suitable as a basis for group division.

6. Factor Analysis Results

This section extracts factors suitable for clustering through factor analysis. First, correlation matrices were calculated for "Instant Messaging (IM)," "P2P," "Blog," "RSS," "Social Bookmarking (Tag)," "Social Networking Software (SNS)," and "Podcast," followed by factor analysis, as shown in .

Bartlett' s test of sphericity yielded a value of 814.265 with significance level 0.000, indicating the data is suitable for factor analysis, as shown in .

Based on , principal component analysis was used to extract communality before factor axis rotation, as shown in .

In factor analysis, principal component analysis was employed. According to Kaiser criterion, two factors with eigenvalues greater than 1 were extracted. After varimax rotation converged in three iterations, the cumulative variance contribution rate was 58.727%, with factor loadings shown in . The rotated factor matrix is shown in .

7. Cluster Analysis Results

The K-means clustering method was used with the following steps:

Step 1: Cluster centers were calculated based on Factor 1 and Factor 2 from , with initial results shown in .

Step 2: Through twelve iterations, optimal center points were calculated, namely: cluster centers, as shown in .

Step 3: After twelve iterations, the positions of cluster centers on the factor coordinate axes are shown in .

Step 4: Based on the four group cluster centers, their corresponding positional distances were calculated, as shown in .

Step 5: The original samples contained after grouping (charts omitted here) and their quantities are calculated (see).

The above five-step K-means cluster analysis yielded cluster centers for four groups and the number of individuals contained in each group. Users were classified through clustering, ultimately forming persona categories. By comparing persona quantities, the most and least populous persona categories were Type 2 and Type 1 respectively, as shown in .

8. Pseudo-Mean and Standard Deviation Comparison

Through comparison of pseudo-means and standard deviations, the degree of Web 2.0 service usage for each user category was determined. The specific method was: based on cluster analysis results, the original 497 samples were grouped, and their means and standard deviations were calculated within each group. Categories were: high usage (values > 2), medium usage (values > 1), and low usage (values < 0), as shown in .

Note: This table treats the categorical variable "usage frequency" as continuous values between 1 and 7. Its mean and standard deviation do not represent true statistical measurements but are used only for intuitive judgment, not numerical calculation.

Instant messaging is the most highly used Web 2.0 service. All four user types commonly use instant messaging, with Type 3 users showing higher usage levels, while Type 4 users use instant messaging services less frequently. Compared with other user types, Type 1 users are more attentive to Web 2.0 services, trying and using each service. Type 4 users have the least contact with every Web 2.0 service.

9. Expected vs. Actual Value Comparison

Through comparison of expected and actual values, persona characteristics and behaviors were determined. Since the goal is to "select" appropriate features to depict personas, the principle of "whether it differs too much from expected values" was adopted. The specific operation was: the proportions of gender, age, education level, and major in were set as expected values, then compared with proportions in . Physiological characteristics and social roles exceeding the overall proportions were selected as persona features for that user type. Through proportion comparison, physiological characteristics and social roles for the four persona types were determined.

Note: This table serves as a reference for intuitive judgment, not for numerical calculation. Its judgment results are adjusted according to actual needs.

10. Persona Depiction

Based on analysis results from , , and , four groups were given virtual names, hometowns, and modified photos to create personas:

Ye Daxiong (pseudonym), male, 24, from Fujian, Master' s student at Institute of Mechanics. Likes almost all Web 2.0 services, especially instant messaging, P2P, and podcasts. Needs comprehensive services—the newer and more novel, the better.

Wang Congming (pseudonym), male, 27, from Jilin, PhD student at Institute of Chemistry. Frequently uses instant messaging, occasionally uses P2P, and sometimes reads blogs. Only needs convenient search and fast transmission network tools.

Chen Yijing (pseudonym), female, 25, from Sichuan, Master' s student at Institute of Genetics and Developmental Biology. Frequently uses instant messaging and P2P, often uses podcasts, and maintains her own blog at www.chenyijing.blog.com. Likes to share her experiences with others.

Lin Fuxiu (pseudonym), female, 26, from Hebei, PhD student at Institute of Electronics. Does not engage with Web 2.0 on a daily basis, only uses email and mobile messages. She is a traditional library user who only uses the network to search for bibliographic information.

Wang Congming is the most important service target, while Chen Yijing is a relatively important service target. Digital library services should prioritize the

needs of these two persona types. Here, the secondary personas (Subordinate Persona)—Ye Daxiong (69/497) and Lin Fuxiu (94/497)—serve to contrast with the main personas (Main Persona)—Wang Congming (187/497) and Chen Yijing (147/497) in terms of behaviors, characteristics, and needs.

11. Conclusion

11.1 Scientific and Practical Significance

Past digital library considerations of user behavior generally followed two directions:

Measurement from user social attributes such as gender, age, major, and occupation;

Measurement from user behavioral attributes such as clicks, browsing, and searching.

Beyond these two approaches, this study develops a third measurement method: using the types and categories of software users habitually employ to describe, explain, or predict user characteristics, behaviors, interaction habits, and expectations. Any library can use this method to establish user models.

11.2 Research Limitations

Questionnaire surveys cannot provide in-depth explanations, which presents certain limitations for illustrating research questions. Additionally, when making research inferences, attention must be paid to the limitations of the survey objects. This study only provides a method for using and creating personas and its specific operational details.

11.3 Future Research

The ultimate goal of Web 2.0 group classification based on user behavior is to more meticulously describe user behaviors, expectations, and needs. Surface-level observed values alone cannot provide complete explanations. Actual interview surveys, tracking observations, and focus groups are needed to verify or revise the actual effects and significance of understanding Web 2.0 users.

Acknowledgments

This paper, based on data analysis, benefited from modification suggestions by anonymous reviewers, for which we express our gratitude.

References

[1] Ball P P. *Critical Mass: How One Thing Leads to Another*. Translated by Bao Yongning. Beijing: Contemporary China Publishing House, 2007.

[2] Buchana M. *The Social Atom: Why the Rich Get Richer, Cheaters Get Caught, and Your Neighbor Usually Looks Like You*. Translated by Ye Weiwen.

Taipei: Commonwealth Publishing House, 2007.
[3] Gu Liping. Digital library interaction interface design based on Web 2.0 personas. *Library and Information Service*, 2007, 52(9): 130-134.
[4] Cooper A. *The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity* (2nd Edition). Translated by Ding Quangang. Beijing: Publishing House of Electronics Industry, 2007.
[5] Mulder S, Yaar Z. *The User Is Always Right: A Practical Guide to Creating and Using Personas for the Web*. Translated by Fan Xiaoyan. Beijing: China Machine Press, 2007.
[6] Maness J M, Miaskiewicz T, Sumner T. Using Personas to Understand the Needs and Goals of Institutional Repository Users. [2008-10-13]. http://www.dlib.org/dlib/september08/maness/09maness.html.

Note: Figure translations are in progress. See original paper for figures.