ANALYSIS OF SEMI-AUTOMATION CLUSTERING TECHNIQUES FOR PERSONA DEVLOPMENT

[1]MD SHAMSHAD ALAM, [2]VINAY KUMAR,

Computer Software, Bharath University, Chennai(India)

sanulacky@gmail.com

vinaykumar.rai97@gmail.com

***ABSTRACT*- Current and future information systems require a better understanding of the interactions between users and systems in order to improve system . Use of personas as design tools is becoming more widespread as researchers and practitioners discover and profit. This paper presents an empirical study comparing the performance of existing qualitative and quantitative clustering techniques for the task of identifying personas and grouping system users into those personas. A method based on Factor (Principal Components) Analysis performs better than two other methods which use Latent Semantic Analysis and Cluster Analysis as measured by similarity to expert manually defined clusters.**

***KEYWORDS*-**

PERSONAS,LSA,FA/PCA,CA,KMS,CALDOL HCI ,GOLD STANDARD CLUSTER.

I.INTRODUCTION

The lack of utility or even failure of many information systems (IS) is well documented and spans the business, psychology, computer science, and management information systems literatures. There are as many dimensions to the problem as there are IS development and design approaches. Since the beginning of modern computing, organizations have continued to struggle with challenges encountered at the nexus of humans and computers as they strive to reap the promised rewards of IS productivity. Given the importance of information systems in our globally networked society, it is no wonder researchers and practitioners pay so much attention to system development, design, and redesign, as well as their consequent impact on system acceptance and use. Information system design is a significant problem for organizations because it affects system use and success. Design teams often collect user data from a number of sources using qualitative and quantitative research methodologies; however, depending on the clustering technique used, either qualitative or quantitative data are used to assign users to clusters for persona development. This paper compares the clustering performance of existing qualitative and quantitative persona clustering methods that lead to persona development.

II.RELATED WORK

A. Qualitative Persona Clustering Methods

Researchers usually cluster participants using manual techniques such as affinity diagrams, card sorting exercises, and expert panels . Researchers advocating the manual approach to clustering prefer the use of rich qualitative data from interviews or observations . Pruitt and Adlin recommend grouping participants by themes and relationships, then determining which ones are most important to the project objectives. Mulder and Yaar suggest a collaborative, qualitative process using various stakeholders gathered "in a room with plenty of whiteboards" to develop clusters based on goals, usage lifecycle, or a combination of behaviors and attitudes. Qualitative versus quantitative data for persona development. Goals, needs, and behaviors are captured in both types of data collection. it has at least two main drawbacks. First, as the number of participants and textual data grows, it becomes difficult for human experts to make objective judgments and trace their findings back to user data . Whereas humans may have qualitative and quantitative data available for making clustering decisions, they tend to focus on qualitative data for insight on a few users at the expense of a broader understanding of users through the analysis of quantitative data . Second, manual clustering methods may require extensive resource commitments from organizations.

B. Quantitative Persona Clustering Methods

Quantitative clustering techniques are capable of determining relationships between multiple variables to find patterns in user data that may be latent, or unobservable to the human eye. The quantitative clustering techniques found in the persona literature are all semi-automated methods because they rely on statistical software to identify clusters of users; the techniques include exploratory factor analysis , principal component analysis (PCA) , and multivariate cluster analysis (CA) . Proponents of quantitative techniques assert that they overcome some of the drawbacks of qualitative clustering: subjective assignment decisions and a lack of rigor, the need for experience in qualitative research training, cognitive limitations of humans, and considerable resource commitments the quantitative clustering techniques can only be found a few times in the literature; therefore, each technique is explained below in general mathematical terms and as implemented by previous authors. Table 1 summarizes the main characteristics of persona clustering techniques. Sinha was the first to propose a semi-automated quantitative technique for persona clustering. His approach implements PCA, a data reduction technique, to identify three components (groups of users) for an online restaurant finder based on 32 dimensions of the restaurant experience .

III.COMPARISON BETWEEN LSA AND CLUSTER TO MEASURED BY SIMILARITY

In the proposed system, empirical study comparing the performance of existing qualitative and quantitative clustering techniques for the task of identifying personas and grouping system users into those personas. A method based on Factor (Principal Components) Analysis performs better than two other methods which use Latent Semantic Analysis and Cluster Analysis as measured by similarity to expert manually defined clusters. PCA reduces the original variables into new components that convey as much of the original data as possible . The author claims his PCA approach provides a direct link from user surveys to the identification of persona clusters while nearly automating the task . The author claims his PCA approach provides a direct link from user surveys to the identification of persona clusters while nearly automating the task . FA is also a data

reduction technique concerned with identifying the latent structure among multivariate data . McGinn and Kotamraju apply FA to identify latent Comparison of Persona Clustering Techniques groupings of system users based on work tasks and demographics. They claim their FA clustering method is "fast and cheap, easy to scope," and results in "data-driven" clusters derived from statistically significant sample sizes . The main limitation of FA as applied to persona clustering is that after identifying clusters, teams still need to conduct interviews or collect qualitative data to explain why users behave in ways identified by FA for subsequent creation of the personas.

IV. SYSTEM REQUIREMENTS

Platform	: Windows Xp,
Front End	: Java JDK1.5.
Back End	: MS SQL 2000
Processor	: Pentium IV
RAM	: 512 MB
HDD	: 80 GB

V.SYSTEM ARCHITECTURE

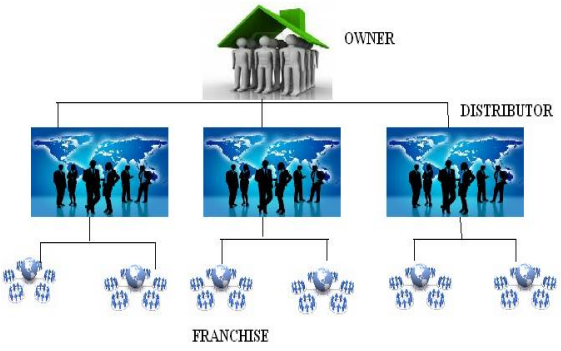

Fig.1 (Architecture digram)

VI.MODULES:

A. PRODUCT OWNER:

The Product is the person who is going to upload the product in the Shopping Server. To upload the product information, the product owner has to enter their username and password. Once it was authenticated, the Owner will enter the product id, product name, amount of the product and the other information.

B. SHOPPING SERVER:

The Shopping Server is place where the product information will be uploaded. Also the shopping server will maintain the Product owner's records and their access details for security purpose. Also the Shopping server will get the amount from the user and take the order for the product that are uploaded in the server. For saling the product the shopping server will get the user information. The Shopping server will also send the product sale details to the product owner in the periodic basis.

C. USER:

To buy the product the from the online shopping server the user have to access the shopping server. To access the shopping server the users have to create an account with the shopping server. For this purpose the user have to enter the username, password, credit card details, phone number of the user.

D. DISTRIBUTORS AND FRANCHISE:

Once the user purchased the product from the shopping server, the information will be send to the corresponding distributor and the product owner. The Distributor will send the product to the concerned franchise and the franchise will deliver the product. To deliver the product, franchise will get a percentage of commission and the Distributor will also get the percentage of commission from the Product Owner.

E. COMMISSION BASED ON THE BUSINESS:

The Franchise will provide business to the Product Owner by marketing more products and also by providing more franchise to the company. For this kind business, the commission will be provided separately for the franchise and as well to the Distributors these information is also stored in the shopping server.

F. ENCRYPTION PROCESS:

For security purpose, we're encrypting the entire transaction details and communication. So that the communication will not be known by the hacker and we can avoid the hacking process.

VII. RESULTS:

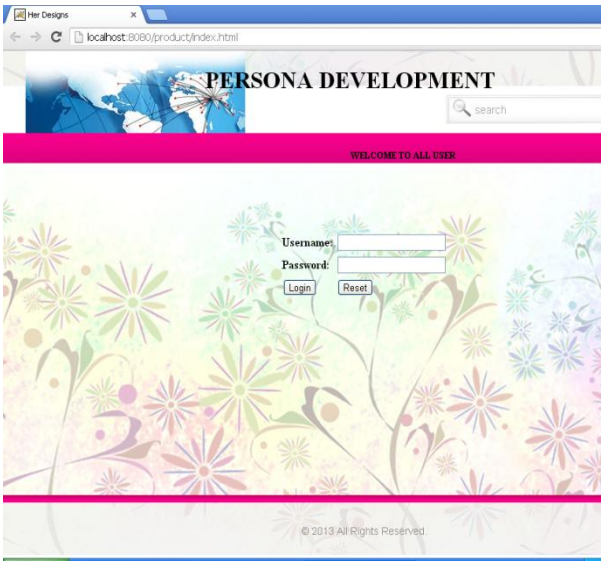

Fig.2- (Login page)

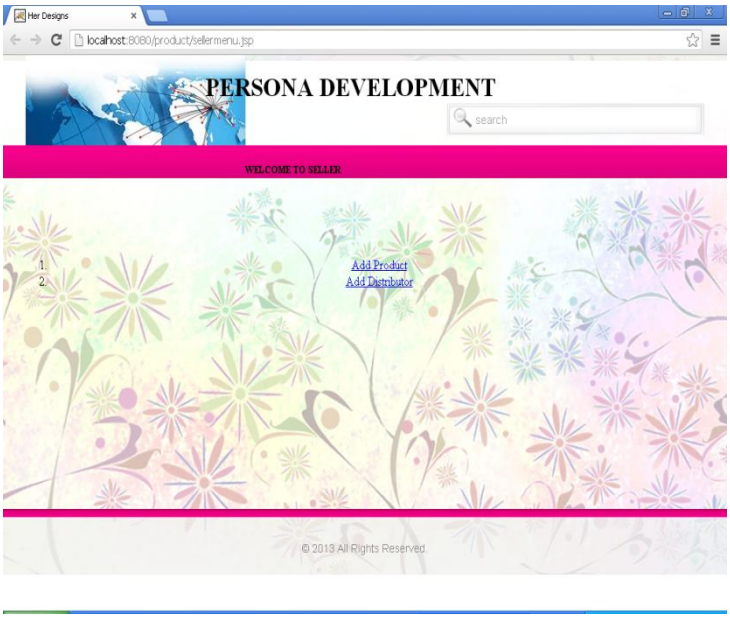

Fig.3-(Select product and distributor)

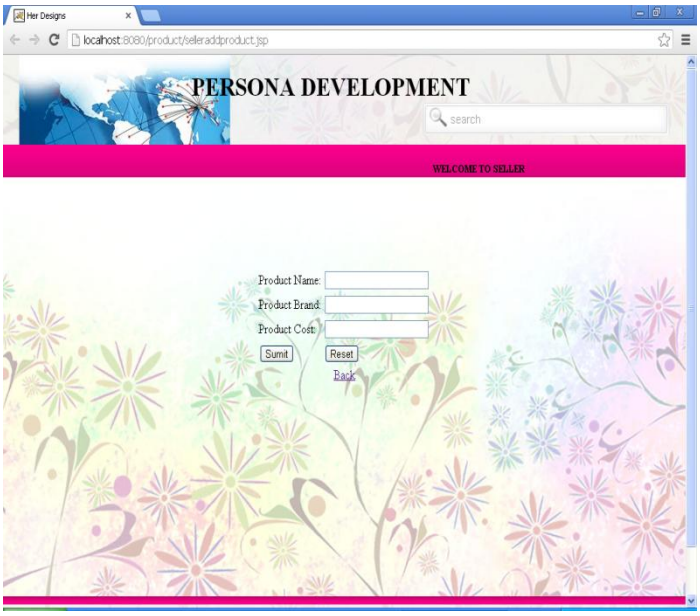

Fig.4 –(Product details)

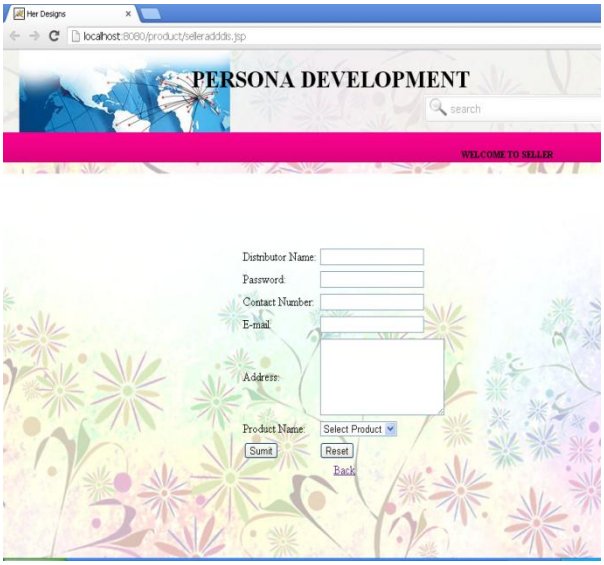

Fig.5 (Distributor details)

Fig.6 (Distributor login page)

VIII. CONCLUSION

As a design tool, personas aim to connect IS designers and users by capturing the goals and needs of groups of users and bringing them to life to evoke empathy and improve communication. A Forrester Research report [46] indicates that systems designed or redesigned utilizing personas experience on average four times the ROI of other systems development projects. However, persona development methods are widely varied and teams need research to determine the parameters and effects of persona design decision such as whether to use qualitative or quantitative data for cluster assignment. In this paper, we identified the differences between qualitative and quantitative methods used for identifying groups of IS users and grouping them into personas to improve system interaction. Additionally, we empirically evaluated the performance of existing semiautomated qualitative and quantitative techniques when compared to manual clustering and found that the quantitative PCA method performed best, but still leaves much room for improvement in clustering methodology. We hope that this work inspires other researchers and practitioners to develop more effective and efficient persona clustering methods.

IX. REFERENCES

[1] M.A. Cook, Build the Enterprise Information Architectures: Reengineering Information Systems. Prentice Hall, 1996.

[2] F.D. Davees, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," MIS Quarterly, vol. 14, no. 3, pp. 319-344, 1989.

[3] V. Sambamurthy, A. Bharadwaj, and V. Grover, "Shaping Agility through Digital Options: Reconceptualizing the Role of Information Technology in Contemporary Firms," Quarterly, vol. 27, no. 2, pp. 237-263, 2002.

[4] V.L. Dreego, B.D. Temkin, and A. McInnes, "Use Personas to Design for Engagementation," Forrester Research, forrester.com/Research/Document/Excerpt/0,7211,45717,00. Html, 2004.

[5] J. Mcc Gin and N. Kotamraju, "Data-Transfer create Persona ," Proc. 25th Ann. SIGCHI Conf. Human Factors Computing Systems, pp. 1521-1534, 2007.

[6] V.L. Drego, B.H. Temkin, "Use Personas to Design for Engagementation," Forrester Research, http://frstweb001. forrester.com/Research/Document/Excerpt/0,7211,45517,00. html,2008.

[7] R. Sinha, "Persona Development for Information-Rich Domains," Proc. Computer Human Interaction Extended Abstracts on Human Factors in Computing Systems (CHI EA '03), pp. 830-831, 2003.

[8] R. Naghshin, A. Seffah, and R. Kline, "Cognitive Walkthrough + Personae = An Empirical Infrastructure for Modeling Software Development," Proc. IEEE Symp. , pp. 239-241, 2003.

[9] N.G. Dixon, N. Allen, T. Burgess, P. Kilner, and Schweitzer, CompanyCommand: Unleashing the Power of the Profession. Center for the Advancement of Leader Development and Organizational Learning, 2005.

[10] B.J. Jansen, I. Taksa, and A. Spink, "Research and Methodological Foundations of Transaction Log Analysis," Handbook of Research on Web Log Analysis, B.J. Jansen, I. Taksa, and A. Spink, eds., pp. 1-15, Information Science Reference, 2008.

[11] J.F. Barrett, H.J. Jarvis, H.N. Macdonald, P.C. Buchan, S.N. Tyrrell, and K.J. Lilford, "Inconsistencies in Clinical Decision in Obstetrics," Lancet, vol. 336, no. 8914, pp. 549-551, 1990.

[12] J. Cohen, "A Coefficient of Agreement for Nominal Scales," Educational and Psychological Measurement, vol. 20, pp. 37-46, 1960.

[13] J.R. Landis and G.G. Koch, "The Measurement of Observer Agreement for Categorical Data," Biometrics, vol. 13, no. 1, pp. 159-
174, 1996.

[14] S. Kalczak and R. Mann, "Utilization and Perceived Benefit for Diverse Users of Communities of Practice in a Healthcare Organization," User Computing, vol. 23, no. 4, pp. 24-50, 2010.

[15] R. Fuller and J.K. de Graaff, "Measuring User Motivation from Server Log Files," Proc. HumanFactorsandtheWeb2Conf.,http://www. microsoft.com/usability/webconf/fuller/fuller.htm, 1996.

[16]S.K.Alag,CollectiveIntelligenceinAction. Manning,2009.