Mining for Culture: Reaching Out of Range

Wanda Eugene
Auburn University
3101 Shelby Center for Eng. Tech.
Auburn, AL 36849-5347 U.S.A.
eugenwa@auburn.edu

Juan E. Gilbert, Ph.D.
Clemson University
821 McMillan Rd.
Clemson, SC 29634-0974 U.S.A.
juan@clemson.edu

ABSTRACT
The goal of this paper is to present a tool that will sustain the development of culturally relevant computing artifacts by providing an effective means of detecting culture identities and cultures of participation. Culturally relevant designs rely heavily on how culture impacts design and though the guidelines for producing culturally relevant objects provide a mechanism for incorporating culture in the design, there still requires an effective method for garnering and identifying said cultures that reflects a holistic view of the target audience. This tool presents culturally relevant designs as a process of communicating with key audiences and thus bridging people and technology in a way that once seemed out of range.

Categories and Subject Descriptors
H.2.8 [**Database Applications**]: Data mining

General Terms
Design

Keywords
Data Mining, Ethnocomputing, Culture, Design

1. INTRODUCTION
Although culture has made its way to the forefront of conversation, constantly a caveat to be accounted for, it still manages to maintain its chameleon image. In every context of its use, it takes on a different meaning or representation, fulfilling a different purpose or function. This lack of uniformity in its functional meaning has become less of a concern than its apparent impact on our changing global society. Yet and still, its increasingly significant presence requires its inclusion in all present and future innovations. Thomas Hughes reflected on the idea that since the beginning of the 20th century, all inventions are fashioned by individuals with a very specific educational and cultural background [10]. He explained that each part of an invention's complex story involves processes that are highly contingent and highly intertwined with social, economic, and political relationships.

Culture and culture identities come from somewhere, have histories, and undergo constant transformation [9]. It can also be said that the nuances of a culture are best understood by its participants. As culture influences action by shaping a repertoire or "tool kit" of habits, skills, and styles from which people construct "strategies of action" [15], it creates culture identities, which are understood within and between the culture participants. These tool kits depict the wealth of knowledge shared and understood within the culture. They also provide an alternative lens of understanding and interpreting data not already associated with one's mental schema. It is suggested that the interaction of culture, affect, and cognition allows a person to develop multiple intelligences, interpersonal intuition and deep knowledge of oneself [11]. However it is also acknowledged that people are unique individuals who belong to several different identity groups [14]. With attempts to impact a broader audience, our efforts converge upon a means to capture a better understanding on these unique culture identities and to bridge people and technology in a way that once seemed out of range.

As of late, there has been several research endeavors purposed for the task of appropriating culture for use in technological design. This paper discusses culture in the context of the design and development of computing technology and further suggests a means for doing culture discovery – the Culture Inquiry Form, which is then introduced and explained. The goal of the Culture Inquiry Form is to present a process that draws upon existing computer science tools such as data mining, and attempts to emulate known methods such as ethnography research, all to serve as a means for better depicting cultures of participation and culture identity.

2. RELATED WORKS
2.1 Culture in the Field
Numerous researchers have studied the use of culture to relate a variety of concepts to diverse learners and to understand subject specific concepts among culture groups. For example, computing as an element within culture can be found in various artifacts and practices within a community of practice. Eglash investigates fractal geometry as in geometric patterns, calculations, and theories, as facets expressed in various African cultures [3]. The use of culture to make connections has been used in several disciplines and domains. For example, Carol Lee uses culture modeling to teach literature. Culture modeling, in essence, provides "instructional organization that makes academic concepts, strategies, and habits of mind explicit that makes ways of engaging in the work of the disciplines familiar and that provides supports for instances where the learner is unsure" [13]. She uses the culture of everyday practices as a lens for understanding the role of perception in influencing actions. Within culture modeling, culture data sets are used, which are

familiar examples that new learning can be anchored and used to provide problems whose solutions mirror the demands of the academic task that we want the learner to discover [13]. Making connections across relevant schemata or clusters of schematic networks helps to create connections between the known and the unknown.

2.2 Cultural Data Mining

Data mining provides a means of transforming large groups of data into information by extracting a pattern and also designates fitting a model to data, finding structure from data, or in general, any high-level description of a set of data [5]. Data mining algorithms' ability to extract patterns from data facilitates a growing need to analyze a subset of data or a model applicable to that subset, within a large data set. As we quickly move from data sets consisting of kilobytes to now terabytes of data, such as that used by the Library of Congress or Wal-Mart, it quickly becomes a daunting task to extract useful information. As computers grow in speed, number-crunching capabilities, and memory, scientific researchers are edging into data overload as they try to find meaningful ways to interpret these data sets [12]. Thinking of the notion of the varying culture identities that exist in our society, data mining offers a means of extracting these unique patterns enumerated from data, as appose to relying upon assumption or sweeping generalizations. For this reason, the idea of cultural data mining is taking root. Manovich observes that until now, the study of cultural processes relied on two types of data: shallow data about many people (statistics, sociology) or deep data about a few people (psychology, ethnography, etc) [2]. Consequently, we can now collect detailed data about very large numbers of people, objects and/or cultural processes and we no longer have to choose between size and depth [2].

2.3 Design of Cultural Relevant Software

Young's Culture Based Model (CBM) reflects a model of culture that evolved from historical and linguistic analyses, in which the findings extrapolated from the analyses reveal a treasure of cultural remnants [16]. The cultural remnants provide an intercultural instructional design framework that guides designers through the management, design, development, and assessment process, while taking into account explicit culture-based considerations [4]. Young's model poses high-level questions to facilitate the big picture of the management of undertaking the design process. However, designers stemming from a technical background, such as computer science, have found this model difficult to navigate for software designers in need of a direct guide to support the design and evaluation of a software artifact.

Cultural Relevance Design Framework assists designers or design teams with creating culturally authentic technology [4]. This framework is designed to uncover the design team's beliefs and biases about their target audience, highlight aspects of about the target audience that might be unknown, and suggest cultural assets that can be investigated to provide building blocks for sound cultural representations [4]. The authors define culture within two dimensions, presenting a wide range of attributes that can be compiled to further capture and illustrate the concept of culture. To guide the design of culturally relevant tools [4] depicted these two dimensions within four themes: Practices, Ontology, Representation, and Tasks. The Cultural Relevance Design Framework is organized such that each of the themes are presented with a definition, an investigative question, and suggested criteria to help the designer explore and better understand the culture of the target audience. The framework provides concrete criteria that correlate to the socio-cultural norms of the targeted group of users [4]. Overall, this framework informs decisions regarding cultural relevance at the onset of the design process as well as a method of evaluating the cultural relevance throughout production processes to help ensure that the goal of a culturally relevant design is produced [4].

3. CULTURE INQUIRY FORM

Much research has focused on the need of culture and more recently how to use culture, but limited research has been proposed on a feasible means for capturing culture in a useful way that can be easily incorporated in design and development. Though the models discussed above address the much-needed frameworks and guidelines to effectively engage in designing culturally authentic technology, still absent from this discussion however, are how to obtain a persons culture identity information. Lee's culture modeling and Eglash's discovery of African fractals came as a result of extensive ethnographic research, that is normally not afforded to a designer engaged in software development. So the challenge becomes how to obtain information, which entails a holistic view of people, where the cultural preferences and differences of the target audience that should influence the design and development of a technology are accessible? Therefore, we present the Culture Inquiry Form (CIF) to serve as an intake of the target audience cultures of participation, and a snapshot of their culture identity.

3.1 CIF

The Culture Inquiry Form (CIF) allows the learner to self identify with the culture(s) in which they participate. CIF collects culture participation information based on "who you are" and "what you do" [8]. The first part, or the "who you are" portion, includes the demographics section of the instrument, which was designed to correlate the data collecting techniques of the U.S. Census Bureau and the Department of Labor, giving a consistent means of measurement. The US Census Bureau demographic categories serve as a model for this study, entailing questions such as age, ethnicity, and gender. The second part of CIF allow us to capture the "what you do" part of the learners' culture of participation. The design of the second part of CIF stems from the Culture Participation Focus Group Protocol and the collected data. The focus group was conducted in collaboration with the Information Management and System Engineering (IMSE) Program in Detroit, Michigan. The IMSE program, under Wayne State's NSF Broadening Participation in Computing Project, is a collaboration of Wayne State, Focus Hope, and several industry partners to support disadvantage students at critical junctures from a GED through the completion of a post secondary degree [1]. The focus group provided a means of understanding the cultures of participation of the targeted audience and their ontology for characterizing their participation in these cultures. The Culture Participation Focus Group Protocol

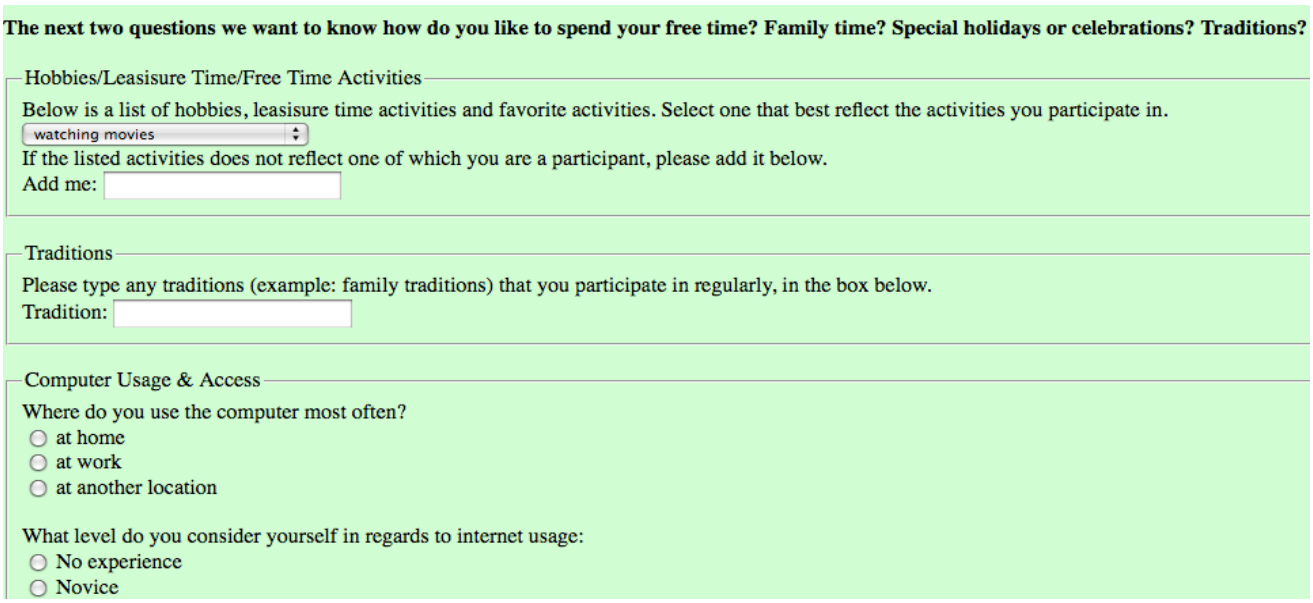

Figure 1. A Sample Portion of the Culture Inquiry Form.

was derived from the focus group protocol designed by the Family Math Team at Stanford University. Thus, the second part of CIF contains selections pertaining to hobbies, employment, and traditions. CIF also entails a field for learners to further describe their participation in said culture(s). If the learner is unable to identify with the listed hobbies another avenue is provided for learners to enter in the cultures in which they participate. A sample portion of CIF is displayed in Figure 1. Also included were questions pertaining to computer usage and perceived level of computer experience. CIF is designed to complement tools such as the Cultural Relevance Design Framework.

3.2 Applications Quest™

The data collected from the CIF is analyzed using cultural data mining, by running a clustering algorithm, Application Quest™, on this data to determine the dominant culture of participation among the participants. Applications Quest™ is a dynamic software tool developed to perform holistic comparisons using hierarchical clustering approach [6]. Applications Quest™ (AQ) takes in numerical values or nominal attributes to determine clusters of similar applications. AQ compares every application to every other application using n C r = n! / [(n-r)! r!], and places the result of each comparison into a database table called the similarity matrix. All numeric attributes are scaled to values between 0 and 1. When considering nominal values, the Nominal Population Metric (NPM) is used. The NPM begins by identifying the nominal attributes within the similarity matrix and then processes them as follows [7]:

1. Compute the total number of combinations for all applications using n C r.
2. Compute the number of unique nominal attribute values.
3. Compute the number of combinations for the unique nominal values using n C r.
4. For those combinations of the nominal attribute value pairs, compute the coverage percentage within the application similarity matrix.
5. The nominal population matrix shows nominal attribute pair coverage across all comparisons. This is an accurate measure of the impact of the nominal attribute value pairs based on their actual existence within the data population. The next step in this process is to adjust the Coverage values if necessary. This is the desired goal when the application is measuring difference vs. similarity.
6. The Coverage values in the nominal population matrix are now the Nominal Population Metrics that can be used in clustering algorithms to accurately compare nominal attribute values.

Using the squared Euclidean distance measure, AQ computes a similarity matrix. To determine the clusters, AQ uses a divisive clustering approach by identifying the two most different applications using the similarity matrix. Using the two most different applications, AQ forms clusters around them based on each individual application's closeness to one or the other.

4. CIF IN PRACTICE

To better illustrate CIF, and further explain the process of identifying cultures of participation, we present the findings of a study recently conducted.

4.1 Demographics

The participants were recruited from the student and faculty population of Auburn University (Auburn, AL). Overall, the study had 104 participants of whom 65% were male and 35% were female. 81% indicated that they were in the age range of 19-24, 16% said they were in the age range of 25-34, and 3% were in the age range of 45-54. Regarding ethnicity, 79% of participants identified themselves as Caucasian, 17% as African American, and 4% as Asian.

4.2 Findings & Analysis

In reviewing the collected data, an approach was selected which grouped attributes such as hobbies and traditions into buckets. All of the questions in CIF were designed as radio buttons or check boxes except for hobbies and traditions. The hobbies question presented the participant with a dropdown list of a several hobbies. These hobbies were those gathered from the focus group study. The participants were also permitted to enter their hobby if it did not appear in the provided list, and it would be added to the list that participants would see when CIF is loaded again. The traditions question was also designed as an entry field, so that participate could enter their traditions and describe them accordingly.

After reviewing the entries for hobbies and traditions, there were several overlaps, thus, it was decided to condense these separate entries into one larger grouping. For example, several entries included sports, football, basketball, and sports in general. All such entries were then identified under the larger category of sports. Another example, watching TV and watching movies, was condensed to the category of entertainment. A similar approach was taken for the traditions attribute. For the traditions attribute, the participants entered more of an explanation of what their tradition entailed. For example, one entry would be "Christmas dinner with the family". In analyzing the entries, we created buckets around the central themes and checked all that apply for each given tradition. Using the example above, holidays, dinner, and family would have been the buckets checked off. The majority of participants in this particular study didn't enter a tradition, and given the amount of variability in the entries, we didn't include it in the analysis. Application Quest™ was then run on the 104 collected responses as shown in the summary in Figure 2.

Upon uploading the data to AQ, the number of clusters (k) that we desired the responses to be grouped into must be determined. As with other clustering algorithms, there is no magic k. So we tried several numbers with the goal that the clusters would remain relevant, where irrelevant entries are not grouped or forced together, and that we don't have several clusters all having one entry. Thus, after several runs and trials, the number of clusters was set to nine.

AU AQ Application Summary (104 applications and 9 clusters)
Difference Index for all Applications 26.97% Standard Deviation 17
Difference Index for Recommended Applications 48.07% Standard Deviation of 14
Age: 19-24 (84), 25-34 (17), 45-54 (3)
CompLevel: intermediate (inter) (58), expert (34), novice (11)
CompUsage: home (89), work (8), another (6)
Education: some college (64), Diploma/GED (16), Graduate or professional (13)
Ethnicity: White (79), Black (18), Asian (4)
Frequency: Daily (96), Weekly (6), (1)
Gender: Male (68), Female (36)
Hobbies: Sports (46), Entertainment (22), Fishing (5)

Figure 2. AQ Summary

Of the nine clusters, 86 participants fell into clusters five, one, and eight. The dominant attributes of those clusters are depicted in Table 1.

Table 1. Table captions should be placed above the table

Cluster #	Number of Applications	Dominate Attributes		
5	44	Age, gender,	Usage, ethnicity, Frequency	
8	22	Usage, frequency	Gender, Level	Age
1	20	Frequency	Level	Gender

The clusters produced by Applications Quest™ provide an accurate and efficient way to determine the cultures of participation. The efforts of determining cultures of participation normally determined after extensive ethnography studies. Thus, AQ tremendously aids in the effort to capture the same essence of a learner in a quantitative approach. In this very basic study, it was easy to see and draw logical conclusions regarding each cluster. In running Application Quest™, industries and traditions were excluded in the specified attributes to simplify the data analysis. The similarities were so few they didn't contribute to the clusters. Given Table 1 along with the data, a designer is provided with fast access to knowledge of the cultures of participation of the participants that lay in the majority clusters.

5. CONCLUSION

The aim of this paper was to introduce a tool and a process for discovering cultures of participation and culture identity. This understanding aids with the inclusion of culture in the design and development of computing technology, such as culturally relevant products. The Culture Inquiry Form uses data mining instead of standard statistical data analysis to determine culture similarities. Standard statistical package can make it difficult to determine the cultures of participation, because they are often grouped based on frequency and give limited information regarding nominal attributes. Using statistics to create content, for example, if only given the information generated in Figure 2, it would not be as clear to detect a majority cluster profile with three similar attributes depicting the participants' culture of participation, especially in terms of large data sets such as culture. Applications Quest™ provides a means for identifying the cultures of participation based on attributes they have selected in common. The Culture Inquiry Form is presented cautiously, for is it not to be viewed as an effort to substitute ethnography research efforts, for it is on this very premise that we have access to the rich data and insight of humanity and culture in the deepest form. Instead, the Culture Inquiry Form is presented as an attempt to bridge the gap and bring together that rich knowledge of culture to the hands of the designers and developers to aid in developing computing technology that reflects the needs of our diverse society.

6. ACKNOWLEDGMENTS

This material is based in part upon work supported by the National Science Foundation under Grant Number CNS-0837580. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

7. REFERENCES

[1] Brockmeyer M. 2007. Broadening Participation in Computing at Wayne State University Retrieved May 20th, 2009. From http://www.cs.wayne.edu/~mab/BPC.html

[2] Digilib. 2009, June 22. Digital Humanities 2009: Opening Plenary by Lev Manovich Retrieved from

http://digilib.bu.edu/blogs/digilib/2009/06/dh09-opening-plenary-by-lev-manovich/

[3] Eglash. R. 1999. African Fractals: Modern Computing and Indigenous Design, Rutgers University Press, New Brunswick, NJ.

[4] Eugene, W. Hatley, L. McMullen, K. A. Brown, Q. Rankin, Y. Lewis, S. 2009. This is Who I am and This is What I do: Demystifying the Process of Designing Culturally Authentic Technology. In HCI International 2009 13th International Conference on Human-Computer Interaction, San Diego, California, July 2009

[5] Fayyad, U., Shapiro. G. P., and Smyth, P. 1996. Knowledge Discovery and Data Mining: Towards a Unifying Framework. KDD-96, 82-88.

[6] Gilbert, J.E. 2006. Applications Quest: Computing Diversity. Communications of the ACM, 49,3, ACM, pp. 99 – 104.

[7] Gilbert, J. E. 2008. U.S. Patent No. 11,053. Washington, D.C.: U.S. Patent and Trademark Office.

[8] Gilbert, J.E., Eugene, W., Arbuthnot, K., Hood, S., Grant, M.M., West, M.L. & Swanier, C. (2009) Culturally Relevant Game Design: A Case Study for Designing Interactive Algebra Lessons for Urban Youth. i-manager's Journal of Educational Technology, 5,3, pp. 54-60.

[9] Hall, S. 1990. Cultural Identities and Diaspora. In Jonathan Rutherford (Ed.), Identity: Community, Culture, Difference, London: Lawrence and Wishart, pp. 222-37

[10] Hughes, T.P. 1993. Networks of Power: Electrification in Western Society, 1880-1930. The Johns Hopkins University Press, Baltimore.

[11] Irvine, J. J., & York, D. E. 2001. Learning styles and culturally diverse students: A literature review. In J. A. Banks (Ed.), Handbook of research on multicultural education (pp. 484-497). San Francisco: Jossey-Bass.

[12] Kamath, C., and Parker, A. 2000. Mining Data for Gems of Information. Research Highlight, Science and Technology Review, September 2000, pages 20-22. UCRL-52000-00-9 Retrieved from https://www.llnl.gov/str/Kamath.html

[13] Lee, C. D. 2007. Culture, Literature, and Learning: Taking Bloom in the Midst of the Whirlwind. New York, New York: Teachers College Press.

[14] Miller F.A. 1998. Strategic Culture Change: The Door to Achieving High Performance and Inclusion. Public Personnel Management. Vol. 27, No.2

[15] Swidler, A. 1986. Culture in action: Symbols and strategies. American Sociological Review, 51, 273-86.

[16] Young, P. A. 2008. The Culture Based Model: Constructing a Model of Culture. Educational Technology & Society, 11 (2), 107–118