On the Integration of Human Computation into Traditional Business Processes

Productivity Games in Microsoft Windows Development

Ross Smith	Dan Bean	Robin Moeur
Director of Test	Program Manager	
Windows Security	Windows Security	2552 E. Roanoke St.
Microsoft Corporation	Microsoft Corporation	Seattle, WA
1 Microsoft Way, Redmond, WA 98052	1 Microsoft Way, Redmond, WA 98052	
425-706-3982		(206) 714-6267
	(425) 703-4238	
rosss@microsoft.com	danbean@microsoft.com	robin.moeur@gmail.com

ABSTRACT

In this paper, we describe how the integration of human computation efforts in the form of a productivity game can achieve results that are not possible or cost effective with traditional business processes.

Productivity Games, as a sub-category of the Serious Games movement, attract players to perform work that humans are good at, but computers currently are not. Although computers offer tremendous opportunities for automation and calculation, some tasks, such as analyzing images, have proven to be difficult and error-prone and therefore lower the quality and usefulness of the output. For tasks such as this, human computation can be much more effective.

We will showcase a real productivity game taken directly from the Windows development process to highlight this integration and its benefits. The "Windows Language Quality Game" encourages native language speakers to perform the job of traditional software localizers and enhances a difficult and expensive business processes with a "serious game". This has resulted in players who enjoy the opportunity to participate and contribute. It has also resulted in a cost-effective way to improve the quality of native language editions of Microsoft Windows.

Categories and Subject Descriptors

H.5.2 [**User Interface**]: Evaluation/methodology
http://www.acm.org/class/1998/

General Terms

Measurement, Design, Economics, Reliability, Human Factors, Languages, Verification.

Keywords

Productivity, Quality, Games, Localization, Language, Cost-Savings

1. INTRODUCTION

The global business challenges of the 21st century require creative approaches and innovative solutions. Traditional methodologies for solving problems are evolving to create hybrid solutions that embrace new collaborative roles for humans and their use of computers. Technology is facilitating these hybrid solutions by enabling a large number of humans to focus on a problem and then easily aggregate their input. This has opened up the opportunity to innovate and creatively solve many business challenges.

One hybrid business solution relies on the use of games to encourage increased participation and productivity from humans. Productivity Games are designed to increase productivity through the use of gaming elements and engaging game play. Play is part of being human and can help bring people together to have fun, work as a group and accomplish a task [1].

Often, this is done within the context of a game. Stuart Brown's research into the concept of play highlights the fundamental elements of human play and showcases the essential roles of trust and community. [2]

A business process can be viewed as a sequence of activities and tasks that are performed to accomplish a specific organizational goal. As we looked at the characteristics of using serious games at work it become apparent that these games were actually variants of business processes. In their August 2008 report, Forrester notes, "the strongest ROI and ultimate adoption will be in serious games that help workers do real work. We are already seeing this with the use of games in product development and collective intelligence, but the real dynamic idea is to pull out the incentive structures and tools of games to boost productivity and employee morale."[3] All of this helped make the case for an increased investment in games.

In a classic statement on the power of working together, Eric Raymond stated in his document The Cathedral and the Bazaar that "Given enough eyeballs, all bugs are shallow" [4]. While this is often the case, the more important challenge for many tasks is how to motivate group participation. If a person gets involved in a software beta program or open source project, they have shown an intrinsic interest in participating. However, if they are not involved in efforts like this, other types of motivation to encourage participation are required. We felt that by designing games that incorporate the fundamental elements of play, people could be enticed to participate. Even better, if the game play was interesting enough to the player, they would be willing to perform productive tasks in order to participate whether they had an intrinsic motivation to accomplish the goal or not. In our experience and game deployments, this has proven to be true.

In this paper, we will look at a productivity game deployed by the Windows engineering team to address a complex software localization problem that could not have been solved in a cost-effective way without massive participation.

2. BASICS of PRODUCTIVITY GAMES

Productivity games are related to crowd-sourcing or human computation efforts, but with some key differences. Similar to recognized crowd-sourcing efforts like Wikipedia, or human computation initiatives such as the ESP Game, productivity games enable employees to have fun participating and feel good about accomplishing productive tasks in the process. The key difference between productivity games and crowd-sourcing is the use of gaming concepts to motivate participation in work-related tasks. The evolution of the ESP game into the Google Image Labeler, and the subsequent production of actual business data for Google is an example of a productivity game.

Productivity games are not a universal solution for every business process or task. Games introduce an alternative incentive system into the workplace as a byproduct of the game architecture and scoring of play. Since the workplace usually already has an incentive system in place – usually in the form of a paycheck, productivity game designers must be careful when, where and how they deploy games that can potentially impact existing incentives and rewards.

2.1 Game Impacts

Work tasks draw upon employee skills that can be grouped into one of three categories: core, unique, or expanding. Employees share "core" skills, such as the ability to type, that may be specific to their industry, but do not differentiate employee A from employee B. Some Employees have "unique" skills that require specialized training or experience. "Expanding" skills are what employees aspire to and acquire over time to help them perform their jobs better.

From an organizational perspective, there are two categories of tasks that relate to the goals of the organization: "in-role" tasks and "organizational citizenship behaviors" (OCBs) [5]. In-role tasks are what Employees are paid to perform. Organizational Citizenship Behaviors are the behaviors that an organization would like Employees to voluntarily do to enhance the workplace culture and environment.

From a productivity games viewpoint, the employee categorization and the organizational classification overlap in a way that can help identify whether or not a game will be successful in modifying behavior and having people "play".

Table 1 illustrates the areas where productivity games can be the most successful. Focusing either on expanding skills in role, or OCB's that require core skills are the best way to ensure the success of the game.

Table 1. Successful Game Deployment

	Core	Unique	Expanding
In Role Behavior			*
Organizational Citizenship Behavior	*		

3. LANGUAGE QUALITY GAME

The Windows Language Quality Game has been a successful

productivity game. It addresses organizational citizenship behaviors by calling on employees within Microsoft to apply their core native language skills to help assess the quality of Windows translation efforts.

The traditional business process uses specific language vendors to perform translation work, and then a secondary vendor to assess the quality. The business challenge has been that, for some languages and locales, finding two independent vendors can be difficult and costly. To address this problem, the Language Quality Game was developed to encourage native speaking populations to do a final qualitative review of the Windows user interface and help identify any remaining language issues. The goal was to ensure a high quality language release and using the diverse population of native language speakers within Microsoft has enabled the pre-release software to be validated in a fun and cost-effective way. The list of Windows languages can be found on Microsoft.com [6]

Table 2. Language Quality Game Pilot Statistics

Players	Tasks	Additional Errors Reported
900	26000	170

Table 2 illustrates the success rate of the initial pilot of the Language Quality Game for Windows XP Service Pack 2.

4. BUSINESS PROCESS CHALLENGES

The Windows Language Quality Game provided a solution to challenging business problems that could not be easily solved through traditional processes.

Software development, particularly at the scale of Windows, requires sensitivity towards cultural and political issues. While language issues like this may not impact the reliability of the application, users may react negatively and seek alternatives. In addition, government purchases can also be impacted by mistakes in language translation. As a result of these risks, it is imperative that the Windows Team develops software in a robust way that eliminates cultural and political defects.

The typical process involves finding two vendors; one to do the translation work, and the other to help with quality assessment. As an example, Galician is the language of Galicia, in Northwestern Spain. Portuguese speakers can understand Galician and sometimes refer to it as a dialect of Portuguese. However, there are cultural and dialectal differences that must be accounted for specifically in the Galician version of Windows.

Translation or geopolitical errors can impact the quality, perception, and sales for a region. In Windows XP, for example, a user can set up a profile by entering details such as their age, sex and number of children. A version distributed in Latin America asked users their gender, giving their options as No especificado (unspecified), varon (male) or hembra (female). Unfortunately in some Latin American countries the term hembra also means …" [7] - has a negative connotation.

5. GAME ARCHITECTURE

The Language Quality Game is built using a SQL Server database of images that are rendered in the game using Silverlight. The Windows International Team uses an automated process to copy

dialog images from the Windows source code into the SQL server database. The dialogs are then augmented with metadata about the language and usage of the image in question.
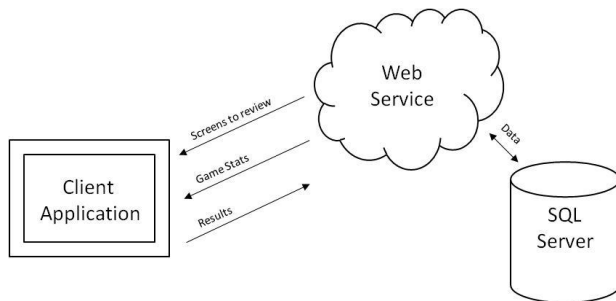
Figure 1 - Language Quality Game Architecture

The dialog images are broken up randomly into groups of 25 to provide multiple "levels" for the player to achieve. As the player works their way through the game, each dialog is presented. The player can use their mouse or a digital pen to circle errors using electronic ink, which is stored numerically as coordinates along with the dialog ID. This not only saves space in the database, but it also improves performance and helps with results reporting.

6. PLAYER POPULATION SELECTION

Finding players to perform the human computation work of reviewing dialogs in the Language Quality Game can be a challenge. It is critical to find native speakers for all the languages supported by Windows versions. For the Language Quality Game, the players are selected using native language speaker social aliases. Invitations were sent via email to groups such as "Persian Speakers at Microsoft"; asking members to visit the Language Quality Game web site and play the game.

Finding the right aliases of potential players was critical to the response rate. We also found that native language speakers typically have friends and relatives who will be using localized copies of Windows. Therefore, it is in the speaker's best interest to play the game and help ensure the quality of the localized version that is important to them.

7. DATA QUALITY AND CHEATING

While it's not possible to completely prevent cheating in a way that scales and keeps people actively participating, it is possible to inject "known defects" and ensure that players find and record them. This helps assess the reliability and validity of an individual player's answers and allows for filtering. In addition, for the Language Quality Game, there is an assumption that as a player works through dozens or hundreds of dialogs, their tendency to cheat diminishes. Furthermore, since there are no significant prizes for "winning", the incentive to cheat is minimized.
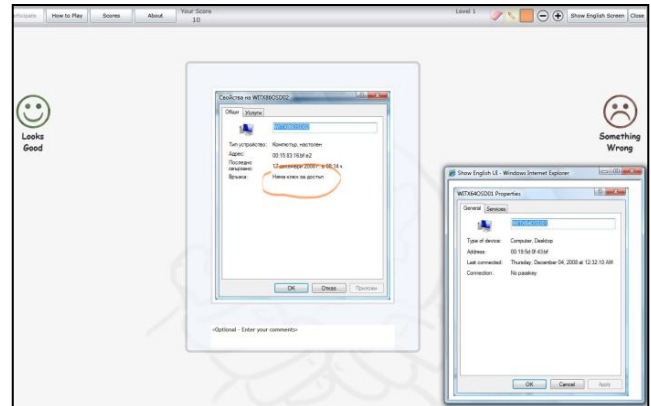
Figure 2 - Language Quality Game Screen Shot

8. GAME ELEMENTS

While the language quality screen review work is not tremendously difficult for native language speakers, it is also not the most interesting or engaging, particularly with a large volume of screens. Consequently, game elements and enticing game play were designed and used to attract players and help motivate them to "play". These are the characteristics of productivity games that help differentiate them from other crowd-sourcing efforts.

8.1 Game Levels

The dialogs are broken up into groups of 25 images and presented as "game levels". Once players review all the images in one level they move to the next higher level and are presented with a new set of 25 images.

8.2 Earn Markup Pen Colors

There are multiple markup pen colors. As a player reviews more and more dialogs, they can earn and use a different color pen.

8.3 Graphical Image Movement

After a player marks up a dialog, they move it to either the "Looks Good" or "Something Wrong" pile. This movement and displaying the next dialog involve some basic Silverlight animation which adds visual interest and a gaming feel to the experience.

8.4 Leader Board

Each person can view a leader board showing all players, their current game level and how many dialogs they have reviewed. Not only does this allow each person to assess their relative effort, but it also provides the basis for some friendly competition. The leader board is divided up in a variety of categories – by language for instance – to encourage participation.

9. LANGUAGE QUALITY GAME RESULTS

There has been 100% language participation – all 36 languages have been sent out for linguistic review and reviews have been received for all of them. The lowest participation was Estonian – 96 reviews, and the highest was Chinese (People's Republic of China) with more than 2600 reviews.

After validation and data quality assessment, an average of 71% dialogs were found to be correct – the highest was Estonian with 93% correct and the lowest was Romanian with 50% correct.

There have been over 900 players, The language with the most had 130 players and the least had 3 players.

Providing review comments is optional, but 29% of the dialogs included comments which identified linguistic issues. The percentage varied across languages.

There were 170 actual bugs identified across all 36 languages.

10. OTHER PRODUCTIVITY GAMES

Microsoft has also tried other styles of productivity games. The most popular were the games used in the Windows Vista Beta program. This experience is covered extensively in chapter 5 of "The Practical Guide to Defect Prevention" [9]. More recently, a Feedback productivity game was created and used to classify freeform text comments as "actionable" or "not actionable". Traditionally, this feedback categorization has been performed manually by the software team and is time-consuming and labor intensive. In some cased, automated machine translation and "text-crunching" tools have been tried with limited success, and still required a human step for final validation.

The strong interest in college basketball tournaments was used to attract potential players. The Feedback productivity game was structured as three phases, one before the tournament started and the other two phases related to subsequent rounds. The goal was to keep game duration short, vary the format, and keep interest levels high.

To participate in the Feedback productivity game, the player had to gain credits by classifying text comments into "actionable" or "not actionable". For each comment classified, one game play credit was received.

The pre-Tournament phase provided each player with random pairs of basketball teams and they could then select the one they thought would win between this hypothetical pairing. Each selection required one comment classification credit.

The next phase of the productivity game had real matchups displayed and the player could then select who they thought would win. Each selection required one comment classification credit.

The final phase of the game focused on the remaining teams in the tournament. To play, each player exchanged four credits for a "team ticket" indicating that team would win it all. Multiple tickets could be obtained for each team and tickets could be obtained for multiple teams. The objective of the game was to obtain tickets for the team that actually won. All players with tickets for the winning team would earn points in proportional to the number of tickets they had.

A total of 150 players participated in classifying 4723 feedback comments and 53% were assessed to be "actionable". These results saved the Windows team a tremendous amount of effort by distributing the work across basketball fans with these core skills.

11. CONCLUSION

In this day and age, many business challenges can benefit from groups of people working together to provide solutions. Recently, crowd-sourcing has been used to distribute tasks out the crowd that can benefit from human computation. This same concept can be utilized in corporations to tackle tasks that they are not resourced to support or require unique skills such as native language proficiency.

A challenge in any of these efforts is how to entice and motivate people to participate. The productivity game concept utilizes gaming elements and engaging game play to help generate that motivation. Through productivity games like the Language Quality and Feedback games, we have shown that people can become engaged in a game and willing to exchange "real work" in order to participate. These results have demonstrated to us the tremendous potential of productivity games to help solve problems that are difficult or impossible to accomplish within traditional organizations and business processes. We look forward to the continued pursuit of that potential.

12. ACKNOWLEDGMENTS

13. REFERENCES

[1] Time Magazine. http://www.time.com/time/specials/2007/article/0,28804,1815747_1815707_1815705,00.html

[2] National Institute for Play www.nifplay.org

[3] It's Time to take Games Seriously, Forrester, August 2008

[4] Eric S. Raymond, "The Cathedral and the Bazaar", 2000

[5] Wikipedia http://en.wikipedia.org/wiki/Organizational_citizenship_behavior

[6] http://support.microsoft.com/kb/292246

[7] http://www.telegraph.co.uk/news/worldnews/1469733/Its-a-tricky-world-in-computers-says-Microsoft-chief.html

[8] www.defectprevention.org or http://productivitygames.blogspot.com

[9] The Practical Guide to Defect Prevention, Marc McDonald, Robert Musson and Ross Smith, Microsoft Press, 2008