ANALYSING QUALITATIVE DATA FROM VIRTUAL WORLDS: USING IMAGES AND TEXT MINING

Brad McKenna Operations and Systems Management, University of Portsmouth, United Kingdom brad.mckenna@port.ac.uk

Michael Myers, Information Systems and Operations Management, University of Auckland, New Zealand m.myers@auckland.ac.nz

Lesley Gardner, Information Systems and Operations Management, University of Auckland, New Zealand l.gardner@auckland.ac.nz

Abstract

There is an increasing interest within both organisational and social contexts in virtual worlds and virtual reality platforms. Virtual worlds are highly graphical systems in which avatars interact with each other, and almost every event and conversation is logged and recorded. This presents new challenges for qualitative researchers in information systems. This paper addresses the challenges of analyzing the huge amounts of qualitative data that can be obtained from virtual worlds (both images and text). It addresses how images might be used in qualitative studies of virtual worlds, and proposes a new way to analyze textual data using a qualitative software tool called Leximancer. This paper illustrates these methods using a study of a social movement in a virtual world.

Keywords: Virtual worlds, qualitative, images, text, leximancer

1 INTRODUCTION

The recent purchase of Oculus Rift, a virtual reality platform, by Facebook for USD $2 Billion indicates the increasing interest in virtual worlds and virtual reality platforms. Companies such as Toyota, IBM, American Apparel, and Reuters (Wasko et al., 2011) are also investing in virtual worlds for business purposes (Messinger et al., 2009). We suggest this increasing interest in virtual worlds presents a new challenge for qualitative researchers in information systems: if we are going to study these virtual worlds, how can the vast amounts of textual and graphical data that characterize virtual worlds be analyzed? Virtual worlds are highly graphical systems in which avatars interact with each other (Castronova, 2007). Not only are images used extensively, but conversations, events, and almost anything that happens online is logged and recorded.

There is literally a flood of qualitative data pouring into the Internet every day, waiting to be interpreted and analysed by the qualitative researcher. Whereas many companies are trying to figure out how it might be possible to mine this huge trove of "big data" for marketing purposes, we suggest there is also an opportunity for qualitative researchers to use this data for research purposes. The purpose of this paper is to discuss the challenges of analyzing the huge amounts of qualitative data that can be obtained from virtual worlds (both images and text). Although the collection of qualitative data from virtual worlds is relatively easy (since in many cases the users of virtual worlds are in effect automatically recording many of their own words and actions themselves), the analysis of this data is more problematic.

Based on our prior study of social movements in virtual worlds, found in McKenna et al. (2012), this paper presents two key issues which emerged during the study. The overarching research problem from the study was to understand how social movements are using virtual worlds. The first key issue we discuss in this paper is to propose a new way to analyze the vast amounts of text in virtual worlds. The second key issue, is how images might be analyzed in qualitative studies of virtual worlds. We

illustrate how images and text can be analyzed using a study of a social movement in a virtual world called World of Warcraft (WoW) a virtual world game, made by Blizzard Entertainment.

2 VIRTUAL WORLDS

A virtual world can be defined as "*a synchronous, persistent network of people, represented by avatars, facilitated by networked computers*" (Bell, 2008). Virtual worlds include social virtual worlds such as Second Life, and gaming virtual worlds, such as WoW. Virtual worlds differ from more traditional social networking technologies because they provide a richer, more immersive experience, and can be deeply engaging (Wasko et al., 2011).

Millions of people have invested their time and energy into using these virtual worlds, creating characters, meeting new people, and engaging in new forms of social interaction. Virtual worlds are of interest to IS researchers for both their business and social aspects (Messinger et al., 2009). Games also provide an interesting research stream. For example, WoW offers an alternative world where social functions, learning, and the development of social skills occur in a virtual environment (Davidson and Goldberg, 2009). Virtual worlds have the potential to become laboratories where experiments in social science can test new norms, values, and institutions (Bainbridge, 2010). Virtual worlds also have the potential to be proving grounds for real world social innovations, cultures, and social movements (Bainbridge, 2009) as well as substituting for social institutions in the real world (Williams, 2006).

Virtual worlds create unique challenges for researchers. One unique aspect of virtual worlds is the field site. The main difference between traditional ethnographic field sites and virtual world field sites is that a virtual world researcher does not physically visit a research site (Kozinets, 2010). Instead, the researcher has to log into the virtual world from wherever they happen to be. This means that online ethnographic research is more like deskwork than fieldwork (Rutter and Smith, 2005). Another unique aspect is the ability for researchers to use hidden observations, without members of the online community even knowing the researcher is there (Guimarães, 2005). Data analysis in virtual worlds also introduces some unique aspects, for example, it may not be possible to determine the true age, sex, or ethnicity of virtual world participants (Kozinets and Kedzior, 2009). People in virtual worlds have the ability to be anonymous or use pseudonyms (Christopher, 2009). Virtual worlds also provide users with the ability to express themselves in ways which extend beyond their body, for example, by modifying the environment around them, altering their identity, and creating multiple avatars (Boellstorff, 2008). They may also inhabit multiple virtual worlds simultaneously.

3 IMAGES

A key difference between traditional and online ethnographic research is that online studies have new types of digital texts . These digital texts include images or photographs (Urquhart and Vaast, 2012). Virtual worlds are highly graphical systems in which avatars interact with each other (Andrade and Arthanari, 2009). Due to their highly graphical nature, images become an important source of data in virtual world studies. We argue that by analyzing images, a researcher can understand how avatars interact with each other and the virtual objects around them.

Screen captures can be very useful for showing the space and inhabitants of a virtual world, while video captures allow researchers to examine how avatars move through space and how their animations appear (Castronova, 2007). However, images are not a common data source in information systems research. In marketing, (Moore et al., 2009) discusses taking screenshots of whatever is on the researcher's computer screen, but he mostly focusses on screenshots with textual data, not images. Images should be captured at key moments of these interactions, for example when the researcher observes something which is theoretically interesting or related to the research question.

(Kozinets, 2010) advocate the use of images as part of the research process. The primary object of inquiry for many visual studies is a set of materials or activities that a scholar finds visually interesting (Andrade and Arthanari, 2009). Such images could be used to allow readers to understand a particular

phenomenon, for example what does a particular area look like, or how did people look a decade ago? (Pauwels, 2011). We believe that presenting images when discussing virtual worlds is absolutely necessary. Virtual worlds differ from the physical world, and the presentation of such images would increase the contextual understanding of the reader.

Another use of images is temporally ordered photographs to show how something has changed over time (Pauwels, 2011). A repeat photograph, or 'rephotograph' is a duplicate photograph used alongside a pre-existing photograph to emphasize certain aspects which may have changed between photographs. The spatial location of the second image is typically repeated, showing the reader the same scene again and inviting comparison (Rieger, 2011). As virtual worlds are constantly changing (McKenna et al., 2012), this becomes important as images can be presented over multiple timeframes to illustrate those changes.

Cartographic (i.e. maps) have been used in the social sciences (Klett, 2011) to represent different territories, illustrate the shape and contour of the land, or to indicate landmarks and features of the area (McKinnon, 2011). Maps have also been used in information systems (see (Spencer, 2011)). As virtual worlds are also virtual landscapes, maps can be used to represent the different virtual locations which exist within that virtual world, much the same as maps are used to represent the geographical locations of the physical world.

4 TEXTUAL ANALYSIS AND LEXIMANCER

Another challenge involved with the study of virtual worlds if the analysis of textual data. With online studies of a virtual world, the amount of data can be huge. One tool which we have found useful for analyzing large amounts of text is Leximancer, which was developed at the University of Queensland, Australia. Leximancer uses machine learning (content analysis) to analyze large qualitative data sets and to display the results in a visual format. There are examples of research using Leximancer in accounting and management (Walsham and Sahay, 1999), conceptual modelling (Crofts and Bisman, 2010), human-computer studies (Davies et al., 2006), risk management (Stockwell et al., 2009), and event management (Martin and Rice, 2007). Leximancer has been evaluated for stability and reproducibility and its results so far have been proven to be reliable.

Leximancer creates visual output in the form of a conceptual map which presents the main themes contained within the text, and information about how those themes are related. The themes are heat-mapped to indicate their importance. Therefore, the 'hottest' (most important) themes appear in red, and the next most important theme in orange, and so on. Leximancer also allows the researcher to extract the actual pieces of text which were used to create the themes.

Content Analysis in Leximancer can be supervised or unsupervised. If using the supervised approach, the researcher will construct a set of key terms (known as concepts) usually with some background knowledge within the domain, or with some theoretical sensitivity. Alternatively, in the unsupervised approach, the algorithm will discover the concepts on its own via reading and re-reading the data. It is the unsupervised approach which is the greatest strength of Leximancer as it is very useful when there is no prior model or set of factors by which to analyze the data (Scott and Smith, 2005). This approach relies on the algorithms in Leximancer to detect the main themes and concepts arising from the data .

The remainder of this paper will introduce the virtual world that we studied, and explain how we analyzed our graphical and textual data using Leximancer.

5 RESEARCH APPROACH

In the original study we decided to focus our exploration on one of the largest social movements within WoW, a Lesbian, Gay, Bisexual, and Transgender movement, hereafter referred to as the LGBT movement. LGBT aims to create awareness for LGBT issues, both in game and out. By early 2013 LGBT has over 7,800 members (players) in WoW and has over 15,000 characters (it is possible for one player to add multiple characters). LGBT was established on a WoW server in October 2006 to

"*better service the LGBT community and offer a safe, inclusive place to game for members of any sexual orientation or gender identity*" (LGBT movement website, 2010).

LGBT has been profiled in a number of gay and lesbian magazines and in a prominent WoW blog website. LGBT is a global social movement with members from many countries. LGBT also maintains a website with discussion forums. LGBT holds many regular activities inside WoW such as an annual pride parade with floats, model competitions, dance parties, group photographs, and events for Valentine's Day. These events are generally organized by the leaders of LGBT, often with input from members via the discussion forum. They also organize member meetings in the real world, and have had meetings in Australia, Canada, and the United States. In (McKenna et al., 2012) we realized that these activities were influenced by patches, which are changes to the game released by the game designers. We followed how these changes to the technical ecosystem influenced the social activities of LGBT. The research question addressed in this study is: how does the technological artifact (the virtual world) and the social world (the social movement) co-evolve?

The lead researcher joined the LGBT movement in WoW and participated in a number of movement activities such as virtual pride parades, dance parties, and group photographs. During the fieldwork, field notes were taken. These notes included digital texts, suggested by Urquhart and Vaast (Davies et al., 2006) to include images. Field notes based on participant observation online took a different form than in traditional ethnographic studies. The lead researcher did not actively take field notes during interactions with LGBT, but rather chose to record details with screen captures. On occasions, LGBT members would meet together and perform social movement activities, such as the pride parade. The lead researcher participated in these activities, for example, by marching in the parade. However, marching in the parade involves controlling an avatar's movements with a keyboard and mouse, which means one's hands are not free for active note taking.

In our study we found that the discussion forum data was also an extremely valuable data source. In some ways it can be argued that the discussion forum data is more "authentic" than interview data. As Myers and Newman point out, the interview is an artificial situation (Palmer, 2013). The interviewee is prompted by a stranger (the interviewer) to answer some questions at a specific point in time, whereas discussion forum data records what people actually thought and posted at the time and is continually updated.

The multiple sources of data collected throughout the research project are listed in Table 1. In total, the lead researcher spent over 1,600 hours engaging with LGBT. The next section of this paper will discuss more specific methodological issues from the original study.

Source of Data	Nature of Data Collected	Quantity Collected	Type
Participant Observation	Screen captures from movement activities.	At least 50 screen captures.	Images
Discussion Forum Posts	Discussion posts from movement website.	128,773 posts dating back to 2006.	Text
Chat Logs	Chat logs from movement in-game chat channels.	Approximately 1.5 years worth of chat logs.	Text
Social movement's website	Textual information relating to background information about the movement and rules of membership.	Approximately 20 pages of information.	Text
WoW Patch Notes	Documents the changes implemented by a patch.	Patches dating back to 2006. 114 patches.	Text
Other WoW websites	Textual information relating to aspects of WoW gameplay.	Not documented but estimated to be over 100 pages of information	Text

Table 1. Data sources used in this study.

6 DATA REDUCTION AND ANALYSIS

The discussion forum data proved problematic due to the large amount of data collected. It was necessary for the researchers to find a way to filter out the data which was not considered important for answering our research questions.

For the analysis of the dataset we first loaded the entire dataset into Leximancer. We soon realized that this was not a good approach as the dataset contained many posts which were irrelevant to the research questions proposed. Therefore we had to find a way to reduce the dataset. To reduce the dataset we read the patch notes of 114 patches in an attempt to discover which patch had an impact on LGBT. After analysis of the patches, we discovered 3 patches which had a strong influence. Second, we had to discover the impact that these patches had on LGBT. Often the patch notes are released before the patch is implemented into the game. Therefore we were able to filter the data from the discussion forum by extracting only those posts made about a certain patch before and after the patch was released. We also performed keyword searches using keywords from our theoretical approaches along with keywords based on our knowledge of the game and the impacts from analyzing the patch notes. Therefore we were able to disregard most of the posts, which gave us a final count of 405 posts. The chat logs also proved to be difficult to extract useful data. Having collected 1.5 years of chat logs, the data set was massive. Therefore we used keyword searches to extract useful text. In total we extracted 10 useful text excerpts from the chat logs.

6.1 Leximancer

Data analysis subsequently involved both the use of NVivo and Leximancer. NVivo is a well-known qualitative software package so a discussion of its use of beyond the scope of this paper. Because the nature of this study was exploratory, the unsupervised approach was used as we were interested to see what emerged from the data. Only the discussion forum data was analyzed in Leximancer due to the large number of posts downloaded.

The entire dataset was first loaded into Leximancer, and the unsupervised approach was executed. This created a set of Leximancer themes which were then filtered out through the researcher's knowledge of the patches which were implemented in WoW. This reduced the dataset to a total of 405 posts which were useful for answering the research question. The reduced dataset was loaded into NVivo for coding manual coding with theoretical sensitivity to actor-network theory (ANT), which is

an appropriate tool for socio-technical research (Callon 1986), and has been advocated for use in information systems research by Walsham (1997). ANT does not distinguish between human and non-human elements, including people, software, hardware, organizations, processes, and treats the social and the technical as inseparable (Walsham 1997). ANT is a useful means for data analysis in this study for its ability to examine the co-evolution of society and technology (Callon 1986). A more detailed discussion of theory choice is beyond the scope of this paper, but can be found in (McKenna et al., 2012). The reduced dataset was also loaded back into Leximancer, and the unsupervised approach was rerun. This created a new set of themes (Figure 1), and knowledge pathways. The text which created those knowledge pathways was then loaded into NVivo and further coded manually with theoretical sensitivity to ANT. Next the codes from both rounds of NVivo coding were compared with each other. This process is illustrated in Figure 1.

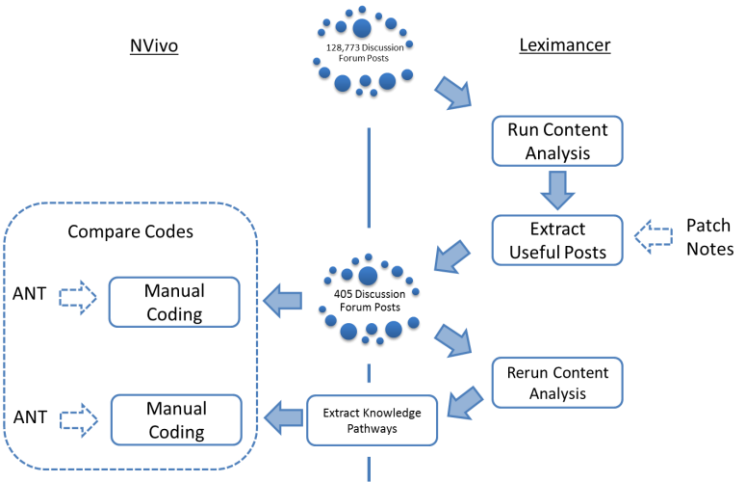


Figure 1. Coding Process

An example of the Leximancer output is illustrated in Figure 2. In the example illustrated, the sub-set of data set was used to extract the common themes evident in the discussion forums when a new patch was released which placed a cap on the size of guilds. On the left hand side is the conceptual map. Each circle within the map represents a theme. We can see from this image that the most important theme was "*guild*", followed by "*members*". Each theme contains multiple concepts (nodes) which make up that theme. The solid line indicates the *knowledge pathway*, which shows the connections between concepts. These pathways were used to empirically link concepts together which provided a very useful way of understanding the data, and analyzing relationships between concepts. The right hand side indicates the actual text extracted from Leximancer which creates the knowledge pathway, i.e. the text which supports the relationship between the concepts *guild* and *Blizzard*. Note that only one knowledge pathway is shown, but can be created between any concepts in the map.

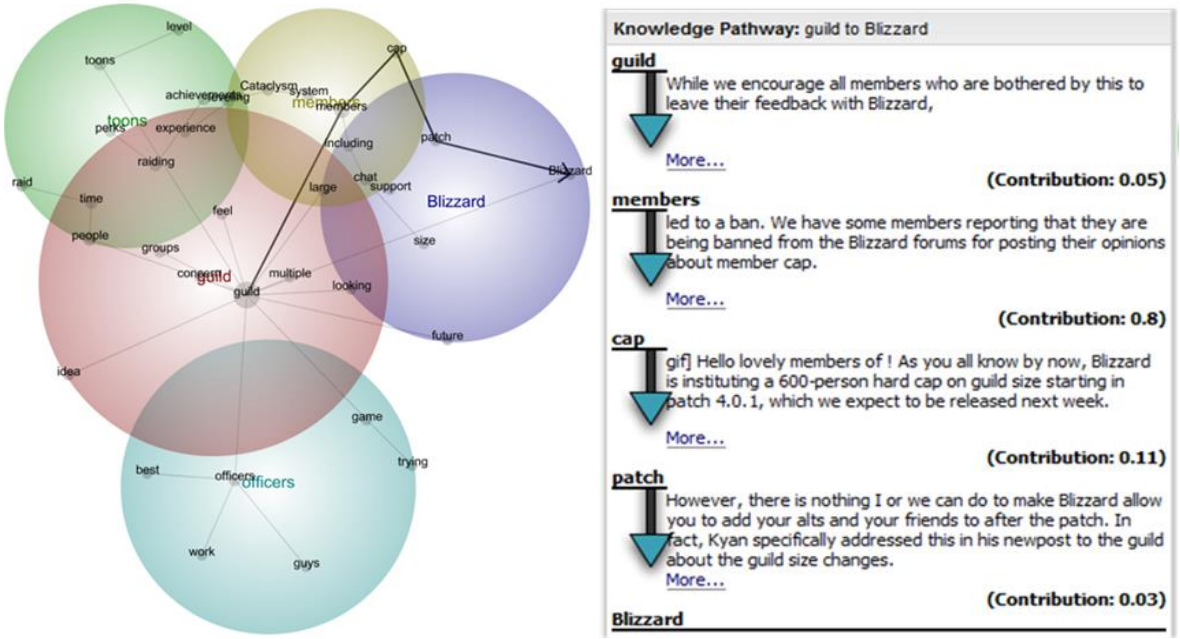


Figure 2. Leximancer Output

6.2 Images

The other main data source was the screenshots (images). We coded the screenshots using NVivo. In the following section, we will briefly review some key points from Andrade and Arthanari (Urquhart and Vaast, 2012) who present three stages for using images in research: image maker, image analyzer, and image presenter.

The first stage of using images in a research report is the information systems researcher playing the role of an *image maker*. Andrade and Arthanari (Myers and Newman, 2007) discuss this role in relation to that of a photographer. However, in our case no camera was needed. The image was taken by the researcher typing a keyboard command on his computer which took a screenshot of the researcher's computer screen (WoW has a built in command for taking screenshots). None of the other players would be aware that the screenshot is taken. These screenshots are saved automatically in a WoW folder within the computer where the game is installed.

The second stage of using images requires the researcher to become an *image analyzer*. In order to interpret an image, the researcher must understand the subject matter and understand the context of the image (Andrade and Arthanari, 2009). In our study images were used to aid the analysis of social movements. For example, the image presented in Figure 3 illustrates players getting their avatars ready to begin the virtual parade. Avatars group together into pre-organized floats, and then begin to form into lines in order to march through the virtual landscape.

Another example of images taken which fits within the context of social movements is illustrated in Figure 4. This image illustrates the actual parade. Here we can see several items of interest to our study. Firstly we can see the parade itself, in which both friendly and enemy players are participating. In WoW, Horde and Alliance players are meant to be fighting each other, but during the parade they come together and fighting is discouraged. In fact the lead researcher had installed an add-on to the game which notifies him when an enemy player is nearby. This add-on is displayed at the bottom right of the image, and is creating many warnings due to many of the parade participants being enemy players.

Players form into groups (floats) before the parade begins

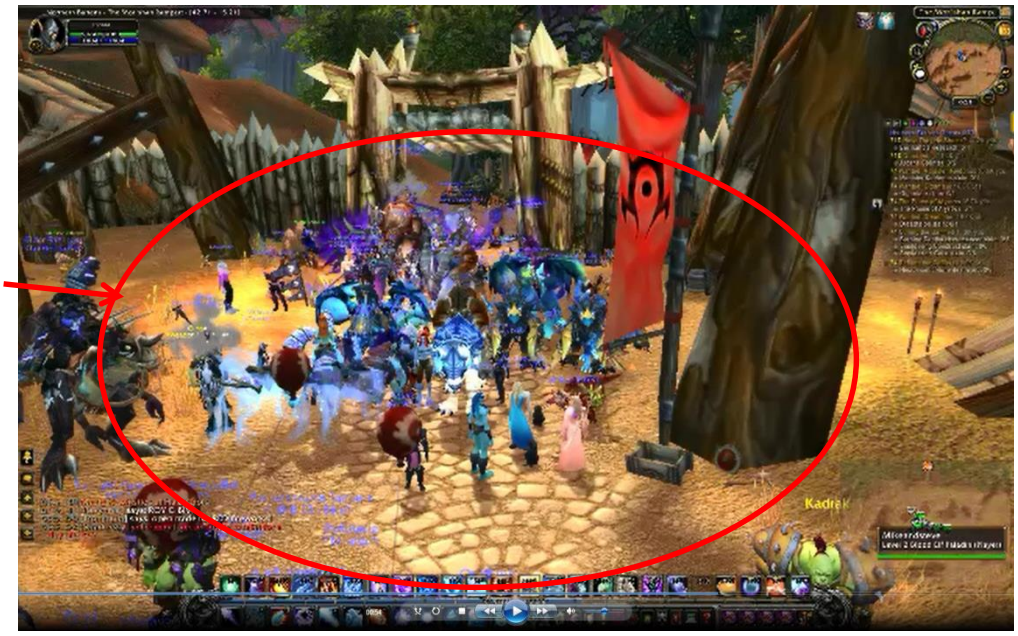


Figure 3. Avatars getting ready to start the parade. Image: ©2004 Blizzard Entertainment, Inc. All rights reserved.

Some parade participants decide to fly above the parade rather than march

Non-members of LGBT observing or participating in the parade.

Both friendly and enemy players marching together in the

Add-on warning that enemy players are nearby. Horde and Alliance marching together.

Figure 4. Virtual Pride Parade. Image: ©2004 Blizzard Entertainment, Inc. All rights reserved.

The image also shows non-members of LGBT who come out to watch the parade. In this image they are standing to the side of the parade, however it is not uncommon for non-members to also participate in the parade. The image also illustrates that not all parade participants choose to walk with the parade, as in the top left of the image some participants choose to fly above it.

The final stage of using images is where the author becomes an *image presenter*, where researchers produce a written report of their findings. Alongside the written text, images can be presented to provide an understanding of a phenomenon which might be difficult to present in textual form (Andrade and Arthanari, 2009).

7 DISCUSSION

According to (Andrade and Arthanari, 2009) "*text and image depend on each other to convey the whole of the content*" (p. 551). The following example illustrates how multiple data sources were used to triangulate our analysis, in order to tell a convincing story.

7.1 Pride Parades

The pride parades were directly impacted by a patch. This example illustrates the use of patch notes, images, chat, and discussion forum data. This patch was implemented in December 2010 and changed the virtual landscape of WoW. Some virtual areas within the game become flooded, while in others giant canyons were formed. The patch notes are illustrated in Table 2. If a reader was unfamiliar with WoW, this would mean absolutely nothing to them. However, if an image was to accompany the text, it would add more value to the full understanding of the context.

Azeroth Shattered: Deathwing's return has had an immeasurable impact throughout the Eastern Kingdoms and Kalimdor. Players will notice drastically altered terrain.

Table 2. Patch 4.0.3a patch notes

Therefore in this example we have used an image to illustrate the changes to the virtual landscape (see Figure 5), or rephotograph as advocated by (Wagner, 2007). In this example we can see the virtual landscape before and after patch implementation. The images at the top are before patch, the images at the bottom are after patch. The image at the top-left shows a peaceful village called Camp Taurajo in the region known as The Barrens which was used as the starting point of the pride parade. The image at the top-right shows a flat plain land where the pride parade used to march through. On the bottom-left, the image shows how Camp Taurajo was destroyed after the patch, and on the bottom-right how the plain land has now become a deep canyon which has split The Barrens into North Barrens and South Barrens. In Figure 6, we illustrate how we used a map (van Leeuwen, 2011) to aid our analysis of how the parade route had to change after the patch was implemented.

We also captured some chat from LGBT members around this event. Note that names have been disguised.

```
[Pradush]: aww poor camp taurajo
[...]
[Hasan]: I'm loving how southern barrens turned out
[Adwardomos]: How does one get from northern barrens to southern
    barrens?
[Sramak]: Go south?
[Zeb]: I had to go all the way around it
[Adwardomos]: To the west or east?
```

This chat was especially important because it illustrates that players were having difficulty crossing the canyon. This was the same route that LGBT had previously taken for their pride parade. Therefore LGBT had to move the parade to a new virtual location for the parade.

Figure 5. Parade starting location before patch (top-left), parade route before patch (top-right), parade starting position after patch (bottom-left), parade route after patch (bottom-right). Image: ©2004 Blizzard Entertainment, Inc. All rights reserved

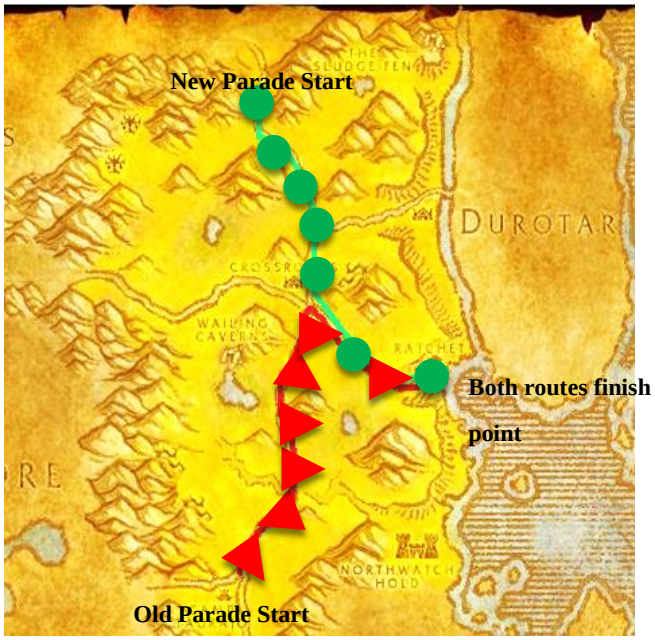


Figure 6. Old parade route (triangles), new parade route (circles). Image: ©2004 Blizzard Entertainment, Inc. All rights reserved.

Together these multiple data sources are able to tell the story of how changes in configurations to a virtual world affect the inhabitants of that virtual world. The next example illustrates a further impact, this time on the size of guilds.

7.2 Guild Size Caps

During beta testing of the expansion pack called Cataclysm, Blizzard became aware that the size of the guilds must be capped at 600 members (later increased to 1,000). Blizzard was implementing new systems to manage guilds, which contained complex systems which track the (gaming) contributions that an individual guild member makes to the guild. Larger guilds would therefore place more strains on these systems. Therefore in order to allow the system to run smoothly, they found that capping the size of guilds was necessary. At the time this patch was implemented, LGBT had approximately 6,200 members. In this example mostly the discussion forum data was used, and NVivo and Leximancer were used for the analysis.

On the left side of Table 3 the Leximancer conceptual map is illustrated which indicates the discovered themes and concepts from the data relating to the guild size caps. This example also illustrates the knowledge pathway (i.e. actual text) which represents the linkage between *guild* and *size*. This pathway connects these two end concepts with the intermediate concepts of *members* and *chat.* On the right hand side the text which creates the knowledge pathway is presented.

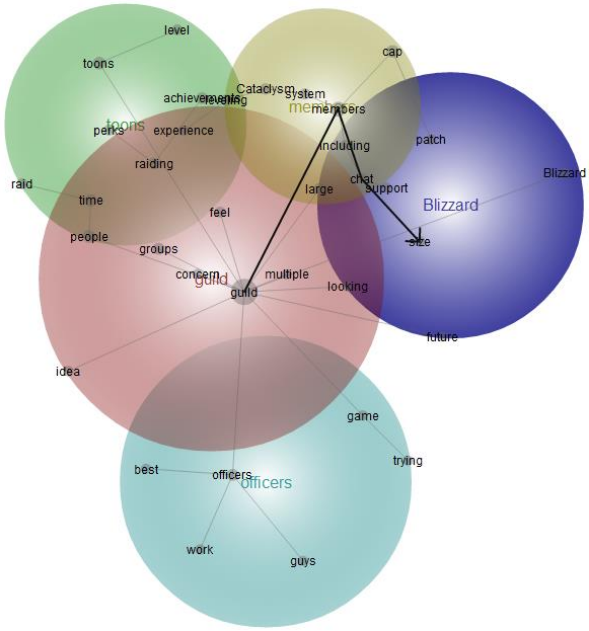
Leximancer Conceptual Map	**Knowledge Pathway**
 Concepts: **Guild** → Members → Chat → **Size**	1. Hello lovely members of LGBT! As you all know by now, Blizzard is instituting a 600-person hard cap on guild size starting in patch 4.0.1, which we expect to be released next week [discussion forum]. 2. Previously, adding players to a guild was as simple as adding people to a chat channel. While approximately 500 members were visible in the UI, there was no real need to limit guild size [discussion forum] 3. While some players have used options including mods and custom chat channels to support large player and guild alliances that number multiple thousands, groups of that size aren't ideally suited to our design philosophy. As always, we're continually looking into adding new features to help facilitate guild management, scheduling, and player communication [Blizzard official announcement].

Table 3. Leximancer Output

We also created manual codes in NVivo based on the discovered Leximancer concepts. The Leximancer output was loaded into NVivo and further coded. This gave us our second round of coding. Table 4 illustrates some of the Leximancer discovered concepts, and the manual coding performed in NVivo from the same text excerpt.

Supporting Text	**Leximancer Concepts Discovered**	**NVivo Codes**
With the advent of the new guild system in Cataclysm we are tracking many more things on each individual player in a guild and in order to support that, we need to limit the amount of members to a reasonable level. The new cap of 600 members is fully supported in the new guild system and that means that everyone will be visible in the UI and able to contribute to all guild functions like experience and reputation gain.	System Support Members Experience Level	Blizzard's reasoning Potential Impact on LGBT Limitations over code
When patch 4.0.1 goes live (most likely next week), we will reactivate the LGBT guild for all future invites, since we won't be able to add any toons to LGBT. We will not be making any other changes or adding any other sub-guilds until we see exactly how the new system works and the results of our advocacy efforts with Blizzard.	Patch Future System Toons (Avatar/Character) Guild	Potential Impact on LGBT Game systems Potential solutions Discussions with Blizzard New sub-guilds

Table 4. Leximancer concepts and NVivo codes

8 CHALLENGES

Our study has presented a number of unique challenges not evident in traditional ethnographic studies. First, huge amounts of data potentially creates a problem of qualitative data analysis. With such a large data set, it becomes necessary for the researcher to weed out data which is not relevant to the study, and to hunt through the data set to find the correct data points relevant to the research question. In some instances this is like finding a needle in a haystack. Our advice to researchers is to have some method to reduce the dataset. For example, we used patches, but other methods could be used such as filtering for research question specific concepts, or theoretical sensitivity.

The second challenge is how to analyze such a large data set. One obvious solution is to use automatic data analysis software. In this study, we realized that coding 128,773 discussion forum posts would be time consuming if using a qualitative data analysis software package like NVivo. Therefore, we decided to use Leximancer, which provides automated qualitative data analysis. Because Leximancer is an automatic text analysis tool, we suggest that researchers also compare the Leximancer outputs with codes created in NVivo. Not only does this give multiple rounds of coding (automatic and manual), but it also ensures that every possible theme of concept is extracted from the dataset.

However researchers need to be aware of the 'garbage in garbage out' problem, which is especially problematic when using live data such as discussion forums. Discussion forum data contains many threads and posts about an unlimited number of topics, and conversations which may start about one particular topic may suddenly go off on a tangent. Therefore we believe that even in the Leximancer approach some filtering is required by the researcher. We found that without this prior filtering, the Leximancer analysis was producing many inappropriate results. It was also necessary to further filter the data after the Leximancer analysis. For example the data from the discussion forums often had HTML tags embedded within it. As these tags are text, they were included in the results. Leximancer contains a pre-defined set of stop words, which are common English language words which will be

skipped over by the algorithm. It is possible to edit the stop word list so the HTML tags were added to the list and the algorithm was run again. This iterative process was completed many times to remove words unnecessary for the analysis. Other words were also added to the stop word list based on the prior knowledge of the researcher. It is claimed that the strength of Leximancer is its unbiased analysis of the data, however this iterative process by definition adds some amount of subjectivity and bias to the results based on the words the researcher decides to exclude from the analysis. Therefore, researchers must still be very familiar with the dataset prior to processing in Leximancer. Researchers should also have background contextual knowledge to ensure that words or not included or excluded from the analysis.

The third challenge is how to take field notes. We decided to capture what was happening in the field by taking screenshots. This requires data analysis similar to that presented in Andrade and Arthanari (Klett, 2011). Because virtual worlds are highly graphical, with avatars interacting with other avatars and virtual objects, we found that that taking screenshots of these interactions with useful. But the researcher must know the appropriate time to take a screenshot. Therefore we recommend that the researcher spend time getting to know the virtual environment and the people and avatars within it. In our study the lead researcher spent over 1,600 hours interacting with members of LGBT. The researcher built up an intimate knowledge of their activities, which enabled the selection of images to ensure they were related to the research question.

The final challenge is combining all of the many different digital texts and analysis (McKinnon, 2011) into the written report to tell a convincing story. We have presented examples of how this can be achieved. It is our hope that our experience in trying to use these alternative data sources might prove valuable for other IS researchers.

9 CONCLUSION

This paper has presented how we analyzed our qualitative data based on a study of a virtual world social. We have discussed how we analyzed images and the huge amount of textual data. As the quantity of textual data in online forms is increasing, we need to find new ways to manage and analyze this data. We have presented our approach to this problem by using a machine learning software tool (called Leximancer) to automatically discover key themes in vast quantities of text.

This study has its limitations. One limitation is that we used only a small subset of the dataset (405 posts) for analysis. Future studies could report on findings from Leximancer with much larger datasets. Because our data was downloaded directly from a discussion forum, it also contained much noise. Other textual based data sources, such as interviews or questionnaires, may contain less noise and would be easier to analyze without the need for a filtering process such as the one we used. Future studies could also look into using object recognition tools to analyze the content found within images.

References

Andrade, A. D. & Arthanari, T. 2009. An Unorthodox Interpretive Approach in Information Systems Research: A Picture is Worth 1000 Words. Proceedings of the Thirtieth International Conference on Information Systems (ICIS), Phoenix.

Bainbridge, W. S. 2009. Etopia. *Networker,* 13**,** 36-35.

Bainbridge, W. S. 2010. Online Multiplayer Games. *In:* Marchionini, G. (ed.) *Synthesis Lectures on Information Concepts, Retrieval, and Services.* Online: Morgan & Claypool.

Bell, M. W. 2008. Toward a Definition of "Virtual Worlds". *Journal of Virtual Worlds Research,* 1.

Boellstorff, T. 2008. *Coming of Age in Second Life,* Princeton, New Jersey, Princeton University Press.

Callon, M. 1986. The Sociology of an Actor-Network: The Case of the Electric Vehicle, in: M*apping theDynamics of Science and Technology,* M. Callon, J. Law and A. Rip (eds.), London: Macmillan Press,pp. 19-34.

Castronova, E. 2007. *Exodus to the virtual world: how online fun is changing reality,* New York, NY, Palgrave Macmillan.

Christopher, T. 2009. In-Game Identities and Meatspace Mistakes. *In:* Cuddy, L. & Nordlinger, J. (eds.) *World of Warcraft and Philosophy: Wrath of the Philosopher King.* Chicago: Open Court Publishing.

Crofts, K. & Bisman, J. 2010. Interrogating accountability: An illustration of the use of Leximancer software for qualitative data analysis. *Qualitative Research in Accounting & Management,* 7**,** 180-207.

Davidson, C. N. & Goldberg, D. T. 2009. *The Future of Learning Institutions in a Digital Age,* USA, MIT Press.

Davies, I., Green, P., Rosemann, M., Indulska, M. & Gallo, S. 2006. How do practitioners use conceptual modeling in practice? *Data and Knowledge Engineering,* 58**,** 358-380.

Guimarães, M. J. L. 2005. Doing Anthropology in Cyberspace: Fieldwork Boundaries and Social Environments. *In:* Hine, C. (ed.) *Virtual Methods: Issues in Social Research on the Internet.* New York: Berg.

Klett, M. 2011. Repeat Photography in Landscape Research. *In:* Margolis, E. & Pauwels, L. (eds.) *The SAGE Handbook of Visual Research Methods.* London: Sage Publications Ltd.

Kozinets, R. V. 2010. *Netnography. Doing Ethnographic Research Online,* London, Sage Publications Ltd.

Kozinets, R. V. & Kedzior, R. 2009. I, Avatar Auto-Netnographic Research in Virtual Worlds. *In:* Wood, N. T. & Solomon, M. R. (eds.) *Virtual Social Identity and Consumer Behavior.* New York: Society for Consumer Psychology.

Martin, N. J. & Rice, J. L. 2007. Profiling Enterprise Risks in Large Computer Companies using the Leximancer Software Tool. *Risk Management,* 9**,** 188-206,.

McKenna, B., Gardner, L. & Myers, M. D. 2012. The co-evolution of the "social" and the "technology" : a netnographic study of Social movements in virtual worlds. 33rd International Conference on Information Systems (ICIS). Orlando, Florida.

Mckinnon, I. 2011. Expanding Cartographic Practices in the Social Sciences. *In:* Margolis, E. & Pauwels, L. (eds.) *The SAGE Handbook of Visual Research Methods.* London: Sage Publications Ltd.

Messinger, P., Stroulia, E., Lyons, K., Bone, M., Niu, R., Smirnov, K. & Perelgut, S. 2009. Virtual Worlds - past, present, and future: New directions in social computing. *Decision Support Systems,* 47**,** 204-228.

Moore, R. J., Gathman, E. C. H. & Ducheneaut, N. 2009. From 3D Space to Third Place: The Social Life of Small Virtual Spaces. *Human Organization,* 68**,** 230-240.

Myers, M. D. & Newman, M. 2007. The Qualitative Interview in Is Research: Examining the Craft. *Information and Organization,* 17**,** 2-26.

Palmer, P. D. 2013. Exploring attitudes to financial reporting in the Australian not-for-profit sector. *Accounting and Finance,* 53**,** 217-241.

Pauwels, L. 2011. An Integrated Conceptual Framework for Visual Social Research. *In:* Margolis, E. & Pauwels, L. (eds.) *The SAGE Handbook of Visual Research Methods.* London: Sage Publications Ltd.

Rieger, J. H. 2011. Rephotography for Documenting Social Change. *In:* Margolis, E. & Pauwels, L. (eds.) *The SAGE Handbook of Visual Research Methods.* London: Sage Publications Ltd.

Rutter, J. & Smith, G. W. H. 2005. Ethnographic Presence in a Nebulous Setting. *In:* Hine, C. (ed.) *Virtual Methods: Issues in Social Research on the Internet.* New York: Berg.

Scott, N. & Smith, A. E. 2005. Use of automated content analysis techniques for event image assessment. *Tourism Recreation Research,* 30**,** 87-91.

Spencer, S. 2011. *Visual Research Methods in The Social Sciences,* Oxon, UK, Routeledge.

Stockwell, P., Colomb, R. M., Smith, A. E. & Wiles, J. 2009. Use of an Automatic Content Analysis Tool: a Technique for seeing both Local and Global Scope. *International Journal of Human Computer Studies,* 67**,** 424-436.

Urquhart, C. & Vaast, E. 2012. Building Social Media Theory from Case Studies: A New Frontier for IS Research. *Thirty Third International Conference on Information Systems (ICIS).* Orlando, Florida.

Van Leeuwen, T. 2011. Multimodality and Multimodal Research. *In:* Margolis, E. & Pauwels, L. (eds.) *The SAGE Handbook of Visual Research Methods.* London: Sage Publications Ltd.

Wagner, J. 2007. Observing Culture and Social Life: Documentary Photography, Fieldwork, and Social Research. *In:* Stanczak, G. C. (ed.) *Visual Research Methods: Image, Society, and Representation.* Thousand Oaks, CA, USA: Sage Publications.

Walsham, G. & Sahay, S. 1999. GIS for district-level administration in India: problems and opportunities. *MIS Quarterly,* 23**,** 39-66.

Walsham, G. 1997. Actor-Network Theory and Is Research: Current Status and Future Prospects, in: Information Systems and Qualitative Research, A. Lee, J. Liebenau and J. DeGross (eds.), London:Chapman and Hall, pp. 466-479.

Wasko, M., Teigland, R., Leidner, D. & Jarvenpaa, S. 2011. Stepping into the Internet: New Ventures in Virtual Worlds. *MIS Quarterly,* 35**,** 645-652.

Williams, D. 2006. Why game studies now? Gamers don't bowl alone. *Games and Culture,* 1**,** 13-16.