Quantitative Content Analysis of Information-Seeking Behavior in Movies, 1950-2010:

A Research Design

Andrew Leigh-Bullard

Kellie Meehlhause

Tiffany Newton

Elizabeth Tschetter

LI 810 Research in Library and Information Science

Emporia State University

Keywords: Technology, Movies, Film, Information Seeking,

Information Technology, Information Behavior

Purpose of Research

The purpose of our research study is to determine the most frequent information seeking methods used by characters in movies and whether those behaviors have changed over time, especially in terms of technology. To date, there is little research that examines information seeking behavior in movies, much less a longitudinal study of how this behavior has changed or been influenced over a specific period of time. Existing research indicates that individual movies, such as *The Big Lebowski* or *Dogma*, illustrate information seeking as being self-defined and intrinsically linked to one's social and self-awareness (Dill & Janke, 2010; Drumheller, 2005). Though these observations provide valuable insight into how some people, albeit fictional ones, fulfill their information needs, they also fail to explore whether these same behaviors are present in other films or if they are case-specific. Moreover, research to date lacks an empirical examination of how advances in information technology (i.e., computers, the Internet, social networking, etc.) within the last sixty years have aided movie characters' information-seeking endeavors, if at all.

This project will use a quantitative framework in order to properly document specific information seeking tools/methods and their frequency of use. It will also allow us to observe and compare whether different behavior was present at different time periods--for example, was a computer more likely to be used to find information in 2002 as opposed to 1963? While completing this project, we will consider the following questions:

- How has information seeking been portrayed in movies over the last 50 years?
- How has information seeking in movies changed in the past 50 years?
- Is there a difference in the way characters search for information in the movies based on when they were made?

- Is using technology to search for information more prevalent in recent movies?

Though more theoretical than practical in application, this project will benefit library science scholarship by emphasizing the complexity of information seeking behavior and the various tools people have and continue to use to make sense of something. Libraries may then use these observations to better understand patron perceptions of certain information tools as well as how to best find and interpret knowledge. Finally, because movies can often be both influential on and reflective of the society it portrays, this project will also benefit film criticism by providing a historical overview of how movie characters find and make sense of information.

Research Paradigm

The backbone of the study hinges on quantitative research methods. This approach was chosen because it will enable better observation of frequency patterns in our samples over time. Bryman (2008) defines quantitative research as "entailing the collection of numerical data" (p. 140), such as statistics and percentages, in contrast to qualitative research's emphasis on word data. Because of its emphasis on measurement over interpretation of social variables, this strategy often centers around deductive reasoning, wherein theory or a hypothesis guides one's research (Bryman, p. 4). Such logic is relevant to this study because it allows researchers to begin by considering ideas related to information seeking behavior technology, and movies, and develop from those observations, the hypothesis that the more recent a film, the more likely it is to employ technology to find information. The study will then test the validity of this theory through the research design described below.

Moreover, the combination of quantitative research and content analysis is already common for the study of film, television, and media in general. Within the last fifty years, the

same aforementioned research methods have been implemented in numerous studies within the fields of mass communications, social and behavioral sciences, political science, and marketing. The scope of such research ranges from TV coverage of the September 11 terrorist attacks (Reynolds & Barnett, 2003), to photographs of African Americans in magazines (Lester & Smith, 1990), from children's roles in television advertising (Riffe, Goldson, Saxton, & Yu, 1987), to female tobacco use in movies (Escarmilla, Cradock, & Kawachi, 2000). The successful techniques used in these studies subsequently establish a precedent for our own research.

Epistemological Issues

For this study, there is no clearly defined epistemology. Mostly clearly linked to quantitative research, positivism is defined as "an epistemological position that advocates the application of the methods of the natural sciences to the study of social reality and beyond. . . . The role of research is to test theories and to provide materials for the development of laws" (Bryman, 2008, pp.13-14). While some aspects of natural science methods, such as the creation of a testable hypothesis, are applicable to our research design, it doesn't address all of the issues and components present in this study. On the other hand, phenomenology, an anti-positivism philosophy centered on "the question of how individuals make sense of the world around them" (Bryman, p.15), does relate to the interest in the tools used for information seeking behavior in the movies. However, it also requires an interpretation of people's "actions and their social world from their point of view" (p. 16) that is beyond the capabilities of quantitative research. Because the content analysis methodology is meant to record and measure the numerical frequency of certain behaviors, it limits researchers from further analyzing why or how a movie character

chooses one tool or method over another. Thus, both positivist and phenomenological epistemology are integral to this research, albeit indirectly.

Ontological Issues

Though not a primary feature of quantitative research, constructionism provides the necessary ontological approach to our research. Bryman (2008) defines this position as "assert[ing] that social phenomena and their meanings are continually being accomplished by social actors" (p. 19). In other words, rather than being constrained and acted upon by society, constructionism posits reality as being in a constant state of change, of "construction and reconstruction" (p. 20). Likewise, our study theorizes that information seeking behavior in movies has experienced continuous change in the last sixty years, especially in terms of technological usage. The focus on movies, too, is due to their documented history of influencing society and taking an active role in redefining social phenomena (McQuaill, 1979).

Research Design

Introduction

To obtain the desired data, this research will use a quantitative content analysis coding sheet to analyze the highest grossing films of randomly selected years between 1950 and 2010. The design of this project involves discussion of sampling procedures, the units of analysis, the instrument used, the permissions needed, and the parameters of this study.

Sample

There are 61 years between 1950 and 2010, inclusive, with many films produced in each year. Thus in the interest of completing this study in a timely fashion, a manageable and representative sample needed to be selected. This was accomplished by limiting the scope, setting a selection interval and taking a random sample.

Rather than consider all movies released between 1950 and 2010, the scope of this study was limited to the highest grossing film in the US each year. For each year this film was determined using the American Movie Classics Filmsite All-Time Box-Office page ("AMC all-time top," 2012), which was confirmed by the Third Millenium entertainment Box Office website to ensure that the films selected were the highest domestic grossing ("TMe box office,").

The next step was to establish a sampling interval. The factors in this decision were the time it would take to carefully analyze each item and the need to ensure that any changes in information seeking behavior could be documented over time. With these factors in mind a sampling interval of three was selected, which will result in twenty units for analysis.

The final consideration was randomness. A random starting point was necessary as 61 is a prime number and it was impossible to select a sampling interval that would cover the entire sample. Thus rather than starting at 1950, a random number was selected using the Research Randomizer website using a range between 1 and 61 (Urbaniak & Plous, 2008). This process resulted a starting point of 57, which in a chronological list of years from 1950-2010 places the first sample as 2006. The remaining sample years are 2009, 1951, 1954, 1957, 1960, 1963, 1966, 1969, 1972, 1975, 1978, 1981, 1984, 1987, 1990, 1993, 1996, 1999, and 2002.

The ultimate step was pairing the randomly selected years, with their highest domestic grossing film. To facilitate observation of the changes over time, the random samples were

arranged in chronological order from 1951 to 2006. In chronological order the first sample from 1951 is *Quo Vadis*. The remaining samples will be *Rear Window* (1954), *The Bridge on the River Kwai* (1957), *Swiss Family Robinson* (1960), *Cleopatra* (1963), *Hawaii* (1966), *Butch Cassidy and the Sundance Kid* (1969), *The Godfather* (1972), *Jaws* (1975), *Grease* (1978), *Raiders of the Lost Ark* (1981), *Beverly Hills Cop* (1984), *Three Men and a Baby* (1987), *Home Alone* (1990), *Jurassic Park* (1993), *Independence Day* (1996), *Star Wars: Episode I: The Phantom Menace* (1999), *Spider-Man* (2002), *Pirates of the Caribbean: Dead Man's Chest* (2006), and *Avatar* (2009). The top film for 1966 was disputed between the two sources. AMC listed three options for that year (*Hawaii, The Bible,* and *Who's Afraid of Virginia Wolf)* without listing what they grossed. TMe listed *The Good the Bad and the Ugly* as the highest grossing at 19 million dollars. AMC was chosen as the more reliable source due to its industry recognized name and the first film listed for that year was presumed to be the highest grossing, thus *Hawaii* was selected for 1966.

Units of Analysis

The primary question of this project pertains to information seeking behavior, which is the overall unit of analysis. To facilitate analysis, information seeking behavior will be subdivided into information need, patron, resource consulted, and resource categories.

Officially information need is defined as "a recognition that your knowledge is inadequate to satisfy a goal you have" (Case, 2008). This project will define an information need as any information a character needs or desires but does not immediately know. For coding purposes the information need will be limited to 1 to 15 words in length, and may be generalized (i.e. from 1993 Smith Way to address) to facilitate comparisons. Ideally the categories of

information needs would be laid out in advance, however this would likely result in missed information needs that did not fit into the categories. When a reviewer is uncertain as to if an information need can be generalized, they will list the full need, and place the suggested generalization in parenthesis next to it for the research team to discuss once data collection is completed.

As defined in the *Dictionary of Library and Information Management,* a patron is "a person or group that encourages and supports an activity, sometimes with money" (patron, 2006). Oftentimes this term is used to reference those who use the library. While this analysis will involve information seekers who are not patrons of any specific resource, this term will be expanded for this project to refer to the individual or character with an information need. The more precise term "information seeker" was considered and rejected due to its length and potential for confusion with the information need. The term patron has the advantage of being brief and easily identifiable as pertaining to a person or character. The patron will be recorded by the character's profession (i.e. teacher, police officer, child, adventurer, etc.).

A resource is "information in a variety of formats that is useful and available" (resource, 2006). This project will list the resource consulted for the separate information needs. If more than one resource is examined for an information need, the resources will be listed in the order they were considered with a semicolon between each resource. This section is to be as detailed as possible, including the specific resource if one is shown, i.e. "*Black's Law Dictionary*" rather than "dictionary".

The resource consulted column will contain the required information for in-depth analysis of the types of places characters go for information. In order to facilitate rapid, quantitative analysis the resources need to be divided into logical sub-categories which cover

most, if not all, places resources might be found. The primary division will be between human resources (HR) and technological resources (TR). Each sub-category will have their own cell on the coding sheet, along with a third cell for specific HR/TR pairs where a technological resource is used to reach a human resource.

The first subcategory will be human resources. Within this category there are casual human resources (CHR), information professionals (IP) and other professionals (OP). CHR includes friends, family, acquaintances and others who are not information professionals. Thus CHR is the catch-all for every human resource that does not fit into the following two categories. The second category, IP, will consist of librarians, archivists and others for whom information work is their primary field. The final human resource is OP, which covers those like lawyers and doctors who have specialized knowledge in a field. By using causal human resources to include everyone who isn't an information professional or other professional these three categories should cover the primary types of people that a person would see as an information resource.

The second subcategory is technological resources. Technological resources includes books (Book) which is any bound print object including books, periodicals, encyclopedias, and dictionaries. The second technological resource will be other print media (OPM) which includes fliers, posters and any other print format that is not bound. The third is contact resources (CR) such as telegraphs, telephones and e-mail that place the patron in contact with a human resource. A fourth technological resource focuses on locations, such as archives and museums (A/M) and would also include library buildings or other places that are considered a resource. The final technological resource will be computer resources (CompR) which includes databases, the internet and any other information source accessed via a computer. The CompR will have an

additional designation of mobile, indicated by an asterisk when the computer is a mobile device ,
such as a web-enabled phone or tablet (ex. CompR*).

In situations where multiple resources are consulted they will be listed in the order they
are used with a semicolon separating them. If a technological resource is used in conjunction
with a human resource both categories will be relisted in a separate column using the format of
"(human resource)/(technological resource)."

To answer the questions of information seeking behavior in the movies, a very broad
topic must be broken down. This is done by recording each information need in a compact form,
considering who the patron is, what specific resource they consult and categorizing the consulted
resources to enable rapid broad analysis.

Instruments

This study will be conducted using a coding sheet designed for this project by the
research team. It will include space to identify what film is being evaluated, what year that film
represents, who is completing the coding and the units of analysis to be used. Each heading will
include either directions or an example or both of what information is needed to complete that
section. The coding sheet will be pre-tested on a random movie that is not in our sample to
ensure that it measures what it is supposed to and that it is possible to fill out accurately during
the film.

Permissions

As this project does not involve human subjects there is no requirement to obtain
permission from the institute review board. Additionally this project does not impede on any

location, department, or individual so no permissions are needed. All films will be watched in private settings to avoid copyright violation.

Parameters

This project began with the questions of: how has information seeking been portrayed in movies over the last 50 years? How has information seeking in movies changed in the past 50 years? Is there a difference in the way characters search for information in the movies based on when they were made?

Is using technology to search for information more prevalent in recent movies?

The first question, how has information seeking been portrayed in the movies over the last 50 years will be evaluated by examining what resources were used in each film and comparing them to the other films from the different periods. This will provide an easy to work with overview for what has occurred in each film.

The second question, how has information seeking in movies changed over time pertains directly to our hypotheses. The null hypothesis (H0) states there is no correlation between the age of a movie and the number of times technology is consulted as an information source. Our active hypothesis (H1) is that there is a negative correlation between the age of a movie and the number of times technology is consulted as an information source. This means that the older a movie is, the fewer times technology will be consulted as an information source. This will be proven by comparing the categories of resources used in each film to the other films to note whether there was no change, a negative correlation, or a positive correlation between the age of the movie and the number of times technology is consulted as an information source.

The third question is about the difference in the way characters search for information over time, and the fourth question, whether technology use is more prevalent in recent films, are answered in the same manner as the previous two questions, by comparing categories of resources used in the movies.

While outside the scope of this study, the instrument used also provides the option for further studies relating to the nature of information resources used by certain types of patrons and for comparing what information needs are portrayed over time.

Conclusion

This project will use a quantitative content analysis coding sheet to analyze the highest grossing films of randomly selected years between 1950 and 2010. The sampling procedure used is a random sample, using a sampling interval of three years to return twenty films for analysis. The specific film for each year was chosen based on their highest domestic gross income, as measured by two websites for confirmation. The primary unit of analysis is information seeking behavior, which is evaluated by considering the information need, the patron seeking information, the specific resource used and a subcategory to facilitate rapid analysis of the wide-variety of resources available. The instrument for this research is a coding sheet written by the design, which will be tested on a random film. There are no permissions needed due to the nature of this design. This study will answer all questions we raised in the literature review, while providing data that could enable further studies. The design of this project involves discussion of sampling procedures, the units of analysis, the instrument used, the permissions needed and the parameters of this study.

Ethical Considerations

This study does not involve the use of human participants, and therefore ethics approval will not be needed from Emporia State University's Institutional Research Ethics Board. All records will only be seen by the researchers and their advisor. The research data will be kept safe in password protected file. Effort will be made to analyze the data as objectively as possible. We are open to questions and concerns related to the data collection, and have no reason to withhold information. No covert methods will be used.

Limitations

Though steps were taken to reduce the limitations such as sample size, the content analysis aspect of the research can have several limitations for our research. The observer could miss details. To compensate for this, there would be at least two observers watching each movie at the same time. There is also quite a bit of interpretation that must be done by observers which leaves room for errors.

This study is focused on top grossing films in the box office for each year. This has the potential to be flawed because, in the case a movie brings in a large amount of revenue in a given year, it does not guarantee that it will have a dramatic influence on the public. At the same time, box office failures might be very influential to the audience. Depending on the year, the highest grossing box office hit may be much less well known than in previous or later years.

Setting of the movies could lead to more limitations. Just because a movie was produced in the past sixty years does not mean that the setting would be current. Movies like *Pirates of the Caribbean: Dead Man's Chest* and *Star Wars: Episode I: The Phantom Menace* have a different time period (and in the case of *Star Wars*, a different planet) than the premiere date. Depending

on genre, location, setting, etc., films may have different information seeking behaviors than present day United States.

Another limitation is with the descriptive nature of quantitative content analyses. The results would simply be stating the observed information seeking behavior in the movie. The next step would be to understand how the presence of said behavior could affect the audience.

Appendix

See second file for Appendix A: Coding Sheet

References

AMC all-time top box-office hits. (2012). Retrieved from

http://www.filmsite.org/boxoffice2.html

Bryman, A. (2008). *Social research methods* (3rd ed.). New York: Oxford University Press.

Case, D. O. (2008). *Looking for information: A survey of research on information seeking, needs and behavior*. (2nd ed.). Bingly, England: JAI Press.

Dill, E., & Janke, K. L. (2010). "New shit has come to light": Information seeking behavior in The Big Lebowski. Retrieved May 27, 2012, from

https://scholarworks.iupui.edu/bitstream/handle/1805/2099/new%20shit.pdf

Drumheller, K. (2005). Millennial dogma: A fantasy theme analysis of the Millennial generation's uses and gratifications of religious media. *Journal of Communication & Religion*, 28, 47-70. www.wtamu.edu/~kdrumheller/Drumheller%20Dogma.pdf

Escarmilla, G., Cradock, A.L., & Kawachi, I. (2000). Women and smoking in Hollywood movies: A content analysis. *American Journal of Public Health*, 90(3), 412-414.

Lester, P., & Smith, R. (1990). African-American photo coverage in "Life", "Newsweek", and "Time", 1937-1988. *Journalism Quarterly*, 67(1), 128-136.

McQuail, D. (1979). The influence and effects of mass media.*Mass Communication & Society*,

patron. (2006). In *Dictionary of Information and Library Management*. Retrieved from http://0-www.credoreference.com.www.whitelib.emporia.edu/entry/acbinfomanage/patron

resource. (2006). In *Dictionary of Information and Library Management*. Retrieved from

http://0-

www.credoreference.com.www.whitelib.emporia.edu/entry/acbinfomanage/resource

Reynolds, A., & Barnett, B. (2003). This just in . . . How national TV news handled the breaking

"live" coverage of September 11. *Journalism and Mass Communication Quarterly*, 80(3),

689-703.

Riffe, D., Goldson, H., Saxton, K., & Yu, Y. 1989. Females and minorities in children's Saturday

morning television commercials. (1987). *Journalism Quarterly,* 66, 129–136.

TMe box office tops. (n.d.). Retrieved from http://www.teako170.com/box60-69.html

Urbaniak, G. C., & Plous, S. (2008). *Research randomizer*. Retrieved from

http://www.randomizer.org/form.htm