Content Analysis: Applying the Research Method for Analysis of the Web Content

Piyali Ghosh

Research Scholar

Dept. of Library and Information Science

Jadavpur University

Kolkata 32

India

Abstract:

Purpose: The purpose of this paper is to introduce the content analysis technique as a research method. For this purpose, the paper seeks to exemplify the use of content analysis on Web-Based contents and describes each steps involved.

Methodology: To illustrate how content analysis may be used, this paper provides a brief overview of a case study that was designed to find out the most suitable encyclopedic reference tool regarding Indian philosophy among three free online encyclopedias, namely *Wikipedia* which was general in nature and other two were *Stanford Encyclopedia of Philosophy* and *Internet Encyclopedia of Philosophy*. In this study the standard approach was taken to conduct the content analysis among those three encyclopedias' websites. Based on that experience, this study makes several suggestions on the benefits of using content analysis.

Practical implications: This paper provides the techniques to apply content analysis of the material from the Websites.

Key words: Content analysis, web content analysis, data coding

1. Introduction

Originated in sociology and widely used in mass communication research, content analysis is now also popular in LIS research, e.g., analyzing messages conveyed in different types of information resources, professional discourses, citation functions in scholarly texts, etc (Lin & Jeng, 2015). Content analysis is a flexible and unobstructed method to analyze the meanings of narratives objectively. The rise of the World Wide Web attracted concerns among those social science and LIS scholars, especially those in the communication school who studied it by various methods like content analysis. However, the dynamic environment of the World Wide Web challenged this traditional research method, and, in turn, scholars tried to figure out valid solutions. A research work by Zhang (2005) conducted a thematic meta-analysis to examine how researchers apply content analysis to the World Wide Web after 2000. A total of 39 studies that used content analysis to study Web sites were identified from three sources. Then data were collected and analyzed. It was found that, from 2000 to 2004, content analysis of the World Wide Web proliferated. Content analytical scholars had created new strategies to cope with challenges posed by the WWW. The suggestions made in this study forms some guidelines in the steps of content analysis research design, potentially aiding the future research of content analysis to Web sites in developing their own valid methods to study the rapid-paced WWW.

2. What is content analysis?

Content analysis is a method applied for document or documentary research to determine certain specified characteristics of communication (Ohdedar, 1993). It is the application of scientific methods to documentary evidence and a method of analyzing written, verbal or visual communication message (Elo & Kyngäs, 2007).Content analysis as a research method is a systematic and objective means of describing and quantifying phenomena. Content analysis provides a structured way of analyzing data that are typically open-ended and relatively unstructured. Content analysis is a systematic, rigorous approach to analyzing documents obtained or generated in the course of research (White & Marsh, 2006).

The aim of content analysis study is to attain a condensed and broad description of the phenomenon, and the outcome of the analysis is concepts or categories describing the phenomenon. Usually the purpose of those concepts or categories is to build up a model, conceptual system, conceptual map or categories (Elo & Kyngäs, 2007).

3. The units involved in content analysis

The whole process of content analysis involves following units (White & Marsh, 2006):

3.1. Units of analysis: The units of analysis are the basis for interpreting data and reporting analysis and these are determined by research questions or hypothesis. So at the early stage of content analysis, the research questions or hypotheses have to be built. To give answers to research questions or to establish those hypotheses the analytical constructs has to develop, which will help to move from the text to the context. The two domains, the texts and the context are logically independent and the researcher draws conclusions by applying the rules of inference from one independent domain (the texts) to the other (the context). The analytical constructs may also be derived from (1) existing theories or practices; (2) the experience or knowledge of experts; and (3) previous researches. Units of analysis involves following steps:

    a. Establishing the coding scheme(that allows for testing hypothesis or answering research questions);
    b. Coding the data;
    c. Analyzing coded data, applying appropriate statistical test(s);

3.2. Sampling Units: These units serve to identify the population and establish the basis of sampling. In these units the sampling considerations must be determined. It is also to be deciding whether to analyze, only the manifest content or the latent content as well. It is important to assure that the sampling must be representative of the universe from which it is drawn. But probability or judgment sampling (purposive sampling) is necessary when a population is too large to be analyzed or the individual's judgment is given importance in providing answers to some research questions.

3.3. Data collection units: These units are the units for measuring the variables, selecting or choosing samples for analysis. Pragmatism determines the sampling and data collection unit.

3.4. Units for reporting: These units represent the results.

All these units are interdependent to conduct a content analysis study. The steps in each unit is differ from one another depending on their approaches whether it qualitative or quantitative in nature. So it is necessary to understand these two types of content analysis.

4. Types of content analysis

Content analysis enables the analysis of data to be structured and may be used in both qualitative and quantitative studies.

4.1. Features of qualitative content analysis

Qualitative content analysis typically based on an individual's perspective and is similar to textual analysis in that it is primarily interpretive in nature and often does not utilize statistics for data analysis. Qualitative content analysis flows from a naturalistic or a humanistic tradition. Here the data are read initially with the intent of trying to see the big picture (White & Marsh, 2006). It has following features:

4.1.1. Research approach: It has inductive research approach where **research questions** guide data gathering and analysis but potential themes and other questions may arise from careful reading of data.

4.1.2. Objective: It has the objective to capture the meanings, emphasis, and themes of messages and to understand the organization and process of how they are presented. It has the focus on the uniqueness of the text and are consciously aware of the multiple interpretations than can arise from a close perusal of it. The need for close, reiterative analysis itself usually limits the size of the sample.

4.1.3. Data Selection: Purposive sampling is used in this type of content analysis to allow for identifying complete, accurate answers to research questions and presenting the big picture and selection of data may continue throughout the project.

4.1.4. Categorization schema: Coding scheme usually developed in the process of close, iterative reading to identify significant concepts and patterns of the document or documents.

4.2. Features of quantitative content analysis

Quantitative content analysis flows from a positivist research tradition. According to White and Marsh (2006) quantitative content

analysis is "a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use". It has the following features:

4.2.1. Research approach: It has deductive research approach and based on previous research, which allows for formulating hypotheses about relationships among variables.

4.2.2. Objective: It is used to make valid and reliable inferences from the data to their context.

4.2.3. Data Selection: The process of data selection involved here is systematic and random sampling is used to allow for generalization to broader population; data selection usually complete prior to coding.

4.2.4. Categorization schema: Coding scheme developed here is a priori in accord with testing hypotheses. If adjustments are made during coding, items already coded must be recorded with the revised scheme.

5. Use of content analysis as a methodology

Content analysis has been applied to a wide range of social science topics including gender and race, violence, media reporting, political communication etc (Rose, Spinks, & Canhoto, 2015). According to Elo and Kyngäs (2007) it was first used as a method for analyzing hymns, newspaper and magazine articles, advertisements, and political speeches in the 19th century. Today, content analysis has a long history to use in communication, journalism, psychology and business and during the last few decades its use has shown steady growth. The rise of Internet creates opportunities for using content analysis techniques to analyze online communication. Pictures on a Website for example are used to convey one or more meanings, often in combination with text and can be subjected to content analysis either by themselves or by looking at the relationships between images and text (White & Marsh, 2006). It can also be used to the Website itself, as for example, it can be used to one or more organization's Websites, Government's official Websites to analyze how efficiently these sites convey massages to the public. In recent times, the content analysis is also used to online information resources to assess or evaluates their quality and the results of these analyses are used to take decisions about the uses of those sites.

6. Use of Content analysis on Web-based content

According to Kim and Kuljis (2010) the method of content analysis has been employed not only in the field of traditional communication, but also in studies of human-computer interaction such as web based applications, norms of behavior and cultural values. In an article, he has described the process of applying content analysis on Web documents. However, applying content analysis to Web-based content faces many challenges such as sampling and coding. The complexity of the mix of various media characteristics within the Web-content affects generalizations and representativeness. The Web is a complex and rich mixture of old and new technologies. Therefore, it provides many opportunities and challenges for researchers who apply content analysis to Web-based content. In particular, the complexity of new features such as mixed multiple media (text, graphics, animation, video and audio etc), interactivity, decentralized and hyperlinked structures, and its continuously evolving nature provide challenges to the development of valid descriptive categories, recording and sampling frames for the method.

7. Steps in content analysis

Kim and Kuljis (2010) had illustrated the process of content analysis through nine steps. Those are:
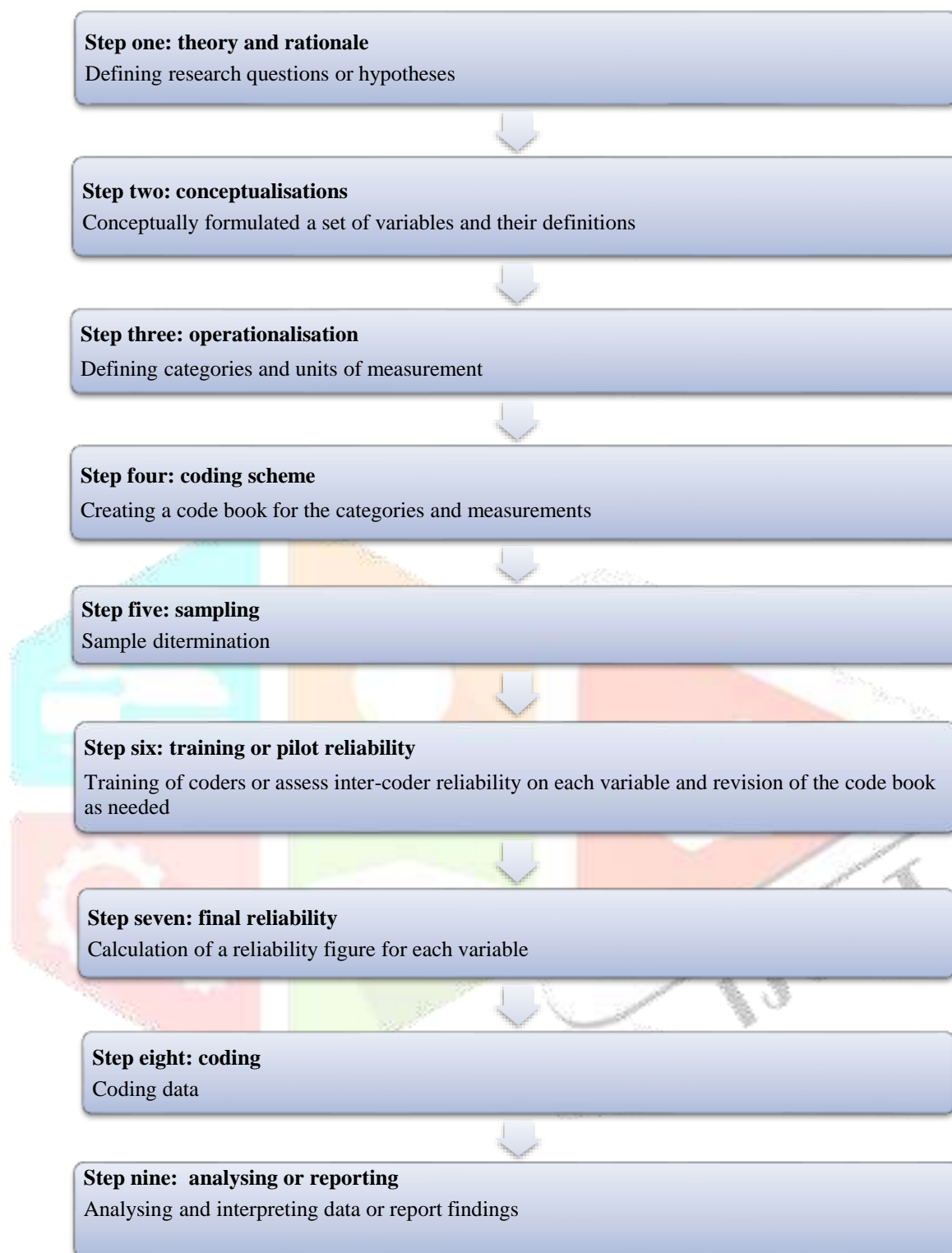
Step one: theory and rationale
Defining research questions or hypotheses

Step two: conceptualisations
Conceptually formulated a set of variables and their definitions

Step three: operationalisation
Defining categories and units of measurement

Step four: coding scheme
Creating a code book for the categories and measurements

Step five: sampling
Sample ditermination

Step six: training or pilot reliability
Training of coders or assess inter-coder reliability on each variable and revision of the code book as needed

Step seven: final reliability
Calculation of a reliability figure for each variable

Step eight: coding
Coding data

Step nine: analysing or reporting
Analysing and interpreting data or report findings

Figure1. Steps of content analysis

8. Exemplifying the process : a case study of three websites' content analysis

The qualitative content analyses was taken as a methodology in previous study to find out better reference source on Indian philosophy among two online subject encyclopedias and one general encyclopedia and above step were followed. This paper gives a clear view of those steps involved in the content analysis which was taken as methodology for that study.

Step one: formulation research question or hypothesis

The problem of the proposed research was the assessment of *Wikipedia* as reference tools regarding the treatment of 'Indian

Philosophy'. *Wikipedia* is a free online general encyclopedia. Two other free online popular encyclopedias on philosophy are *Stanford Encyclopedia of Philosophy* and *Internet Encyclopedia of Philosophy*. These two encyclopedias were selected. Is *Wikipedia*, the best reference source for Indian philosophy in comparison with other two free online encyclopedias *SEP* and *IEP*? The study was designed to find out the answer of above research question. Objective was to satisfy academics' concern about the quality of these reference works *SEP, IEP, Wikipedia*, to explore the completeness in covering the subject by these reference works; to explore the comprehensiveness of the articles; to determine the satisfactory level of Wikipedia, *IEP* and *SEP* in expressing any particular topic; finally to provide reference librarians with evidence regarding the quality of *Wikipedia* articles compared with respected reference resources on Indian philosophy. Depend on these objectives the above mentioned research question was broken into following six questions:

i. How well Indian philosophy is represented in *SEP*, *IEP*, *Wikipedia*?
ii. How far *SEP*, *IEP*, *Wikipedia* give authoritative information on Indian philosophy?
iii. How far *SEP*, *IEP* and *Wikipedia* provide the scope for further study or research on Indian philosophy?
iv. How far the articles on Indian philosophy from *SEP*, *IEP* and *Wikipedia* are comprehensive?
v. How well the articles on Indian philosophy from *SEP*, *IEP*, *Wikipedia* are understandable?
vi. How far *SEP*, *IEP* and *Wikipedia* provide up-to-date information on Indian philosophy?

Step two: identifying variables

For finding out the answers of the above research questions, it was necessary to identify some especial aspects of articles from each encyclopedia separately. Those are:

i. Representation of the articles;
ii. Authoritativeness of the articles;
iii. Scope for further study or research in the articles;
iv. Comprehensiveness of the articles;
v. Understandability of the articles;
vi. Uptodatedness of the articles.

Step three: defining the units of measurements

The proposed study had to find out answers of the above research questions. For that purpose the above variables were chosen. The next step was to identify the units of measurement. It may also call parameters on which the articles from different encyclopedias would vary to each other. Following categories or parameters were chosen for each variable:

Table.1. Table of parameters

Variables	Units of measurements or parameters
Representation	Number of images
Authoritativeness	Author's name Author's credential Author's contact address
Scope for further research	Total number of references Types of references
Comprehensiveness	Average number of words in an articles Coverage of the article
Understandability	Readability grade of the article
Uptodateness	The date of publication or last updation of the article

Step four: creating coding scheme

Data were coded according to the above parameters separately for each encyclopedia and the measurements were recorded in the code book.

Step five: sampling

For each encyclopedia, the articles related to Indian philosophy were the population of observation for this study. Indian philosophy is generally subdivided into nine philosophical schools, i.e., Buddhist, Jain, Cārvaka, Nyāya, Vaiśesika, Sāṁkhya, Yoga, Mīmāmsa, and Vedānta. So it was convenient to divide the population of Indian philosophy articles into the strata like articles on Buddhist philosophy, Jain philosophy, Cārvaka philosophy, Nyāya philosophy, Vaiśesika philosophy, Sāṁkhya philosophy, Yoga philosophy, Mīmāmsa philosophy, Vedānta philosophy.

After dividing the articles into above nine strata, the sample drawing was started. But it was found that there were many articles in each encyclopedia which was not fitted into above groups. To accommodate those articles another group was created which was given the name "Miscellaneous articles". Therefore all the articles were subdivided into three main groups; those are Articles from Nāstika Schools, Articles from Āstika Schools and Miscellaneous articles. Since the population under observation had different strata so in this study stratified sampling was chosen to draw articles from those encyclopedias. As because it was expected that the subject encyclopedia should cover all aspect of a particular subject as much as possible, so it was convenient for the study to choose the subject encyclopedias first to identify the Indian philosophy related articles. It was found that *IEP* has 24 articles and *SEP* has 19 articles on Indian philosophy. The next step was to collect similar articles from *Wikipedia*. There are 28 articles in Wikipedia.

Step six: data coding

Data coding, i.e. the coding of the sample articles was done by the coding schemes defined in the code book. The study considered the following methodology to code data according to the previously mentioned parameters:

Table 2. Data coding methodologies

Parameters	Coding methodologies
Number of images	Counting the total images used in an article.
Author's name Author's credential Author's contact address	Checking out whether the article have mentioned author's name, credential, contact address
Total number of references Types of references	Analyzing the type of references like books, periodical, dictionaries, encyclopedias etc. used in an article and counting down the number of those references in each type.
Average number of words in an articles Coverage of the article	Counting the total words of an article to find average words of an article from each encyclopedia. The coverage was coded by checking the articles in a way that whether they have covered the fundamental topics regarding the concepts or not.
Readability grade of the article	To code the readability of the articles (and also to count the total number of words in the articles) online readability count software tool was used.
The date of publication or last updation of the article	Considering the date of publication and last date of updation (if available) of an article to identify the most updated articles from each encyclopedia.

Step seven: data analysis and findings

This stage was very crucial, because the answers of the pre-mentioned research questions would have to be come out from the analysis. Qualitative content analysis method was applied to the three online encyclopedias' articles. After the analysis of contents of selected articles, it was found that among three encyclopedias, *Wikipedia* is the only source which best represents the topics of the selected subject and provides authoritative articles. It provides best scope for further research on the selected subject by providing lots of references along with the articles. Wikipedia's articles were found more comprehensive than other two encyclopedias' articles. Moreover it provides articles which are most understandable for the students and this encyclopedia updated regularly than other two encyclopedias. So it was easy to decide that among *SEP*, *IEP*, *Wikipedia* which one should be considered as a best reference source on Indian philosophy.

9. Conclusion

The steps mentioned above were found easy and useful to conduct content analysis as a research methodology. Moreover it is flexible and can be modified or restructured as per researcher's necessity. Therefore it may be concluded that applying content analysis to Web-based content is a relatively easy process that allows researchers to perform and prepare data at their convenience. So it is expected that more researchers will start investigating Web-based contents and content analysis will become a method of choice in such studies.

REFERENCES

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. 2014. Qualitative content analysis: a focus ontrustworthiness. SAGE Open Feb 2014, 4 (1).

Ghosh, P. 2016. The treatment of Indian philosophy in Wikipedia, Stanford Encyclopedia of Philosophy and Internet Encyclopedia of Philosophy: a comparative study. (Unpublished master's thesis). Jadavpur University, Kolkata, West Bengal, India.

Kim, I., & Kuljis, J. 2010. Applying content analysis to web-based content. Journal of computing and information technology, Vol. 18 (4), 369-375.

Lin, C., & Jeng, W. 2015. Using content analysis in LIS research: experiences with coding schemes construction and reliability measures. Qualitative and Quantitative Methods in Libraries (QQML), Vol. 4, 87—95.

Ohdedar, Dr. A. R. 1993. Research methodology. Calcutta: Bengalore Library Association.

Rose, s., Spinks, N., and Canhoto, A. 2015. Quantitative content analysis.

White, M. and Marsh, E. 2006. Content analysis: a flexible methodology. Library Trends, Vol. 55 (1), 22-45.

Zhang, J. 2005. Content analysis of web sites from 2000 to 2004: a thematic meta-analysis. Submitted to the Office of Graduate Studies of Texas A&M University.