

---

## **UNIT 2    CONTENT ANALYSIS : QUANTITATIVE AND QUALITATIVE ASPECTS**

---

### **Structure**

- 2.0 Objectives
- 2.1 Introduction
- 2.2 History of Content Analysis
- 2.3 Use of Content Analysis
- 2.4 Types of Content Analysis
  - 2.4.1 Quantitative or Conceptual Analysis
  - 2.4.2 Qualitative or Relational Analysis
  - 2.4.3 Advantages of Content Analysis
  - 2.4.4 Disadvantages of Content Analysis
- 2.5 Evaluating Web Documents
- 2.6 Summary
- 2.7 Answers to Self Check Exercises
- 2.8 Keywords
- 2.9 References and Further Reading

---

### **2.0    OBJECTIVES**

---

After studying this unit you will be able to:

- understand what is content?
- analyze a document using quantitative as well as qualitative techniques;  
and
- evaluate a web document using various techniques.

---

### **2.1    INTRODUCTION**

---

Content analysis is a tool to identify the occurrence of concepts in an exposition. It is nothing but quantification of text or its meaning. An exposition can be an article, book, lecture, interview, conversation, Web page or any other reading material. To conduct a content analysis of any text, the text may be coded or broken down, into manageable meaningful categories on a variety of levels—word, word sense, phrase, sentence, or theme—and then examined using any one of the two basic methods of content analysis viz. conceptual analysis or relational analysis.

---

### **2.2    HISTORY OF CONTENT ANALYSIS**

---

Content analysis is an important job for developing a product in the library parlance, particularly, when we think of bringing information consolidation packages for the user community. It not only consists of analysis of text matter but also the idea that the text carries. Initially it used to be done manually but

due to the development and deployment of technology this process could be automated to a great extent. During the 1950s, information scientists started working on analyzing the concepts of the contents. Content analysis explores mental models, and their linguistic, affective, cognitive, social, cultural and historical significance.

---

## 2.3 USE OF CONTENT ANALYSIS

---

The scope of content analysis ranges from marketing and media studies, to literature and rhetoric, ethnography and cultural studies, gender and age issues, sociology and political science, psychology and cognitive science, and many other fields of inquiry. Berelson has suggested the possible use of content analysis as it:

- reveals international differences in communication of content;
- detects the existence of propaganda;
- identifies the intentions, focus or communication trends of an individual, group or institution;
- describes attitudinal and behavioral responses to communications; and
- determines psychological or emotional state of persons or groups.

---

## 2.4 TYPES OF CONTENT ANALYSIS

---

We will discuss two general categories of content analysis:

- 1) Conceptual or quantitative analysis and
- 2) Relational or qualitative analysis.

Conceptual analysis can be thought of as *frequency of concepts*. Concept can be represented by texts as well as pictures where text is very common. To analyze the concept we look for the appearance of words in the text. It is not necessary that the same word appears always, there may be synonymous terms present. For example, if we are analyzing with a hypothesis that a certain document is about *freedom* then one should look for the related words like *liberation, independence*, etc. In contrast, relational analysis goes one step further by examining the relationships among concepts in a text. In relational analysis we look for what are the related words appearing next to the word in question. For example we see what are the words that appear next to *freedom* and then determine the related concepts.

### 2.4.1 Quantitative or Conceptual Analysis

Traditionally, content analysis has most often been thought of in terms of quantitative analysis. In quantitative or conceptual analysis, a concept is chosen for examination, and the analysis involves quantifying and tallying its presence, which is also known as thematic analysis. The focus here is on looking at the occurrence of selected terms within a text, the term could be found as synonyms (explicit) or as a concept using a cluster of different words (implicit). The identification of the second category is a little difficult.

## Methods of Quantitative or Conceptual Analysis

Quantitative or Conceptual analysis begins with identifying research questions and choosing a sample or samples. In the sample, text is sought and put in the different categories or the content is codified in different categories. This method is known as *selective reduction*. By reducing the text to categories consisting of a word, set of words or phrases, the researcher can focus on, and code for, specific words or patterns that are indicative of the research question.

An example of a quantitative or conceptual analysis would be to examine several library websites, and code them for the existence of certain words. For example take '*agricultural economics*'. Examining the related web pages, the answer to the information needed might involve examining the number of key words of '*agricultural economics*'. The researcher would be interested only in quantifying these words, not in examining how they are related with other concepts. For example, terms like '*pest control*' or '*inflation rate of country and its impact on agricultural economics*' which is a part of relational analysis. In conceptual analysis, the researcher simply wants to examine the presence with respect to his/her research question, i.e. is there a stronger presence of positive or negative words used with respect to proposed question?

Once the research question has been established, the researcher can make his/her coding choices with respect to the eight category coding steps indicated by Carley (1992).

The eight category coding steps are as follows:

- a) Decide the level of analysis
  - b) Decide how many concepts to code for
  - c) Decide whether to code for existence or frequency of a concept
  - d) Decide on how you will distinguish among concepts
  - e) Develop rules for coding your texts
  - f) Decide what to do with "irrelevant" information
  - g) Code the texts
  - h) Analyze your results
- a) **Decide the level of analysis**

First, one should decide upon the level of analysis. Level of analysis means the word or word set which is chosen for coding. To continue with the example, the researcher must decide whether to code for a single word for example, "*farming*" or sets of words or phrases "*farming of rice*".

- b) **Decide how many concepts to code for**

Once the concept is decided he/she should decide how many other concepts should be taken into consideration. That means he/she should pre-define the categories. He/she must also set the level of flexibility allowed while picking the words for these categories. The researcher should also decide on the way of categorization whether he/she should be allowed to pick any other category

if he/she feels it is relevant. It also must be decided if each and every word which appears should be coded or not or which words should be taken into consideration.

**c) Decide whether to code for existence or frequency of a concept**

Another major decision the researcher needs to take is whether he/she should count the word i.e. frequency or the cluster of words which denotes the concept of word in question. If he counts the frequency he/she may lead to a limited perspective of concept. If he/she takes care of the concept as well as related terms he/she may get varied perspective of concept. Say for example, the term "*farming*" gives a very broad perspective but if we consider the words "*cultivation of rice*" we understand that the concept is about rice production.

**d) Decide on how you will distinguish among concepts**

Now the researcher should decide whether the concepts are to be coded exactly as they appear, or if they can be recorded as the same even when they appear in different forms for example, "*farming*" can appear as "*harvesting*". The researcher needs to decide whether both words mean the same or they are different so that both could be kept in different categories.

**e) Develop rules for coding your texts**

This process belongs to idea plane. After taking the generalization of concepts into consideration, a researcher will want to create translation rules or in other words use of numbers that will allow him/her to streamline and organize the coded terms to understand that he/she is coding for exactly what he/she wants. Developing a set of rules helps the researcher to ensure that he/she is coding things consistently throughout the text, in the same way every time. Say for example, "*The use of fertilizers*" and "*farming*" are two different categories then one cannot put the previous one under the second, as it will violate the rules. With construction of these rules one can understand the relationship of related concepts and the importance of their meaning and use. This provides consistency and coherence. One of the major approaches to it is the use of classification schemes as they provide consistency in codification of terms.

**f) Decide what to do with "irrelevant" information**

Now the researcher should decide what should be done with irrelevant information. He/she should decide if findings have some kind of irrelevancy then whether to bear it or re-do the process. Usually the puff words should be left out for example, *and, of, the*, etc. as they add nothing to the quantification and thus can be left.

**g) Code the texts**

This is the key process for notational plane where we codify the term in an artificial language, may be in numbers. Once these choices about irrelevant information are made, the next step is to code the text. This is done either by hand, i.e. reading through the text and manually writing down concept occurrences, or through the use of various computer programs. When coding is done, a researcher can recognize errors far more easily. A computer is only a

tool and can only code based on the information given. This problem is most apparent when coding for implicit information, where category preparation is essential for accurate coding.

**h) Analyze your results**

After coding is done, the researcher should try to conclude from the data, but before that he/she should decide what should be done with the remaining uncoded data. The options are whether to delete that part of text or re-codify all the text changing the coding scheme. Once this part is over he/she should look for a conclusion, by looking at the results or trends. For example, if one finds that the frequency of 'harvesting of rice' is more, could he conclude that there is a shift in farming and more people are going for farming of rice?

**Self Check Exercise**

- 1) Define the steps involved in Conceptual analysis.

.....  
.....  
.....  
.....

**2.4.2 Qualitative or Relational Analysis**

Qualitative or Relational analysis is an extension to the Quantitative analysis. However, qualitative analysis seeks to go beyond presence of concepts by exploring the relationships between the concepts identified. Relational analysis has also been termed as *semantic analysis*. In other words, the focus of relational analysis is to look for semantic, or meaningful, relationships. Individual concepts are viewed as having no inherent meaning. Rather, meaning is a product of the relationships among concepts in a text.

Basically there are two key theoretical approaches for qualitative content analysis linguistics and cognitive science.

**Linguistics**

Linguistic approach to content analysis focuses on the analysis of texts on the level of a linguistic unit, typically single clause units. One example of this type of research is Gottschalk (1995), who developed an automated procedure, which analyzes each clause in a text and assigns it a numerical score based on several emotional/psychological scales.

**Cognitive Science**

Approaches that derive from cognitive science include the creation of decision maps and mental models. Decision maps attempt to represent the relationship(s) between ideas, beliefs, attitudes, and information available to an author when making a decision within a text. These relationships can be represented as logical, inferential, causal, sequential, and mathematical relationships. Typically, two of these links are compared in a single study, and are analyzed as networks. This methodology is thought of as a more generalized cognitive mapping technique, rather than the more specific mental models approach.

Mental models are groups or networks of interrelated concepts that we make in our mind at conscious or subconscious level. Cognitive scientists say, these internal mental structures are made as people draw inferences and gather information about the things around them. Mental models are more specific approaches of mapping because conceptual entities can be numerically and graphically analyzed. Semantic nets fall under this category. Studies based on this approach follow five general steps:

- i) Identifying concepts
- ii) Defining relationship types
- iii) Coding the text on the basis of i) and ii)
- iv) Coding the statements
- v) Graphically displaying and numerically analyzing the resulting maps.

**Relational Analysis: An Overview**

The basic objective of Qualitative analysis is to identify and understand the concept in question. Then to look for the sample and its size. The larger the sample, the less the clarity of result but if the sample is small, the reliability of the result may be less. Thus with a larger sample contextual information must be given for the clarity in result. Computers are very helpful in Qualitative analysis because of use of rigorous statistical techniques.

**Self Check Exercise**

- 2) What is Relational analysis? How does cognitive science and linguistics affect it?

.....  
.....  
.....  
.....

**Steps for Conducting Qualitative or Relational Analysis**

The following are the steps for Qualitative analysis:

- a) Identify the Question
- b) Choose a sample or samples for analysis
- c) Determine the type of analysis
- d) Reduce the text to categories and code for words or patterns
- e) Explore the relationships between concepts (Strength, Sign & Direction)
- f) Code the relationships
- g) Perform Statistical Analyses
- h) Map out the Representations

**a) Identify the Question**

The first step is to identify the question or in other words the identification of to problem the because this guides the researcher throughout the process even when analysis becomes difficult because different interpretations are possible

for the same observation. From our earlier example of library web pages we can for?

What information is sought out of a library site?

**b) Choose a sample or samples for analysis**

Once the question has been identified, the researcher must select content for sample. For relational content analysis, the primary consideration is how much information should be preserved for analysis. One must be careful not to limit the results. The researcher must take special care not to take on so much that the coding process becomes too heavy and extensive to supply worthwhile results.

**c) Determine the type of analysis**

Once the sample has been chosen for analysis, it is necessary to determine what type or types of relationships you would like to examine. There are different subcategories of relational analysis that can be used to examine the relationships in texts. The three subcategories of Relational Analysis are as follows:

**Affect extraction**

This is a kind of emotional evaluation of writer or speaker. It is basically the psychological state of the writer or speaker. It is very difficult to ascertain the emotional/psychological level of a person. Gottschalk (4) says that this can be ascertained by verbal behavior of speaker or writer. He further gives example, by assigning concepts identified, a numeric value on corresponding emotional psychological scales that can then be statistically examined.

**Proximity analysis**

Here we identify a string, which represents a concept. This string is known as a window. This window is scanned through the document for co-occurrence of concepts. Finally, a matrix is created of all interrelated concepts, which is known as *concept matrix* to represent overall concept. This technique is problematic because the window records only explicit concepts and treats meaning as proximal co-occurrence. Other techniques such as clustering, grouping, and scaling are also useful in proximity analysis.

**Cognitive mapping**

This approach is one that allows for further analysis of the results from the two previous approaches. It attempts to take the above processes one step further by representing these relationships visually for comparison. While effective and proximal analysis function primarily within the preserved order of the text, cognitive mapping attempts to create a model of the overall meaning of the text. This can be represented as a graphic map that represents the relationships between concepts.

It lends comparison of semantic connections across texts. This makes us understand the possible meanings of the text. This can derive a variety of mental models i.e. from writers' or speakers' point of view; or any other if possible, depending on the research.

Once the subcategory of analysis is chosen, the selected text must be reviewed to determine the level of analysis. The researcher must decide whether to code for a single word, or for sets of words or phrases.

d) **Reduce the text to categories and code for words or patterns**

The next step is to determine the categories and codify the words as we do in Quantitative analysis. The standardization of terms is brought about by removing ambiguity found among the terms such as problems of homonym or synonymous terms. To remove ambiguity of terms a thesaurus can be used. One can create his/her own authority file, or use any existing thesaurus.

e) **Explore the relationships between concepts (Strength, Sign and Direction)**

Once words are categorized and coded, the relationship among the concepts is seen. This relationship is based on three factors- strength, sign and direction of relation. The strength of relation is chiefly through the occurrence of words like *must, should, perhaps, may be, unless*, etc. The sign basically shows whether it is negative or positive. In a stock exchange, bear and bull are typical examples of it; where the bear shows negative while the bull shows positive. The flow or direction of relationships is another important aspect of the content analysis. This can be typically seen in programming language where decisions regarding various conditions are shown as *if-then-else*.

f) **Code the relationships**

One of the main differences between quantitative and qualitative is that the statements or relationships between concepts are coded. In other words the relation among the terms is identified and denoted with symbols.

g) **Perform Statistical Analyses**

This step involves conducting statistical analyses of the data coded during qualitative or relational analysis. This may involve exploring for differences or looking for relationships among the variables one has identified in his study.

h) **Map out the Representations**

In addition to statistical analysis, qualitative or relational analysis often leads to representation of text in mental model i.e. mapping of text. Qualitative analysis is also performed by a variety of different theoretical approaches: linguistic content analysis, decision mapping, and mental models.

**Content Analysis: Reliability and Validity**

The Reliability of content analysis depends on stability, reproducibility, and accuracy. Stability means, a coder, if he/she re-codes the same text he/she should get the same results. Reproducibility refers to the tendency for a group of coders to classify categories membership in the same way and accuracy is the extent to which the categorization conforms to standards. According to Gottschalk (1995) (4), getting the same level of reliability is not possible. So he suggests 80% as an acceptable margin for reliability.

The validity corresponds to *categorization, conclusions* and *generalizability*. Valid *categorization* should generate the same set of category whenever and whoever categorizes the text.

The *conclusion* should be looked into with the degree of challengeability that means how the conclusion can be explained. Is it based on data or can it be explained on some phenomenon or other factors? One such example can be shown with the use of homonyms. The term "mine" is a personal pronoun but the same word is also a noun, the explosive device. And if we are working with

*mine* as personal pronoun we will get the result of “mine” as explosive. Such result may affect the result severely.

The *generalization* from conclusion depends on the categories, the conclusion is brought and their reliability.

### 2.4.3 Advantages of Content Analysis

Content analysis offers several advantages to researchers who consider using it. The following advantages could be understood (2):

- It provides a social communication among scholars via texts or transcripts.
- A text can be analyzed both qualitatively and quantitatively. Thus understanding the relationship of concepts.
- Insight of old civilizations can be understood by analyzing the text of inscriptions.
- It allows closeness to text, which can alternate between specific categories and relationships and also statistically analyzes the coded form of the text.
- It can be used to interpret texts for the development of expert systems to develop the rule base or inference engine showing the relationship among the concepts.
- It can be used to understand the way human mind thinks.

### 2.4.4 Disadvantages of Content Analysis

There are some disadvantages associated with content analysis:

- It takes time to be done.
- Relational analysis is a matter of high interpretation, hence is prone to inaccuracy.
- It is based on few steps; there is no sound theory to date, behind doing content analysis, which may reduce the precision of study sometimes.
- In content analysis a large concept is reduced to a code. Often there can be mistakes while coding, particularly dealing with complex texts and can result in ‘information loss’.
- Often content analysis is criticized to be as word count.
- Often we need to understand the context while drawing a conclusion from an analyzed text because there could be many possible interpretations.

#### Self Check exercise

3) Define the steps involved in conducting Relational analysis?

.....

.....

.....

.....

---

## 2.5 EVALUATING WEB DOCUMENTS

---

Search engines generate results in millions in which only a few are relevant. In such an environment it is highly recommended that pages must be evaluated.

Louis Shores has given a checklist for evaluation of reference books and the same can be applied for Web documents. Librarians can use these checkpoints for evaluation of Web documents, which they keep in digital library or subject gateway collection. Commonly these points are known as ASTAFS.

- 1) **Authority:** The very first step is to identify the author and publisher of the web document. On Internet it is very difficult to find the author and the publisher of a website. But one can easily find out the organization, which has hosted the site and is responsible for the content generation. For example, the most authentic source for syllabus of National Eligibility Test (NET) conducted by University Grants Commission (UGC) is the UGC website (<http://ugc.ac.in>).
- 2) **Scope:** What is the subject covered by the website? Which authors have contributed to the website in case there are many contributors? What is the recency level of the website? Whether the links to other sites are given; if so, whether they are working or not? For example, the Berkeley Sunsite Digital Library, at California has an exclusive collection of images in different subject areas.
- 3) **Treatment:** How accurately is the subject covered? Whether additional links are provided and if not, whether additional bibliography is provided?
- 4) **Arrangement:** The common feature of web documents is non-linearity. But still one can find out a level of arrangement of information on the website. Besides arrangement the most common feature is the style of rendering information. The style must be eye catching. Keeping newly added services in side of screen must be avoided. The best way to render the information is by giving a menu on user interface. A user must be given a search window for collection search. Besides to display the collection browsing facility must be given.

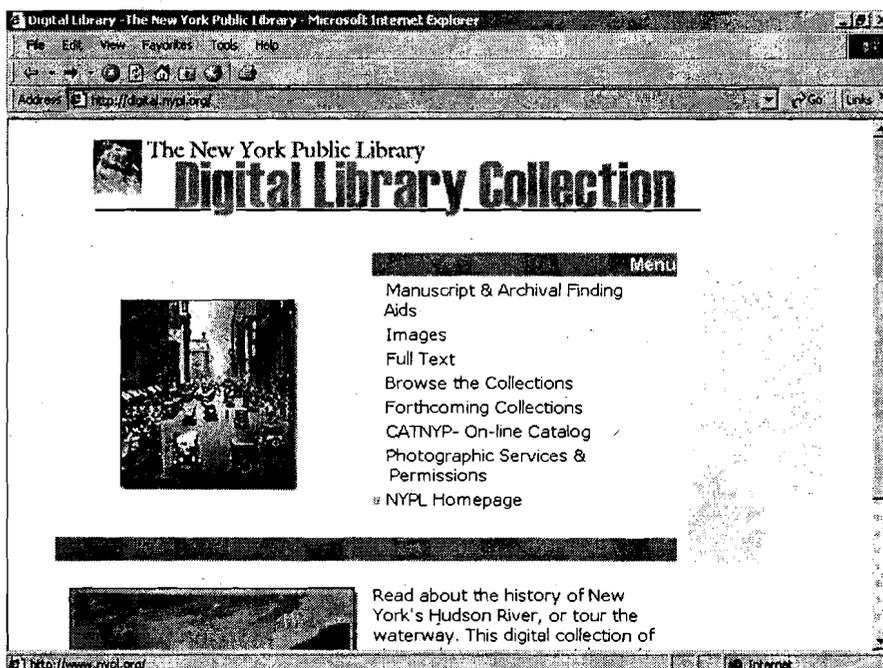


Fig. 1: Web interface of New York Public Library

Subject gateways are other information resources available on Internet and these can be very well managed by librarians because of their professional skills and specialized knowledge. A subject gateway is a subject portal and thus could be a launch pad for browsing in a specific subject discipline. It covers the different subtopics of a subject, the list of learned bodies and available online journals. Besides subject gateways provide searching facility. Information is well classified in such gateways and arranged in helpful sequence of directories.

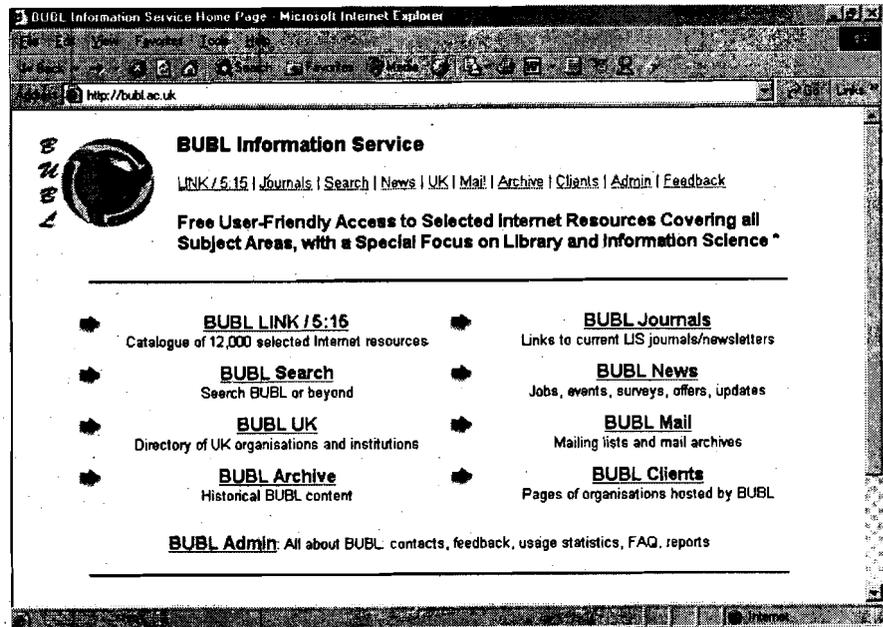


Fig. 2: BUBL Information Service a subject launch pad for librarians

- 5) **Format:** What is the format of a document? Here format means basically audio, video, text and image format. One should also look at the kind of file format it is i.e. whether it is MP3, WAV, MID etc. in case of sound files?
- 6) **Special Feature:** If there is any specific feature associated with the document? Specially, has indexing been done? Download given etc.

There are many projects which have recommended parameters for evaluation of Net resources. (5) (6) A few more parameters in addition to the classical ones given above are suggested for the evaluation of Internet resources. They are:

- Accessibility
- Currency
- Response Time
- Stability
- Accuracy
- Target Audience
- Coverage

- Language
- Completeness
- Style

**Self Check Exercise**

4) How does checklist of points of Louis Shores hold good for evaluating a web document?

.....  
.....  
.....  
.....

---

**2.6 SUMMARY**

---

Content analysis is the most important part for identifying, representing and understanding the content of any kind of document. In this unit we have discussed that Content analysis is basically of two types:

- Quantitative or conceptual analysis
- Qualitative or relational analysis

Quantitative content analysis is simply counting the occurrence of terms in a document and bringing out a certain inference out of it.

Qualitative content analysis is the application of text-based information analyzing techniques on the document. In addition, here we identify the relation among the terms or concepts in question. In other words, here we are more concerned with semantics of the document and the terms appearing in it.

Besides the above mentioned, the check-list of Louis Shores can also be adopted for evaluation of a document to identify its quality. Additionally other check points like file sizes and access time has to be kept in mind as they are pertinent to online net resources.

---

**2.7 ANSWERS TO SELF CHECK EXERCISES**

---

- 1) There are eight steps in conceptual analysis:
  - i) Decide the level of analysis
  - j) Decide how many concepts to code for
  - k) Decide whether to code for existence or frequency of a concept
  - l) Decide on how you will distinguish among concepts
  - m) Develop rules for coding your texts
  - n) Decide what to do with "irrelevant" information
  - o) Code the texts
  - p) Analyze your results

- 2) Relational Analysis has two parts. The first step involves quantitative technique for analysis and the second step is identification of the relation among the concepts, which are quantitatively analyzed.

Basically, linguistics is involved in analyzing the verbal content of the document. That means one can pick up various clauses appearing in a document and perform quantitative analysis, while cognitive science basically deals with the mental model. Here once we identify the term or the concept, which has to be analyzed, we try to develop the relation among the concepts and present. This relation can be the link in a sequence to perform a particular task or a decision tree. The mental model involves the following steps:

- i) Identifying concepts
  - ii) Defining relationship types
  - iii) Coding the text on the basis of step 1 and step 2
  - iv) Coding the statements
  - v) Graphically displaying and numerically analyzing the resulting maps.
- 3) The following are the steps for Qualitative analysis:
    - a) Identify the question
    - b) Choose a sample or samples for analysis
    - c) Determine the type of analysis
    - d) Reduce the text to categories and code for words or patterns
    - e) Explore the relationships between concepts (Strength, Sign & Direction)
    - f) Code the relationships
    - g) Perform Statistical Analyses
    - h) Map out the Representations.
  - 4) Louis Shores has defined the checklist for evaluation of reference sources. Due to advent of Internet most of the reference sources are now available over Internet. His checklist can be easily modified for evaluation of web resources:
    - i) **Authority:** Who is the publisher of the website? Or from which institute or server been the website been launched. Who are the contributors of the website?
    - ii) **Scope:** What is the subject area the web resource covers? What is the frequency of updating? When was the page last modified?
    - iii) **Treatment:** Whether the page is biased to a particular kind of product or thought content? Who are the target users? Is the matter information provided recent?
    - iv) **Arrangement:** How is the information arranged on the website? Is there any kind of classification followed to arrange the topics or it is alphabetically arranged.

- v) **Format:** What kind of content is available through the web resource. Whether it is text, audio, video, movie file or has it a downloadable repository for software.
- vi) **Special Features:** Is there any new feature associated with the web resource?

Besides these there are a few more points, which should be taken care of:

- Accessibility
- Currency
- Response Time
- Stability
- Accuracy
- Target Audience
- Coverage
- Language
- Completeness
- Style

---

## 2.8 KEYWORDS

---

- Case Study** : The collection and presentation of detailed information about a particular participant or small group, frequently including the accounts of subjects themselves.
- Content Validity** : The extent to which a measurement reflects the specific intended domain of content.
- Covariate** : A product of the correlation of two related variables times their standard deviations. Used in true experiments to measure the difference of treatment between them.
- Data** : Recorded observations, usually in numeric or textual form.
- Deductive** : A form of reasoning in which conclusions are formulated about particulars from general or universal premises.
- Electronic Text** : A “paper” or linear text that has been essentially “copied” into an electronic medium.
- Existence or Frequency** : The number of count.
- Inductive** : A form of reasoning in which a generalized conclusion is formulated from particular instances.

- Inductive Analysis** : A form of analysis based on inductive reasoning; a researcher using inductive analysis starts with answers, but forms questions throughout the research process.
- Interval Variable** : A variable in which both order of data points and distance between data points can be determined, e.g., percentage scores and distances.
- Level of Analysis** : Chosen by determining which word, set of words, or phrases will constitute a concept.
- Mental Models** : A group or network of interrelated concepts that reflect conscious or subconscious perceptions of reality.
- Parameter** : A coefficient or value for the population that corresponds to a particular statistic from a sample and is often inferred from the sample.
- Phenomenology** : A qualitative research approach concerned with understanding certain group behaviors from that group's point of view.
- Population** : A set of entities.
- Precision** : In survey research, the tightness of the confidence limits.
- Probability** : The chance that a phenomenon has of occurring randomly.
- Qualitative Research** : Empirical research in which the researcher explores relationships using textual, rather than quantitative data.
- Quantitative Research** : Empirical research in which the researcher explores relationships using numeric data. Survey is generally considered a form of quantitative research.
- Random Sampling** : Process used in research to draw a sample of a population strictly by chance, yielding no discernible pattern beyond chance. First numbering the population, then selecting the sample according to a table of random numbers or using a random-number computer generator can accomplish random sampling.
- Randomization** : Used to allocate subjects to experimental and control groups. The subjects are initially considered not unequal because they were randomly selected.
- Range** : The difference between the highest and lowest scores in a distribution.

- Reliability** : The extent to which a measure, procedure or instrument yields the same result on repeated trials.
- Sample** : The population researched in a particular study.
- Selective Reduction** : Text is reduced to categories consisting of a word, set of words or phrases, on which the researcher can focus. Specific words or patterns are indicative of the research question and determine levels of analysis and generalization.
- Validity** : The degree to which a study accurately reflects or assesses the specific concept that the researcher is attempting to measure.
- Variable** : Observable characteristics that vary among individuals.
- Variation** : The dispersion of data points around the mean of a distribution.

---

## **2.9 REFERENCES AND FURTHER READING**

---

Berelson, B. (1971). *Content Analysis in Communication Research*. New York: Hafner Publishing Company.

Overview: content analysis.

<http://writing.colostate.edu/references/research/content/index.cfm>

Carley, K., & Palmquist, M. (1992). Extracting, representing and analyzing mental models. *Social Forces*, 70 (3), pp.601-636.

Gottschalk, L. A. (1995). *Content Analysis of Verbal Behavior: New Findings and clinical applications*. Hillside, NJ: Lawrence Erlbaum Associates, Inc.

BIOME Special Advisory Group on Evaluation. How to evaluate an Internet-based information source. <http://biome.ac.uk/guidelines/eval/howto.html>

Ormondroyd, (Joan). *Critically Analyzing Information Sources*. Ed. by Michael Engle, and Tony Cosgrave. <http://www.library.cornell.edu/okuref/research/skill26.htm>

Abernethy, A. M., & Franke, G. R. (1996). The information content of advertising: a meta-analysis. *Journal of Advertising*, Summer 25 (2), 1-18.

Allen, Bryce and Reser, David (1990). Content analysis in library and information science research. *Library and Information Science Research*, 12 (3) July-Sept 90, 251-262.

Brigitte Endres-Niggemeyer, B. Endres-Niggemeyer (1989). Content analysis — a special case of text compression. Information, knowledge, evolution. (Proceedings of the 4Fourth FID Congress, Helsinki, 28 August-1 September 1988) Edited by Sinikka Koshiala and Ritva Launo, Amsterdam, Elsevier Science Publishers, 1989, 103-112.

- Budd, R. W., Thorp, R.K. and Donohew, L. (1967). *Content Analysis of Communications*. New York: Macmillan Company.
- Burden, C. (1994). Content analysis of computer search request forms in a teaching hospital medical library from 1979 to 1986. *Library and Information Science Research*, 16 (2) Spring 94, p.182-4.
- Carley, K. (1989). *Computer Analysis of Qualitative Data*. Pittsburgh, PA: Carnegie Mellon University.
- Carley, K. (1990). Content Analysis. In R.E. Asher (Ed.), *The Encyclopedia of Language and Linguistics*. Edinburgh: Pergamon Press.
- Carley, K. (1992). *MECA*. Pittsburgh, PA: Carnegie Mellon University.
- Carney, T. F. (1972). *Content Analysis: A Technique for Systematic Inference from Communications*. Winnipeg, Canada: University of Manitoba Press.
- de Sola Pool, I. (1959). *Trends in Content Analysis*. Urbana, Ill: University of Illinois Press.
- Duncan, D. F. (1989). Content analysis in health education research: An introduction to purposes and methods. *Health Education*, 20 (7).
- Ellis, B. G. & Dick, S. J. (1996). 'Who was 'Shadow'? The computer knows: applying grammar-program statistics in content analyses to solve mysteries about authorship. *Journalism & Mass Communication Quarterly*, 73(4), 947-963.
- Fan, D. (1988). *Predictions of Public Opinion from the Mass Media: Computer Content Analysis and Mathematical Modeling*. New York, NY: Greenwood Press.
- Franzosi, R. (1990). Computer-assisted coding of textual data: An application to semantic grammars. *Sociological Methods and Research*, 19(2), pp. 225-257.
- Gottschalk, L. A., Stein, M. K. and Shapiro, D.H. (1997). The application of computerized content analysis in a psychiatric outpatient clinic. *Journal of Clinical Psychology*, 53(5), pp. 427-442.
- Isaacs, J. S. (1984). *A Verbal Content Analysis of the Early Memories of Psychiatric Patients*. Berkeley: California School of Professional Psychology.
- Jolliffe, L. (1993). Yes! More content analysis! *Newspaper Research Journal*, 14(3-4), pp. 93-97.
- Jarvelin, K. and Vakkari, P. (1990). Content analysis of research articles in library and information science. *Library and Information Science Research*, 12 (4) Oct-Dec 90, 395-421.
- Krippendorff, K. (1980). *Content analysis: An Introduction to its Methodology* Beverly Hills, CA: Sage Publications.
- M. Perry, C. Bodkin (2000). Content analysis of Fortune 100 company Web sites. *Corporate Communications: An International Journal*, 5 (2) 2000, p.87-96.
- M. Soloman.(1993). Content analysis: a potent tool in the searcher's arsenal. *Database*, 16 (2) Apr 93, p.62-8.

Moshoeshoe, Matseliso Mamahlape. (1989). Content analysis and description of documents: evolution challenge to Lesotho. Information, knowledge, evolution. (Proceedings of the 4Fourth FID Congress, Helsinki, 28 August-1 September 1988) Edited by Sinikka Koshiala and Ritva Launo, Amsterdam, Elsevier Science Publishers, 1989, pp. 133-143.

McFavish, D.G. and Pirro, E. (1990) Contextual Content Analysis. *Quality and Quantity*, 24, 245-265.

Moeller, L. G. (1963). *An Introduction to content analysis—Including Annotated Bibliography*. Iowa City: University of Iowa Press.

O. M. Mommoh. (1997/1998). Content analysis of Library Focus: 1983-96. *Library Focus*, 15 and 16 1997/1998, pp.12-21.

O'Dell, J. W. & Weideman, D. (1993). Computer content analysis of the Schreber case. *Journal of Clinical Psychology*, 49(1), pp. 120-125.

Palmquist, M. E. (1990). *The Lexicon of the Classroom: language and Learning in Writing Class Rooms*. Doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.

Palmquist, M. E., Carley, K.M., and Dale, T.A. (1997). Two applications of automated text analysis: Analyzing literary and non-literary texts. In C. Roberts (Ed.), *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Tanscripts*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Pamela Snelson, S. Anita Talar. Content analysis of ACRL conference papers. *College and Research Libraries*, 52 (5) Sept 91, 466-472.

Riffe, D., Aust, C. F., & Lacy, S. R. (1993). The effectiveness of random, consecutive day and constructed week sampling in newspaper content analysis. *Journalism Quarterly*, 70 (1), 133-139.

Riffe, D., Lacy, S., & Drager, M. W. (1996). Sample size in content analysis of weekly news magazines. *Journalism & Mass Communication Quarterly*, 73(3), 635-645.

Roberts, C.W. (1989). Other than counting words: A linguistic approach to content analysis. *Social Forces*, 68, 147-177.

S. F. Weiss. Content analysis in information retrieval. Information storage and retrieval, Ithaca, NY, Cornell University, Department of Computer Science, Oct 70, 46p.

Smith, C. P. (Ed.). (1992). *Motivation and Personality: Handbook of Thematic Content Analysis*. New York: Cambridge University Press.

Solomon, M. (1993). Content analysis: a potent tool in the searcher's arsenal. *Database*, 16(2), 62-67.

Stahl, N. A. (1987). Developing college vocabulary: A content analysis of instructional materials. *Reading, Research and Instruction*, 26 (3).

Swetz, F. (1992). Fifteenth and sixteenth century arithmetic texts: What can we learn from them? *Science and Education*, 1 (4).

Thomas, S. (1994). Artifactual study in the analysis of culture: A defense of content analysis in a postmodern age. *Communication Research*, 21(6), 683-697.

Walsh, K., et al. (1996). Management in the public sector: a content analysis of journals. *Public Administration* 74 (2), 315-325.

Weare C. and Lin, W.-Y (2000). Content analysis of the World Wide Web. *Social Science Computer Review*, 18 (3) Fall 2000, pp.272-92.

Weber, R. P. (1990). *Basic Content Analysis, Second Edition*. Newbury Park, CA: Sage Publications.

Zollars, C. (1994). The perils of periodical indexes: Some problems in constructing samples for content analysis and culture indicators research. *Communication Research*, 21(6), 698-714.