Referência completa para citação:
FREITAS (H.) e MOSCAROLA (J.). Content analyzing qualitative data on Information Systems. **Association for Information Systems Americas Conference 1998**, Technology research in progress. Baltimore, MD, EUA, August 14-16, 1998, p. 941-943.

Content Analyzing Qualitative Data on Information Systems

Henrique Freitas
Associate Professor, GESID - PPGA - Escola de Adm. da UFRGS and CNPq Researcher (Brazil)
Visiting Researcher at ISRC, University of Baltimore (USA)
Docteur "nouveau régime" – Université Pierre Mendès France (Grenoble, France)
Rua Peru, 211 - Canoas/RS - 92420-300 - Brazil - Tel/fax: 55-51-4773610 - hf@ea.ufrgs.br

Jean Moscarola
Director of Ecole de Commerce, Vente, Economie et Gestion, Irege, Univ. de Savoie, Annecy (France)
Doctor in Management (Université de Paris), Sphinx's software Author, moscarola@univ-savoie.fr

Abstract

This paper presents general notions about content analysis, and its application on management and in the IS field, considering the use of computer software. The IS world community debated the use of qualitative research: our goal is to point out the potential contribution of computer-based content and lexical analysis as a method of qualitative data analysis.

1. Introduction

Document analysis, which includes content analysis and lexical analysis, follow classic methods like the judicial and sociological research. It presents a major common rational characteristic, being more or less intuitive, personal, and subjective. Like others, the historical method has validity problems, such as the authenticity of the text, interpretation validity, and the truthfulness of statements of fact. All of these analyses, have the defect of a non-systematized work, depending a lot on the value and competence of the researcher. With the proliferation of the means of communication (TV, radio, etc.), and especially of advertising on the Internet, some things became more obvious: the need to look for meanings, relationships, or even laws to measure influence factors; and the need to foresee certain events.

The notions presented are set in the European texts of Grawitz (1976, 1993) and Gavard-Perret & Moscarola (1995). The American literature is very recent (Krippendorf, 1980). Weber (1990) authored a short book. Silverman (1993) described methods for analyzing talk, text, and interaction. Miles and Huberman (1994) authored a book on Qualitative Data Analysis. The English Journal of Applied Management Studies published two papers on appplying content or lexical analysis, and we found some in the MISQ during 1997. The French literature is also receiving a breath of fresh air from Bardin (1996), Lebart and Salem (1994), and Moscarola (1990). In Section 2, general notions of content analysis are presented; its application is emphasized in Section 3. Some technical stages are in Section 4, and some cases are illustrated in Section 5. The usefulness of content analysis is presented as conclusion in Section 6.

2. General Notions about Content Analysis

The resources that the social sciences offers for our reflection are essentially composed of oral (speeches, annotations and reports of interviews, and conversations) and written communications (official texts, newspapers articles, letters, everything in the literary field, even history and politics). During the research, we gather data to understand, to explain opinions, conduct, or actions. It is very important that the researcher be able to analyze these data in a scientific way (standardized procedures, quantification) and not be satisfied to have a personal opinion or a guess. The text will be decomposed; it will be studied as a function of the words that it contains or the ideas it represents. Berelson (apud Grawitz, 1993, p.534) defines content analysis as: "… a research technique for the objective, systematic and quantitative description of the obvious content of the communications, having for objective to interpret them". Grawitz describes its characteristics:

- Objective: The analysis should proceed according to the pre-established rules, obeying guidelines sufficiently clear and precise in a way that allows different analysts, working on the same content, to obtain the same results.
- Systematic: The whole content should be ordered and integrated in the chosen categories as a function of the objective pursued. Elements of information related to the objective should not be overlooked.
- Quantitative: to evidence the significant elements, to calculate its frequency, etc.; sometimes not so important if looking more for themes than for exact measures.

There are some tools for content analysis associating the qualitative with the quantitative data (Weitzman and Miles, 1995; Kelle, 1995). They offer several ways to

navigate (or surf) the text to identify themes, to define new variables, which can be treated quantitatively; performing two-way or multivariate analysis with other contextual or more objective variables or data, facilitating more robust and precise inferences and conclusions.

The Internet provides a wealth of "free" data for researchers and curious data analysts to conduct a variety of investigations, which could generate unique information leading to useful conclusions. Grawitz (1993, p.536-538) develops three approaches: exploration or hypotheses definition versus hypotheses verification; quantitative (try to accumulate the frequency of themes, words, or symbols) versus qualitative (based on the presence or absence of a given characteristic); and direct (more commonly used, consists of counting the responses just as they appear) versus indirect (beyond what is had as clear and obvious result, to obtain for inference, even that the author left implied). The qualitative and quantitative *worlds* are not (or should not be) mutually exclusive; and yes, they are supposed to complement each other.

3. Content Analysis Application and "Careful" Use

We can apply this kind of data analysis technique in several ways, which can involve different kinds of source, storage or media... Who speaks? To say what? To whom? How? With what result? The content analysis can lift or even to solve problems in the social sciences as well as in different areas of management.

We can easily "surfing" inside a content, and we can quickly have some statistics regarding the lexicon (all the words within a text with their frequency of appearance). Content analysis tools make available to the researcher methods combining the statistical lexical and the data analysis methods, the syntactic analysis, and the lexical surfing and the reading assisted by computer. Moscarola (1995, p. 78) describes the levels of analysis: we can reduce the body of the original text to the top of the lexicon (examining the most frequently words we can have an idea of its content). We can also transfer some of the contents to a report to illustrate something; or still group words in a new word or concept; or establish the statistics of the words of the text according to an external non-textual variable, like "department", gender, age, etc.

4. Content Analysis Technical Stages and Value

As in almost every research, the first step is the idea itself, as well as its objective. The means of obtaining an answer is to ask a question, as in an interview. As the questionnaire allows us to guide an interview, the research analyst will select the categories to guide the analysis based on the data in the documents. We first need to choose categories, and then start the main analysis. Content analysis is a refined technique, therefore delicate, and it demands much dedication, patience, and time to satisfy the investigator's curiosity, in addition to intuition, imagination to notice what is important, and creativity to choose the categories. The analyst should have discipline and perseverance, rigidity when decomposing content or to count results of analysis. These are only some of the main aspects that Grawitz (1993, p. 553-558) points out, along with fidelity, logical validity, inference, and empirical validity. Is the study prediction precise? Instead of convictions, prudence and humility are recommended when drafting the conclusions. Grawitz believes that the experience and the analyst's training will countersign the value of its analysis.

5. Application of Qualitative Analysis

We will now analyze a different and curious case involving an American presidential election debate, and others examples using the Sphinx® software (Moscarola, 1990), a friendly toll which helps users to conceive a questionnaire or database, entry or import, and analyze quanti and especially qualitative data. The goal is to allow the manager or the researcher to "touch and explore the data" by himself or herself.

5.1 - Data Mining with Speech Qualitative Data: The Election Debate "Dole versus Clinton" This is a very interesting field that will continue to be developed for far beyond what it is today. Using the Internet, we can very quickly download the speeches from a political, social, or whatever discussion, and make many different lexical analysis. Let us explore some data regarding the Dole and Clinton elections. We have at least two possible applications of content analysis in these cases: the content itself or the main ideas, and some statistics based on the texts, or the speech-acts.

We can know very easily how many words each one spoke, what was the average length in words of each speech or contribution in the discussion, whether they repeated each one of their main arguments, whether they said something "unique", whether they were "original" in their speech, etc. We can even look only for the verbs, meaning the actions; same thing regarding adjectives or nouns. We will then be able to identify the objects or figures of each speaker, their main keywords, their main messages to the people, their main ideas concerning some important subjects like education, social security, health, etc. and then compare them this way!

Are they really discussing something? Are they only playing a game where they have known very well what they would like to say long before the debate, which means were they ignoring their opponent during the discussion? We can find answers to these questions by making lexical analysis of their speeches! If we explore the data from the Dole-Clinton debate, where Mr. Lehrer was the moderator, we can point out that 17,407 words were spoken, with a mean of 16 words in each time someone made a contribution to the debate (or the mean phrase

length was 16 words), and we can still know more: Clinton spoke 7,962 words, with a mean of 19.5 words in each intervention; while Dole spoke 8,488 words, with a mean of 14.3 words in each intervention. We can identify the main repeated segments or expressions very easily, they were for example: Clinton - worked hard, crime bills, Middle East, better of than we were four years ago. Dole: economic package, people watching, United Nations, Mr. President.

If we divide the Dole-Clinton speech data into five equal parts, creating a new variable (PART), and, after that, we create another new variable, merging the options Dole and Clinton (variable SPEAKER) with each one of the five parts (variable PART), crossing the new data with the main arguments we have a chart that shows us that they never really discussed! Each one gave the people his own message. How many television viewers or analysts even noticed that it was not a debate?

A chart (available upon request or in the Conference) was produced by the technique Correspondence Analysis (CA) - (Greenacre, 1984; Burt, 1950; Carrol, Green and Schaffer, 1986, 1987; Hoffman and Franke, 1986), and can also be interpreted as a MCA, main component analysis (Lagarde, 1995). According to Moscarola (1990), the technique of CA is the best way to visualize a crossing table. A single geometric interpretation of CA is enough to understand its results (Greenacre and Hastie, 1987). Anyway, it is perceptual and intuitive maps (Hair, 1994; Fox, 1988). We have "a posteriori" better conditions to define names for each one of the axes, looking after the positive and the negative contributions.

5.2 - Examples: Data Mining, Hospital, IS Planning, etc

Moscarola (1993, 1994, 1995) conducted several different studies, like the **internal communication** in a bank compared with the customers' expectations (totally different!) by Bachelet and Moscarola (1995).

Data Mining Proceedings of AIS 97 and of ICIS 96: world conferences are now available via the Internet or on CD-ROM. We defined a database structure (like "meeting", "session", "authors", "title", abstract", etc), and have available data in order to compare the main subjects of each conference; identifying some subjects of our interest, as in a literature search.

Stumpf and Freitas (1996) conducted a two-year study inside a university **Hospital**, using a Focus Group technique to collect data from six different groups of professionals: medical care and teaching; nursing care and teaching; management; and research; defining a new patient record, representing a lot on savings.

Domenjoz, Gavard-Perret, and Moscarola (1995), in "Price and communication: how do they interrelate? an analysis of car **advertisements** published in English", analyzed the content of 52 car advertisements using an observation protocol combining descriptive and subjective observations of the media and the visuals coded into a grid with recording of all the text.

Crossing qualitative and quantitative data **to improve the organizational IS planning**, in a large organization, we asked some questions to people from each department or managerial level, and then tried to point out what were the main critical factors of success, the pitfalls, the barriers, etc. M. Jenkins, from ISRC-University of Baltimore (USA) is applying these techniques.

In a cross-cultural exploratory survey study about **decision-making** process, involving respondents from Brazil, France, and USA, we collected 72 different data elements from each one of the almost 300 respondents (Macadar, Costa, Freitas, Becker, and Moscarola, 1997).

We did a research in 1994 to access the **Quality Programs** in the industry of the State of Rio Grande do Sul (RS)/Brazil (Ruas, Pires, Freitas, Cunha & Antunes, 1994), involving 120 different companies (30 small, 30 medium and 60 large). Almost 300 different variables were stored after interviews. Large companies were using indicators that can be quantified (productivity, participation, financial), while the small companies were using sources more related to qualitative factors (customers, salespersons, etc).

6. Conclusion: Useful Content Analysis

The access and the "surfing" regarding the text or document data or the qualitative research methods are then facilitated. The tools are now powerful. However, the analysis of textual data does not change the meaning of the data. Even if analysis reduces the "noise" in the data, the reduction of the long, and sometimes annoying readings, the faster production of conclusions and reports drives everybody to an interpretive reading. Interpretation is often dangerous since it is fast and it appears to be falsely objective. We should so be careful in our investigation and conclusions. Due to its complexity, it should be investigated in which case content analysis or lexical analysis should be applied. When sufficiently defined and detailed, content analysis allows us to pass of the simple description and to reach the objective of every scientific research: the discovery of explanations and causal relationships. The value of content analysis depends on the quality of the conceptual elaboration done *a priori* by the researcher, of the exactitude with which it will translate itself in variables, of the analysis outline or categories, and also of the agreement among the reality to analyze and these categories. It is time to go ahead with many more qualitative studies, and to educate our managers, starting with our children, that the world is not only quantitative, but also qualitative. At least, what good quantitative study could not having been preceded by a qualitative one?

Keywords: Lexical & Content Analysis - Qualitative & Quantitative Data Analysis

References, charts and a **full paper** are available upon request from author (hfreitas@portoweb.com.br).

The software Sphinx Lexica® (e-mail: f&c_cons@ portoweb.com.br – http://www.freitasecunha.com.br) for quantitative and qualitative data analysis served as support to the data analysis in this study. This work is supported by Brazilian Government Research Agencies (CNPq, Fapergs, Propesp/UFRGS, and Capes/Cofecub).