

Using Content Analysis in LIS Research: Experiences with Coding Schemes Construction and Reliability Measures

Chi-Shiou Lin¹ and Wei Jeng²

¹National Taiwan University, Dept. of Library & Information Science, Taipei, Taiwan

²University of Pittsburgh, School of Information Sciences

Abstract: Content analysis is now also popular in LIS research. It is a flexible and unobtrusive method to analyse the meanings of narratives objectively. However, the objectivity of analysis relies heavily on a well-constructed coding scheme as well as the consistency of coding behaviour. In this paper, we shared our experiences with content analysis in three very different LIS research. We emphasized on the development of mutual exclusive and comparable coding schemes in each study in order to enhance the internal validity and extensibility for future research. We also discussed on the reliability considerations. Hopefully it may encourage more discussion on the methodology and contribute to the vigour of the content analysis based investigations.

Keywords: Content analysis; Coding scheme; Coding reliability

1. Introduction

Originated in sociology and widely used in mass communication research (Riffe, Racy, & Fico, 2005; Zen, Ding, & Milojević, 2013), 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, to name a few. Content analysis is a flexible and unobtrusive method to analyze the meanings of narratives objectively. However, the objectivity of analysis relies heavily on a well-constructed coding scheme as well as the consistency of coding behavior. Coding schemes determine the validity of the analysis, while standardization and consistency of coding procedures safeguard the reliability of study results. Many factors can influence the research design decisions in undertaking a content analysis. For example, the research inquiry and the complexity of the studied objects (e.g., textual

narratives, images, Web content, citations) significantly influence the construction of the coding scheme. The amount of data to be analyzed may influence how reliability will be established and measured (e.g., intercoder or intracoder reliability, single or multiple coders). While methodology literatures have offered general guidance on how to conduct content analysis, there are many conceptual and technical decisions to be made that can vary greatly depending on the research purposes and research contexts.

In this paper, we share our experiences on how to build a sound and valid coding scheme and how to establish coding reliability. Specifically, we discuss how research goals and theoretical perspectives shape the codes that signify the attributes and meanings of the analyzed objects as well as how research contexts influence the adoption of reliability measures. We use our previous research as examples to highlight the rationales behind the decision. The projects include one study analyzing government Web content (will be referred to as “government Web content study” in the following text), one analyzing the abstracts of social sciences research journals (the “abstract study”), and one analyzing citation contexts in social sciences and humanities journals (the “citation context study”). With the highly diverse research topics, studied objects (texts), study goals, and research contexts, we hope to caution future researchers with potential challenges and possible solutions in undertaking content analysis.

2. Coding Schemes Construction

A major step in the instrumentation process of a content analysis research is to construct an informative and valid coding scheme. While the determination of categories (codes) depends on the substantial questions being investigated, an important decision to make is whether the categories are mutually exclusive. In most cases, mutual exclusiveness is required or mandated for later statistical analyses. Codes that are not conceptually distinct may suffer from problematic interpretations (Weber, 1990).

We want to further emphasize the importance of comparability of the coding scheme with the previous and possible future research. The development of the categories should be based on previous relevant literatures so that the findings will be comparable across different empirical studies. Better yet, codes generated from adequate theorizing will further enhance the depth and insights of the investigation and may inform future extended research.

In our government Web content study, one of the study goals was to identify the intentions of the messages conveyed in the homepages of the government agencies' Web sites (Lin, 2010). While there was no similar content analysis study like ours in the previous literatures, we developed the categories from referencing two previous government Web content studies (Eschenfelder & Miller, 2006; Mahler & Regan, 2007) and derived that the crafting of government Web content may be motivated by three different intentions: public relation, policy advocacy, and disclosure of business-related information. Upon further consideration, we considered that the intentions were not mutually exclusive, for a message may be crafted to serve more than one of the purposes.

As a result, the categories were treated as independent measures, and each government agency homepage in our sample was evaluated in terms of its strength in the three intensions.

Similarly, when we developed the coding scheme for the abstract study (Lin & Chen, 2014), not much similar previous research existed. The study goal was to determine the quality of the abstracts of journal articles in social sciences, and we approached quality assessment from inspecting the content structure and richness of information conveyed in the sample abstracts. Previous papers, however, suggested possible ways for examining the structure of an abstract, for examples, the IMRD (introduction-methodology-result-discussion) structure (Milas-Bracović & Zajec, 1989), the ANSI/ISO standard for abstract writing (Tibbo, 1993), and other abstract content elements suggested in previous research (e.g., Hahs-Vaughn & Onwuegbuzie, 2010; Šauperl & Klasinc, Lužar, 2008). We finally chose the IMRD structure as the basis for the coding scheme for its better applicability across different social sciences disciplines. Here, our goal was to produce coding results that would be comparable within our own study and to previous research findings whose coding may easily correspond to ours. We used the content element scheme (IMRD) to observe whether a specific element is present in an abstract and how rich the information is. We drew sample abstracts from six subject disciplines (sociology, psychology, law, economics, management, and LIS) and coded 600 English abstracts and 600 Chinese abstracts. The results showed that a rather small portion of the abstracts was structurally sound. Most of the social sciences abstracts lacked one or more structural elements and were weak in conveying certain information.

In the third study, the “citation context” study, we again based on previous literatures to develop the coding scheme. Citation context analysis is essentially a content analysis of citations involving careful reading of the text surrounding a citation so as to determine what functions it serves for that paper. This type of research is needed because many bibliometrics-based research evaluations have been founded on problematic assumptions about citation behaviour and motivations. Our study therefore addressed two specific problematic assumptions and examine to what degree they may affect the validity of citation-based evaluation in social sciences and humanities (SS&H) disciplines, i.e., the assumption that each citation in a paper has equal contribution to the work, and the assumption that the cited work influences the citing paper positively (Lin, 2013; Lin, Chen, & Chang, 2013).

Different from the two aforementioned studies, however, is that there had been previous citation context analyses in existing literatures. But upon closer examination of their coding schemes, several problems emerged. First, many of the coding schemes contain codes that are not mutually exclusive (e.g., Voos & Dagaev, 1976; Peritz, 1983; Frost 1989). The problem of mutual exclusiveness is particularly important in citation context analyses because citation behavior is complicated and multidimensional. An author may use a citation to serve multiple purposes, given it functional, rhetorical, or social (Liu, 1993; Bornmann & Daniel, 2008). Therefore, we strived to develop a coding scheme in which codes are conceptually distinct and mutually exclusive. We settled on

Murugesan & Moravcsik's (1975) classification of citation functions and used their idea to develop our codes. They summarized the functions of citations along four dimensions that appeared to be the most applicable for analyses across subject disciplines:

1. Conceptual vs. operational: whether a citation offers conceptual or operational support to the paper.
2. Essential vs. perfunctory: whether a citation is essential to the paper.
3. Evolutionary vs. juxtapositional: whether a citation directly contributes to the construction of a major thesis or simply offers minor supportive information.
4. Confirmative vs. negational: whether the author agrees or disagrees with the citation.

Upon closer examination, two of the dimensions were conceptually more ambiguous and may not be mutually exclusive; they were the conceptual vs. operational dimensions and the evolutionary vs. juxtapositional dimension. In contrast, we found that the essential vs. perfunctory dimension and confirmative vs. negational dimensions were operationally easy to distinguish and can be turned into mutually exclusive categories, so we decided to use them to form a two dimensional coding scheme that are both operationally viable and conceptually distinct (Table 1).

		Confirmatory	Negational
Organic (Essential)	Concept	OCC	OCN
	Factual Statement	OFC	OFN
	Methodology	OMC	OMN
Perfunctory		PC	PN

Table 1 The classification scheme of citation functions

The operational definitions of the eight citation functions in the matrix are as follow:

Organic usages:

- Organic-concept-confirmative (OCC): an essential citation that supports a major concept in the paper; author attitude being positive to the cited source
- Organic-factual-confirmative (OFC): an essential citation that supports a factual statement in the paper, e.g., references of dates, historical evidence; author attitude being positive to the cited source
- Organic-methodology-confirmative (OMC): an essential citation that support a major methodological move in the paper; author attitude being positive to the cited source
- Organic-concept-negational (OCN): an essential citation related to a major concept in the current paper; author attitude being negative to the cited source
- Organic-factual-negational (OFN): an essential citation related to a factual statement; author attitude being negative to the cited source

- Organic-methodology-negative (OMN): an essential citation related to a methodological issue of the current paper; author attitude being negative to the cited source

Perfunctory usages:

- Perfunctory-confirmative (PC): a minor citation that does not influence the major theses of the paper, e.g., insignificant citations about research background, extended readings, or for rhetoric purposes; author attitude being positive to the cited source
- Perfunctory-negational (PN): a minor citation that does not influence the major theses of the paper; author attitude being negative to the cited source

The coding scheme is applicable to analyses of scholarly texts in different formats and of different subject disciplines. The codes can also be matched to the citation function categories used in previous works so that comparisons across studies are possible. We believe that systematic analyses generated from such a coding paradigm may shed lights on the usages of citation as manifested in the scholarly literatures and show the strength, weakness, and validity of bibliometrics research in the SS&H domains.

3. Reliability Measures

Another methodology issue that influences the quality of content analysis research is the reliability of data coding. The methodology literatures suggest that the reliability of content analysis is established in the following steps:

3.1. Decide appropriate indices and the level of reliability:

A researcher should first select one or multiple appropriate statistics for measuring reliability of content analysis. Common methods included: percentage agreement, Holsti's Method, Scott's Pi (π), Cohen's Kappa, and Krippendorff's alpha (Neuendorf, 2002; Lombard, Snyder-Duch, & Bracken, 2002). In addition, one should decide a minimum acceptable level for the preferred indices. For a simple percent agreement, ".90 or greater are nearly always acceptable, .80 or greater is acceptable in most situations, and .70 may be appropriate in some exploratory studies for some indices" (Lombard, Snyder-Duch, & Bracken, 2004, p.3). Cohen's kappa coefficient is generally considered to be a more conservative measure index (Lombard, Snyder-Duch, & Bracken, 2004). Researchers may apply a relatively lower criterion when comparing with a simple percent agreement calculation: Cohen's kappa of .81 or greater are almost perfect agreement, .61 to .80 is substantially reliable in most cases (Viera & Garrett, 2005).

3.2. Coder training & pilot testing

Secondly, the researcher will have to create a procedure instruction and guide book for the participating coders. The reliability of coding may be assessed informally with a small number of units. For example, select a subsample (N=30) for a pilot test of intercoder reliability indices (Lombard, Snyder-Duch, & Bracken, 2004). Piloting the coding scheme can be considered as a critical stage before applying it to the full data: The result of pilot testing helps find out inconsistencies of coders or other inadequacies in terms of the

category construction (Prasad, 2008). A researchers may refine the instrument and coding instructions, or even replace one or more coders until the informal assessment suggests an adequate level of agreement.

3.3. Assess reliability formally

When the level of reliability obtained in the pilot test is adequate, a researcher should use another representative sample to assess reliability for the entire sample. The appropriate size of this subsample “should not be less than 50 units or 10% of the full sample, and it rarely will need to be greater than 300 units” (Lombard, Snyder-Duch, & Bracken, 2002, p.601). Note that larger reliability samples are required when the full sample is large. Kaid and Wadsworth suggested that “when a very large sample is involved, a subsample of 5-7 percent of the total is probably sufficient for assessing reliability.” (as cited in Lacy & Riffe, 1996).

3.4. Report intercoder reliability and dealing with those “disagreement”

Lombard et al. evoke authors report more than one measure of reliability. Common methods included: percentage, Cohen’s Kappa and Krippendorff’s alpha. They further suggest the disagreements of coders be either resolved by

- 1) randomly selecting the decisions of the different coders
- 2) using a “majority rule” if there are an odd number of coders
- 3) having another researcher or an expert who serves as a “tie-breaker” to reach the final decision
- 4) multiple coders discuss and build consensus for the final decision.

A researcher may choose one or more approaches above, as one should be able to justify and report which procedure is adopted (Lombard, Snyder-Duch, & Bracken, 2002, p.601).

To further enhance coding reliability, two additional considerations may be added to the reliability testing framework:

1. Report multiple reliability indices

Researchers suggest that researchers avoid use sole indicator to report reliability. For example, one should not only use simple percent agreement to calculate reliability (Lombard, Snyder-Duch, & Bracken, 2002). Lombard et al. further proposed researchers use a second statistic that accounts for agreement expected by chance (Lombard, Snyder-Duch, & Bracken, 2004).

2. Using random sample to assess the possible level of agreement

As Riffe et al. (2005) indicated, “if two researchers using randomly sampled content achieve a 90% level of agreement, the actual agreement they would achieve coding all material could vary above and below that figure according to the computed sampling error. Therefore, if the desired level of agreement is 80%, and the achieved level on a coder reliability test is 90% plus or minus 5 percentage points, the researchers can proceed with confidence that the desired agreement level has been reached or exceeded. However, if the test produced an 84.0%, the 5% sampling error would include a value of 79% that is below the required standard.”

Another reliability-related problem that content analysis researchers must consider is whether it is permissible for the coders to discuss on their independent coding results and change the original codes assigned. Our

recommendation is to consider the merits of such discussions for the study goal in order to determine if discussions and changes are desirable. This is particularly important if the operational definitions of the codes are still subject to varying interpretations. In all of our three studies, discussions and changes are permitted, but rules varied. In the government Web sites and abstract studies, the categories of government intentions and IMRD elements were not viewed as mutually exclusive codes, but as independent criteria by which coders judge if each element was present in an analyzed text and to what extent it was manifested (from non-presence, weak, strong, very strong). Coders were permitted to discuss on whether a specific criterion was present or absent as misreading, personal bias, or situational factors may influence coders' initial judgment. Discussion may alert coders to a previously neglected aspect of the analyzed text and voluntary code change is viewed as enhancing coding reliability. But the degree of manifestation of a criterion was left to individual judgment. In the citation context analysis where the categories became mutually exclusive, discussion post independent coding was allowed particularly to address the decisions on confirmative/negational citations. This was due to the fact that an author may first cite a particular citation in a seemingly neutral tone in a particular paragraph, but in a later paragraph the author began to disagree with what was expressed by that citation, and therefore the citing motivation for that citation in the whole article's context was negational. Discussions helped to correct the previous misjudgments and may enhance reliability of the result.

4. Conclusions

In this paper, we shared our experiences with content analysis in three very different LIS research. We emphasized on the development of mutual exclusive and comparable coding schemes in each study in order to enhance the internal validity and extensibility for future research. We also discussed on the reliability considerations including reliability measures and procedures recommendations from various methodology guides as well as the adoption of coder negotiation for correcting miscoding in the initial coder judgment. We hope our experiences may encourage more discussions on the methodology of growing popularity and may contribute to a vigorous future of the content analysis research.

References

- Bornmann, L., & Daniel, H.-D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1), 45-80.
- Eschenfelder, K. R., & Miller, C. A. (2006). Examining the role of Web site information in facilitating different citizen-government relationships: A case study of chronic wasting disease Web sites. *Government Information Quarterly*, 24(1), 64-88.
- Frost, C. O. (1989). The literature of online public access catalog, 1980-1985: An analysis of citation patterns. *Library Resources and Technical Services*, 33(4), 344-357.
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage.
- Lacy, S., & Riffe, D. (1996). Sampling error and selecting intercoder reliability samples for nominal content categories: Sins of omission and commission in mass

communication quantitative research. *Journalism & Mass Communication Quarterly*, 73, 969–973.

Lin & Chen (2014). An analysis of the guidelines, structures and readability of the abstracts for the social science scholarly journals. Manuscript submitted for publication.

Lin, C.-S. (2010). *Exploring the business models of the Web content management in Taiwan's government agencies* (NSC99-2410-H-002-258). Taipei: National Taiwan University, Library & Information Science.

Lin, C.-S. (2013). *A study of citation functions in Taiwan's humanities and social sciences research journals* (NSC102-2410-H-002-202). Taipei: National Taiwan University, Library & Information Science.

Lin, C.-S., Chen, Y.-F., & Chang, J.-Y. (2013). Citation functions in social sciences and humanities: Preliminary results from a citation context analysis of Taiwan's history research journals. In *Proceedings of the ASIS&T 2013 Annual Meeting*.

Liu, M. (1993). A study of citing motivation of Chinese scientists. *Journal of Information Science*, 19, 13-23.

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research*, 28(4), 587-604.

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2004). *Practical resources for assessing and reporting intercoder reliability in content analysis research projects*. Retrieved April, 19, 2004.

Mahler, J., & Regan, P. M. (2007). Crafting the message: Controlling content on agency Web sites. *Government Information Quarterly*, 24(3), 505-521.

Milas-Bracović, M., & Zajec, J. (1989). Author abstracts of research articles published in scholarly journals in Croatia (Yugoslavia): An evaluation. *Libri*, 39(4), 303-318.

Moravcsik, M. J., & Murugesan, P. (1975). Some results on the function and quality of citations. *Social Studies of Science*, 5(1), 86-92.

Neuendorf, K. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.

Prasad, B. D. (2008). Content Analysis. A method in Social Science Research. *Research methods for Social Work*. New Delhi: Rawat, 173-193.

Peritz, B. (1983). A classification of citation roles for the social sciences and related fields. *Scientometrics*, 5, 303-320.

Riffe, D., Lacy, S., & Fico, F. G. (2005). *Analyzing media messages: Using quantitative content analysis in research*. Psychology Press. Chicago.

Šauperl, A., Klasinc, J., & Lužar, S. (2008). Components of abstracts: Logical structure of scholarly abstracts in pharmacology, sociology, and linguistics and literature. *Journal of the American Society for Information Science and Technology*, 59(9), 1420-1432.

Tibbo, H. (1992). Abstracting across the disciplines: A content analysis of abstracts from the natural sciences, the social sciences, and the humanities with implications for abstracting standards and online information retrieval. *Library and Information Science Research*, 14(1), 31-56.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam Med*, 37(5), 360-363.

Voos, H., & Dagaev, K. S. (1976). Are all citation equal? Or, did we op cit your idem? *The Journal of Academic Librarianship*, 1(6), 19-21.

Weber, R. P. (1990). *Basic content analysis*. Thousand Oaks, CA: Sage.

Zhang, G., Ding, Y. and Milojević, S. (2013), Citation content analysis (CCA): A framework for syntactic and semantic analysis of citation content. *Journal of American Society for Information Science and Technology*, 64(7), 1490–1503.