Chapter Eight

The completest knowledge of the laws of nature does not carry with it the power of prediction, nor of mastery over nature. If the universe is a machine its levers and wheels are too fine for our hands to manipulate. We can learn and guide its large-scale motions only. Beneath our veiled sight it quivers its eternal quest. (Max Born)

Keyword-Associated-COntent-Variation-Analysis Model (KACOVA Model): An Overview

Content analysis is a highly flexible research method that has been widely used in library and information science with varying research goals and objectives. As a research methodology, content analysis has its roots in the study of mass communications in the 1950s, which was based on a basic communications model of sender / message / receiver. Initially researchers emphasized making inferences based on quantified analysis of easily identifiable aspects of text content. Since then, researchers in many fields, including anthropology, library and information studies (LIS), management, political science, psychology and sociology, have used content analysis. This research method is applied in qualitative, quantitative, and sometimes mixed modes of research frameworks and employs a wide range of analytical techniques to generate findings and put them into context. This model characterizes content analysis as a systematic, rigorous approach to analyzing research papers obtained or generated in the course of research. In this model, the content analysis technique is adopted to measure the variation in content of a subject from its research output to contemplate the unique needs of the research queries and strategies. This model has developed a set of techniques and approaches for analyzing keywords grouped under its four major properties. The predecessor of keyword analysis is textual analysis. The syntactic, syntagmatic, and pragmatic aspects of text have

been considered while selecting keywords. There are so many procedures of content analysis in terms of both analytical goals and the means or processes developed to pursue them, which include, for example, besides content analysis, conversational analysis, discourse analysis, ethnographic analysis, functional pragmatics, rhetorical analysis, and narrative semiotics. Although these approaches are alike in their reliance on communicative material as the raw material for analysis, they vary in the kinds of questions they address and in their methods. This model applies the content analysis method in the micro-domain of a subject to encompass the variation in content over a stipulated time span.

There exist various definitions of content analysis that reflect its historical development. For the purpose of this model, content analysis is a research technique for selecting appropriate keywords from titles, abstracts and texts of the research papers of a particular subject over a time span of twenty years. The notion of inference is especially important in content analysis. The researcher uses analytical constructs, or rules of inference, to move from the text to answer the research queries. The two domains, the texts and the context, are logically independent, and the researcher draws conclusions from one independent domain (the texts) to the other (the context). The context of this model highlights the study of the pattern of temporal variation of the subject-content over a time span with the aid of the keywords. The keywords are reckoned as yardsticks to measure the subject-content and its temporal variation in this model. The collective occurrence of keywords sketches the complete portrait of the subject from the fundamental level. The collective occurrence has been interpreted by numerical analysis and various statistical inferences are drawn. In LIS studies the analytical constructs are not always explicit. The analytical constructs may be derived from (1) existing theories or practices; (2) the experience or knowledge of experts; and (3) previous research. In this model, the analytical construct has been derived from the experience only due to unavailability of both existing theory and previous research as well.

The content of a particular research paper published at a particular time is fixed, but as a research paper is the representative of a subject; therefore the content of a collection of research paper over a stipulated time span may be recognized as the development of the subject concerned for the said time span. As far the content of the whole subject is concerned, it does not remain steady over the entire span but varies. This model also encounters the quantitative and qualitative analysis of the variation in content of a subject. It briefly describes the steps involved in content-variation analysis, differentiates between quantitative and qualitative analysis, and shows that content-variation analysis serves the purposes of both quantitative research and qualitative research.

Content analysis by keywords is based upon the assumption that a paper's keywords constitute an adequate description of its content. The keywords also indicate the links a paper establishes between different subjects. Two different keywords co-occurring within the same paper are an indication of a link between the topics to which they refer (Cambrosio et al., 1993). The presence of many co-occurrences around the same word or pair of words points to a locus of strategic alliance within papers that may correspond to a research theme. Content analysis by keywords reveals patterns and trends in a specific discipline by measuring the association strengths of key terms representative of relevant publications produced in this area. The main feature of content analysis by keywords is that it visualizes the intellectual structure of one specific discipline into maps of the conceptual space of this field, and that a time-series of such maps traces the changes in this conceptual space.

Keywords are the most important research elements in this model of content-variation analysis. There are two ways to extract keywords from journal articles, conference papers, reports or even chapters of books. The scope of this model is restricted to the journal articles only. One way is to extract keywords from keyword lists, title, abstract, and sometimes even including classification codes. Many journals, abstracting services and databases already provide such

keywords. The resulting lists of keywords have been standardized to eliminate different spellings and variants of the same terms. Coulter et al. (1998) selected keywords chosen by professional indexers. They believed that it is useful to study a fixed system that imposes a common nomenclature. Professional indexers' experiences assure standard application of that taxonomy. Looze and Lemarie (1997) conducted co-word study based on the keywords proposed by the experts. Some researchers downloaded keywords from online databases, which are added by database indexers and authors (Courtial, 1994; Law & Whittaker, 1992; Courtial, Cahlik, & Callon, 1994). Noyons and van Raan (1998b) mapped the overall structure in the field of neural networks by using the co-occurrence of classification codes.

This model focuses not only on content analysis, but also on content-variation analysis with the age of a subject. This model deals with content-variation analysis not in all forms of textual analysis, but in the form of keyword-cluster analysis. This model emphasizes quantitative, rather than qualitative approaches to content analysis though both are used in information studies. Content analysis is a flexible research method that can be applied to many problems in information studies, either as a method by itself or in conjunction with other methods. After defining quantitative keyword-cluster analysis, this model goes through the basic steps in a content analysis study to orient the subject-analysis. The most important component of the content-analysis study is data that provide useful evidence for testing hypotheses or answering research queries. The keywords are central data for this model. Another key factor is that the data communicate; they convey a message from a sender to a receiver.

The selection of keywords from a text is a vital task, and there is another crucial step prior to this task, that is to understand the text for recognizing the central theme. Beaugrande and Dressler (1981) suggest seven criteria for defining a text, which is the more common form of data for content analysis:

1) Cohesion
2) Coherence
3) Intentionality
4) Acceptability
5) Informativity
6) Situationality
7) Intertextuality

In other words, text appropriate for content analysis is composed of linguistic elements arranged in a linear sequence that follows rules of grammar and dependencies and uses devices such as recurrence, anaphora and cataphora, ellipsis, and conjunctions to cause the elements to "hang together" to create a message (cohesion). The text has meaning, often established through relationships or implicature that may not be linguistically evident, and draws on frameworks within the recipient for understanding (coherence). The writer or speaker of the text intends for it to convey meaning related to his attitude and purpose (intentionality). Conversely, recipients of the message understand the text as a message; they expect it to be useful or relevant (acceptability). The text may contain new or expected information, allowing for judgments about its quality of informing (informativity). The situation surrounding the text affects its production and determines what is appropriate for the situation and the culture (situationality). The text is often related to what precedes and follows it, as in a conversation (one interpretation of intertextuality), or is related to other similar texts, for example, others within a genre, such as transcripts of chat sessions (another meaning of intertextuality).

The texts used in the present study are generated in connection with the research project, as the journal articles are only taken under consideration. The text generated from research output reflects dynamic content of the subject, and the keywords selected from such texts can trace the true locus of the time-dependent evolution of the subject. This is a keyword-associated dynamic model

for qualitative and quantitative investigation of the characteristics of a subject. The keywords are the tool, the content-analysis is the methodology and the temporal variation study is one of the major objectives involved in this model, i.e. KACOVA model.

The main steps involved in the KACOVA model are illustrated below:

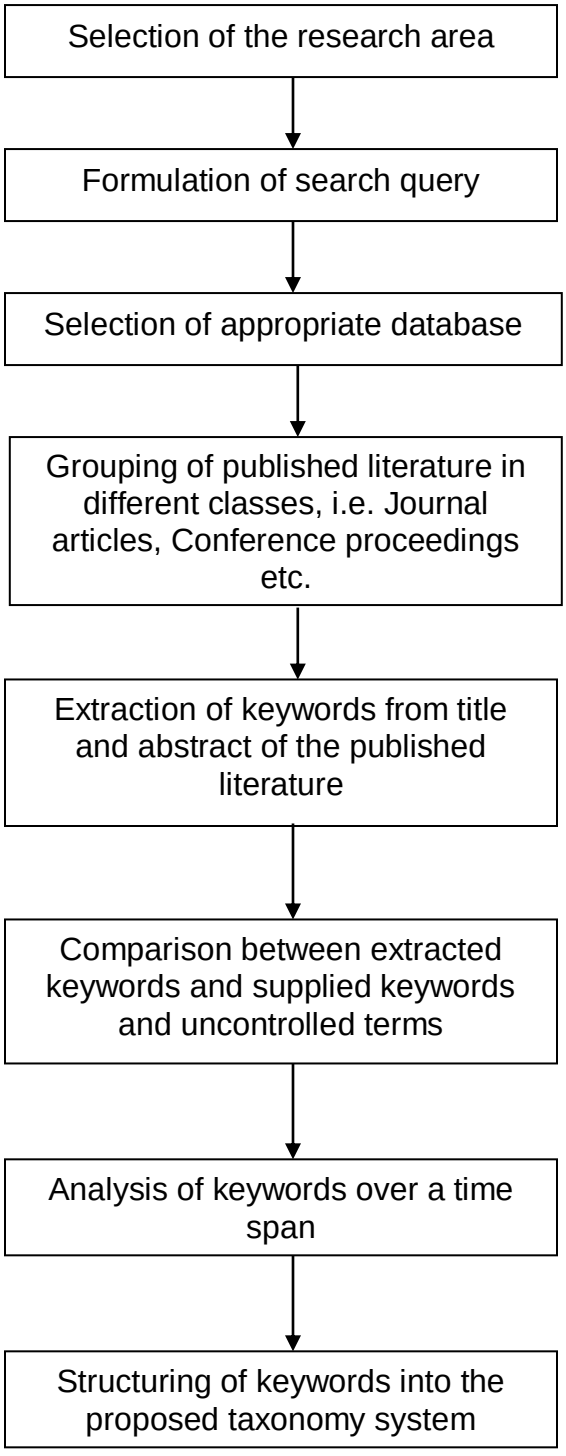

Diagram 8.1: KACOVA model

References

Cambrosio, A., Limoges, C., Courtial, J.P. and Laville, F., 1993. Historical scientometrics? Mapping over 70 years of biological safety research with co-word analysis. *Scientometrics* 27 (2): 119–143.

Coulter, N., Monarch, I. and Konda, S., 1998. Software engineering as seen through its research literature: A study in co-word analysis. *Journal of the American Society for Information Science* 49 (13): 1206–1223.

Courtial, J.P., 1994. A coword analysis of scientometrics. *Scientometrics* 31 (3): 251–260.

Courtial, J.P., Cahlik, T. and Callon, M., 1994. A model for social interaction between cognition and action through a key-word simulation of knowledge growth. *Scientometrics* 31 (2): 173–192.

De Beaugrande, Robert-Alain, and Wolfgang Dressler. 1981. *Introduction to text linguistics.* London; New York: Longman.

Law, J. and Whittaker, J., 1992. Mapping acidification research: A test of the co-word method. *Scientometrics* 23 (3): 417–461.

Looze, M.D. and Lemarie, J., 1997. Corpus relevance through co-word analysis: An application to plant proteins. *Scientometrics* 39 (3): 267–280.

Noyons, E.C.M. and van Raan, A.F.J., 1998. Advanced mapping of science and technology. *Scientometrics* 41 (1–2): 61–67.

Noyons, E.C.M. and van Raan, A.F.J., 1998. Monitoring scientific developments from a dynamic perspective: Self-organized structuring to map neural network research. *Journal of the American Society for Information Science* 49 (1): 68–81.

Van Raan, A.F.J. and Tijssen, R.J.W., 1993. The neural net of neural network research. *Scientometrics* 26 (1): 169–192.

Van Raan, A.F.J., 1997. Scientometrics: State-of-the-art. *Scientometrics* 38 (1): 205–218.