A Semiotic-Conceptual Analysis Framework as a Qualitative Method

Uta Priss[0000-0003-0375-429X]

Ostfalia University, Germany
www.upriss.org.uk

1 A Framework for the Initial Steps of a Data Analysis

Formal Concept Analysis (FCA) is a mathematical method of data analysis and knowledge representation that was invented in the 1980s (Ganter & Wille 1999). It provides a mathematical formalisation of the notions of "concept", "extension", "intension", "formal object", "formal context", "formal attribute" and "conceptual ordering" and supports exploring data sets, clustering, determining implications and more. The translation of these notions into a mathematical method produces both a loss and a gain of information: on the one hand, the richness, amount of detail, associations and embedding which these notions have in their original domains is lost. On the other hand, the mathematical formalisation facilitates computer-supported applications to large data sets and has developed into a lively mathematical field of research[1].

Recently, Semiotic-Conceptual Analysis (SCA) has been proposed as an extension of FCA (Priss 2017) which enhances the FCA vocabulary to include notions of "sign", "representamen", "denotation", "interpretation", "polysemy", "synonymy", "icon", "index" and "symbol". In the same manner in which the FCA notions are abstract versions of their original philosophical cousins, the SCA notions are similarly abstract and formal. Furthermore, although inspired by Peirce, some of the notions diverge from Peirce. This is because the focus of SCA is not on philosophy but on computational applications with FCA. Nevertheless, applying FCA (or any other form of computer-supported data analysis) always starts with human judgement with respect to how to deliminate the data and to decide what aspects will be represented or derived in what manner.

In this paper we are proposing to employ SCA explicitly for this initial aspect of data analysis. In particular, we have developed an SCA framework as a method of qualitative analysis (more precisely content analysis) for this initial step. Other researchers have already advocated developing qualitative analysis methods based on semiotics: Souza et al. (2010) show for their Semiotic Inspection Method (SIM) that such methods present a valid means for obtaining knowledge in the field of Human Computer Interaction. SIM and SCA are somewhat similar and use a similar semiotic vocabulary, but the notions in SCA are mathematically formalised and, furthermore, the process of application and focus of investigation is different for the two methods. Nevertheless we suggest that Souza et al.'s arguments also apply to SCA. Souza et al. state that because computer software is highly structural, it is also possible to employ structural analyses. These

[1] A query on Google Scholar for the string "formal concept analysis" retrieves 23000 results.

can be expected to not just answer ”what” questions but to some degree also ”why” questions.

Content analysis (CA) existed at least since the early 20th century “as a flexible method for analyzing text data” (Hsieh & Shannon 2005), for example for medical health studies or media analyses. It usually consists of deriving a list of “categories” and determining where these categories can be detected within a text. The categories are either derived from the texts (“inductive CA”) or, as “deductive CA”, compiled from prior knowledge and the intention of the study (Elo & Kyngas 2008). “Quantitative” CA evaluates the occurrence of the categories within a text statistically which is “sometimes referred to as quantitative analysis of qualitative data” (Hsieh & Shannon 2005). The validity and usefulness of CA is not undisputed but similar methods exist in other fields, for example discourse or conversational analysis in linguistics. In this paper we are proposing to use an SCA framework (containing the FCA and SCA notions mentioned above) as categories in the sense of a qualitative deductive CA in order to conduct an initial analysis of a set of data. The results from this analysis can then be utilised to determine the preprocessing of the data for further computerised analyses (for example with FCA) and to formulate relevant research questions.

2 SCA Core Definitions

Def. 1 presents the core definitions of SCA. Much further detail cannot be provided in this abstract and is presented by Priss (2017).

Definition 1. *For a set R (called* representamens*), a set D (called* denotations*) and a set I of partial functions $i : R \nrightarrow D$ (called* interpretations*) a* semiotic relation *S is a relation $S \subseteq I \times R \times D$. A triple $(i, r, d) \in S$ with $i(r) = d$ is called a* sign.
For a semiotic relation S with a tolerance relation $\sim_D$ on D, a tolerance relation $\sim_{D\cap R}$ on $D \cap R$, an equivalence relation $\approx_R$ on R and a partial function $f : R \nrightarrow D$:
- (i_1, r_1, d_1) *and* (i_2, r_2, d_2) *are* synonyms $\Leftrightarrow d_1 \sim_D d_2$*;*
- (i_1, r_1, d_1) *and* (i_2, r_2, d_2) *are* polysemous $\Leftrightarrow r_1 \approx_R r_2$ *and* $d_1 \sim_D d_2$*;*
- (i_1, r_1, d_1) *and* (i_2, r_2, d_2) *are* homographs $\Leftrightarrow r_1 \approx_R r_2$ *and* $d_1 \not\sim_D d_2$*;*
- $(i, r, d) \in S$ *is an* icon $\Leftrightarrow r \sim_{D\cap R} d$ *(i.e., describable by a unary relation)*
- $(i, r, d) \in S$ *is an* index $\Leftrightarrow f(r) = d$ *(i.e., describable by a binary relation)*
- $(i, r, d) \in S$ *is a* symbol $\Leftrightarrow (i, r, d)$ *is neither icon nor index.*

Examples for representamens are words, gestures, pictures, videos and so on. The notion “denotation” is employed instead of Peirce’s “object” because it points to the role of the meaning of a sign independently of what kind of meaning that is. In SCA, denotations are usually modelled using FCA concepts. Interpretations encode information about the context (space and time) of a sign and its sign users. Interpretations are different from Peirce’s “interpretants” in that a single interpretation can be applied to many different signs. Interpretations are from the viewpoint of an observer who prepares data for an analysis. Def. 1 covers structural aspects but also presents an abstraction. For example, Def. 1 defines a sign as a triple in which representamen and interpretation together uniquely determine a denotation. Whether or not it is actually interesting to consider a triple that fulfils Def. 1 as a sign is not a purely structural question but also depends on the purpose of an application and is decided by an observer who includes a triple in the relation. This is a human judgement which cannot be formally derived. In a sense an observer is also a sign user which could lead to an infinite chain of sign interpretation, but for the purpose of SCA the chain stops with the observer.

More human judgement is encoded in the additional relations and functions which have to be determined for each application using background knowledge. A tolerance relation is just a mathematical description of similarity. Def. 1 states that icons are unary because r or d alone determines the sign. Indices are binary because they do not require i. The function f (which presents for example an algorithm for determining d from r or a lookup table) encodes background knowledge that holds for many signs of an application domain. Thus a sign is an index if it is binary and a connection between d and r exists according to decisions made by someone who uses SCA for modelling.

3 The SCA Framework

SCA allows to describe and investigate partial orders on signs, translations between different semiotic relations and in general how structures amongst sign components or semiotic relations are maintained or modified (Priss, 2017). It is not the intention of SCA to re-invent the wheel. Details which are already provided by other formal theories (FCA, formal ontologies, linguistics, mathematical category theory etc) are not defined in SCA but can be incorporated from such theories if desired. As mentioned above, the focus of this paper is on using SCA for qualitative investigations. We would argue that some aspects of signs are more suited for quantitative analyses (for example, detecting polysemy and synonymy) whereas other aspects require qualitative judgements (for example, determining whether a sign is an icon or index).

The SCA framework consists of the categories “representamens”, “denotations”, and so on as mentioned in the first section. So far we have applied the SCA framework to a type of mathematical diagram and a computer interface. In the first application, the SCA framework highlights a conflict between iconicity and structure: users consider certain diagrams as iconic even though they are not a very good match from a structural viewpoint. In the second application, the analysis points to a high degree of underlying conceptual and structural complexity.

References

1. De Souza, C. S., Leitao, C. F., Prates, R. O., Bim, S. A., & da Silva, E. J. (2010). *Can inspection methods generate valid new knowledge in HCI? The case of semiotic inspection.* International Journal of Human-Computer Studies, 68, 1-2, p. 22-40.
2. Elo, S., & Kyngas, H. (2008). *The qualitative content analysis process.* Journal of advanced nursing, 62, 1, p. 107-115.
3. Ganter, B.; Wille, R. (1999). *Formal Concept Analysis. Mathematical Foundations.* Berlin-Heidelberg-New York: Springer.
4. Hsieh, H. F., & Shannon, S. E. (2005). *Three approaches to qualitative content analysis.* Qualitative health research, 15, 9, p. 1277-1288.
5. Priss, U. (2017). *Semiotic-Conceptual Analysis: A Proposal.* International Journal of General Systems, Vol. 46, 5, p. 569-585.