How to Gauge the Relevance of Codes in Qualitative Data Analysis? – A Technique Based on Information Retrieval

Alexander Keller[1]

[1] University of Passau, Chair of Information Systems (Information- and IT-Service-Management), Passau, Germany
alexander.keller@uni-passau.de

Abstract. Qualitative research has experienced broad acceptance in the IS discipline. Despite the merits for exploring new phenomena, qualitative methods are criticized for their subjectivity when it comes to interpretation. Therefore, research mostly emphasized the development of criteria and guidelines for good practice. I present an approach to counteract the issue of credibility and traceability in qualitative data analysis and expand the repertoire of approaches used in IS research. I draw on an existing approach from the information science discipline and adapt it to analyze coded qualitative data. The developed approach is designed to answer questions about the specific relevance of codes and aims to support the researcher in detecting hidden information in the coded material. For this reason, the paper contributes to the IS methodology with bringing new insights to current methods by enhancing them with an approach from another discipline.

Keywords: qualitative data analysis; information retrieval; relevance of codes; qualitative methods; information science.

1 Introduction

Qualitative research is an accepted methodology in the information system discipline and since a significant growth in 2005, one can see a stable amount of publications in information system (IS) journals like Management Information Systems Quarterly (MISQ), Journal of Management Information Systems (JMIS) and Journal of the Association for Information Systems (JAIS) [1]. Besides studies which use qualitative research methods to answer specific research questions (e.g. [2–4]), much has been written about recommendations and guidelines regarding how qualitative research is conducted within the IS discipline (e.g. [5–7]). Especially in the IS field qualitative research and the findings it generates often have been seen as a minor discipline accompanied with biases regarding the common quality standards [8]. Therefore qualitative research has a strong need to declare itself and reply critics with its own quality criteria (e.g. [9]).

In their MISQ guest editorial on qualitative research in IS Sarker et al. (2013) introduce the principle of transparency and state that "there is a need for clarity in the logic underlying data analysis" [10]. This paper follows their advice and adds a standardized

13th International Conference on Wirtschaftsinformatik,
February 12-15, 2017, St. Gallen, Switzerland

Keller, A. (2017): How to Gauge the Relevance of Codes in Qualitative Data Analysis? - A Technique Based on Information Retrieval, in Leimeister, J.M.; Brenner, W. (Hrsg.): Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017), St. Gallen, S. 1096-1110

technique to the repertoire of approaches that are used to analyze data in qualitative IS research. Therefore, I have applied an approach from the information science domain and adopted it for the use in qualitative research. With the technique results are generated that grant additional and hidden insights regarding the relevance of codes derived from textual data. This new technique is called *code appearance and relative frequency index* (CARFI). Besides this methodical contribution the technique especially contributes to the field of IS, where relevance aspects are of special interest when studying technical and behavioral issues regarding the role of information technology (IT) (see actual studies like [11, 12]).

The underlying research process started with the need for an appropriate approach to investigate the relevance of codes and is structured in three phases. In the first orientation phase, the qualitative research discipline as well as the IS discipline have been screened for techniques. Due to no appropriate result the scope has been widened to other disciplines. With finding the information retrieval approach the third phase started that aims to adopt the technique for the usage with qualitative data. The results of this phase are presented in this work.

The paper is structured as follows: First, it is explained which methods, data collection techniques and modes of analysis are used in qualitative research and how they collude in the IS discipline. Besides that, I am looking at the coding process and the different techniques used. This is supplemented with existing quality criteria and shortcomings in regard to the interpretation of qualitative data that motivates the development of new analysis procedures. Based on this, I propose a new technique that helps to identify the relevance of codes in textual data sources. I illustrate the technique with real data from interviews with experts. The paper concludes in a discussion of the presented approach and gives an outlook towards future research.

2 Qualitative Research in IS

In the IS domain qualitative research has a long history and aims to study managerial and organizational issues related to innovations in information and communication technologies [13]. Qualitative research methods have its origin in social science and were developed to understand cultural and sociological phenomena [13]. In contrast to quantitative research the focus of qualitative techniques lies on the observation of complex phenomena and situations which cannot be grasped with quantitative measures. Besides qualitative research being used as the only instrument for studying a research topic, it can be combined with quantitative techniques in the same inquiry in a mixed methods approach. Because of both methodologies having different strengths they can rest upon each other to generate richer results in combination compared to being used alone [14, 15].

2.1 Methods, Data Collection and Modes of Analysis

In qualitative research one distinguishes between various methods which specify the research design and strategy. In Table 1 we shortly present five approaches which are

widely used in the IS domain and give references which explain the methods from a theoretical perspective (for the information given in Table 1 also see [13] and [16]).

Table 1. Qualitative Research Methods

Method	Description	References
Action Research	The method is a design-orientated approach and aims to widen the stock of knowledge in the specific domain. The goal is the development of solutions in a collaborative manner. Therefore, several cycles consisting of analysis, action and evaluation steps are needed.	[17–19]
Case Study Research	With case studies, complex phenomena are investigated within their real-life context. The approach focuses on cases which represent instances of a particular phenomenon. Therefore, each regarded case can be seen as a research object, which explicates and illustrates the research subject.	[20–23]
Ethnography	The behavioral method aims to generate findings through participatory observation within the direct environment of the research subject. Therefore, a strong and intensive involvement of the researcher is required to study the phenomena in its social context.	[24–26]
Grounded Theory	Grounded theory is a behavioral method with the goal to systematically develop theory from the interaction between data gathering and analysis. Therefore, the approach consists of a process with several rounds of coding (open, axial, selective) to analyze data in which the theory is grounded.	[27, 28]
Qualitative Content Analysis	Qualitative content analysis is based on an iterative process, which is designed for coding data in relation to a category system. The codes, also called categories, of this system are developed in an inductive (from the information sources) and/or deductive (theory orientated) manner.	[29, 30]

Besides the presented methods of qualitative research, different techniques of data collection are distinguished to gather empirical material for the purpose of data analysis. Qualitative material covers mostly non-numeric information which is collected directly in form of interviews, observations and fieldwork or indirectly from secondary sources like published and unpublished documents [13]. This results in textual information sources like interview transcripts and case descriptions which serve as the foundation for further analysis.

The non-numeric nature of qualitative data results in different modes of analysis compared to quantitative methods. In qualitative research it is essential to understand and interpret the given information in regard to the underlying research questions to ensure that conclusions are drawn in the right context. Therefore, the textual data is processed and text passages are coded with a system of categories to create a foundation for analysis and interpretation [31]. Besides approaches which result from the methodology perspective (e. g. grounded theory), there exist additional modes of analysis that can also be seen as philosophies for interpreting qualitative information. Among others the hermeneutic idea is a way of understanding and interpreting textual information and aims to create an understanding for the "[…] underlying coherence or sense" [32]. Therefore, both understanding the whole text on the one side and interpreting the different parts of it on the other side is relevant for analyzing the given information [33]. The approach targets the understanding of the research subject in general and focusses on the relation between different objects on the basis of textual information.

The semiotic approach aims to investigate the implicit and explicit meaning of texts and is divided into three different categories - the content, conversations and discourse analysis [13]. With content analysis the researcher structures the textual information sources and identifies patterns in regard to the given context [34]. In conversation analysis the background of social interaction and behavior is uncovered with the researcher immersing in the particular situation [25]. Discourse analysis "[…] concerns itself with critically analyzing language […] in the context of social interactions" [35]. With this concept not only the provided content of a person, but also their way of talking is investigated.

Figure 1 shows a summary of methods, data collection techniques and modes of analysis used in IS based on Myers (1997)[13]. As the three different parts are coherent

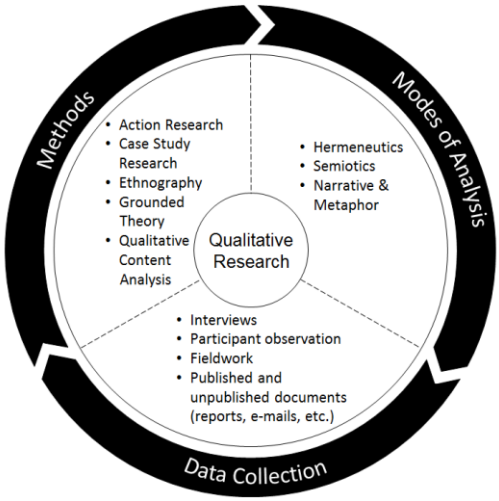

Figure 1. Qualitative Research

and mutually dependent on one another, qualitative research designs evolve from a combination of these aspects.

2.2 Codes and Coding Techniques for Qualitative Data Analysis

With the occurrence of grounded theory methodology the technique of coding emerged in the late 60s and is now said to be one of the most popular techniques of data analysis [36]. Referring to Sarker et al. (2012) about 60% of qualitative studies in the IS discipline use coding techniques to analyze empirical data [1]. Coding is a procedure to structure text and aims to mark segments of textual raw data that contain a specific information in regard to the underlying research question [37]. In the coding process the researcher identifies relevant information in the textual raw data (e.g. interview transcripts, reports, etc.) and attaches an existing or new code to a text segment [38]. A code represents the information given in the data and functions as an empirical generalization of relevant information in regard to the research question.

Looking at code-based qualitative data analysis Gläser and Laudel (2013) distinguishes between two important variations. First the difference in coding themes or coding content and second "[…] the extent to which preexisting theory is used in the coding process […]" [36].
Typically coding results in indexing themes, i.e. a set of codes that "[…] represents the structure of raw data in the text […]" [36]. The intention is to capture the information given in a text segment. In contrast coding content focusses on the content that should be expressed with a particular statement. In this case a code is used as a representation for a typical phenomenon of interest [36]. Besides that, codes can directly be generated from the textual raw data without considering existing theoretical aspects. This approach ensures that the derived codes are not distorted by prior theoretical considerations [39, 40]. In contrast some scholars object to this view by mentioning the importance of ex ante considerations [41, 42]. Miles and Huberman (1984) state that it is not possible to conduct any qualitative analysis without having a conceptual framework that emerges from considerations prior to the data analysis [43]. Therefore, coding techniques either can be used on textual data without any prior theoretical considerations (e. g. grounded theory [40]) or with a theory orientated conceptual framework as a starting point (e. g. deductive qualitative content analysis [30]).

Besides theoretical considerations the researcher has to consider specific guidelines while the data gathering an coding process to ensure a solid data basis for further analysis (see [31, 36, 44]). Negations or intonations are just two sources for biases that occur within the coding processes and can be handled with separate coding iterations.

Although coding can be done manually, software tools (e.g. *NVivo, Atlas.Ti, HyperResearch, MaxQDA*) are used to support the researcher in the coding process. Besides the main functionality of attaching codes to text segments the tools provide options for quantification and visualization of results. These different types of illustrations and representations are valuable sources for data interpretation.

2.3 Quality Criteria and Shortcomings in Qualitative Research

The essential and conventional research quality criteria validity and reliability also apply for qualitative research. Notwithstanding the criteria do have other characteristics when used in this domain [45].

Validity describes the quality and information value an approach is able to measure in regard to what it claims to measure. To exclude interfering influences quantitative methods are highly standardized. One can see this standardization in an understated form in qualitative research as well. Especially the approaches for data collection and processing are equipped with rules to ensure a standardized process (e.g. corpus construction [46], coding in grounded theory [27]). But as Flick (2014) mentions the strength of approaches in the qualitative domain often lies in their flexibility [47]. In addition, he indicates that communicational effects are not controllable in a reasonable way. Besides standardization, the credibility and accuracy in qualitative research can be ensured via communicative validation [45, 48]. This can be performed in form of direct member checks that require a validation of the involved people. For example, an interview partner validates the transcript to ensure its correctness. Additionally, the accuracy of the qualitative approach can be confirmed via peer validation with other researchers or experts in the field to ensure that the right methodology is used.
Reliability stands for the robustness of findings and the overall consistency of an approach which is repeated with consistent conditions. In qualitative research identical results not always represent reliable findings [49]. E.g. identical responses in interviews are not a decent indicator for reliability but point to prepared answers which result in falsified findings. For reliable results in qualitative research the context in which the data is collected and the analyses are performed must be described in very detail. This intersubjective tractability helps to understand each part of the study and should lead to other researches drawing consistent conclusions [50, 51].

Besides reliability and validity additional quality criteria are important that are addressed with the proposed approach. Lincoln and Guba introduce credibility as a construct that ensures the internal validity of qualitative research [45]. They emphasize the trustworthiness of findings and recommend approaches like triangulation and negative case analysis. In addition, the intersubjective traceability should be given to ensure that the results and interpretations are objective and confirmable [42]. Although there exist guidelines on how to meet this quality requirement, critics blame the domain for its strong subjectivity when it comes to interpretation of analyzed data [52]. Since the researcher is the one who is collecting, processing, analyzing and interpreting the data some degree of bias is introduced because of his subjectivity [53]. To control this bias, the research progress should be enhanced with techniques that enable a standardized evaluation and interpretation of qualitative data.

3 An Approach to Generate a Deeper Insight into the Relevance of Coded Qualitative Data

Regarding the lack of techniques for objective interpretation and the need to meet the quality criteria credibility and transparency an approach is provided to gather hidden information from coded qualitative data. The presented technique embraces the numerical structure of coded data which results from the assignment of text passages to codes. With the technique the relevance of codes (i.e. an empirical generalization of relevant information in regard to the research question) can be identified. Based on an information retrieval method an index is calculated to compare codes concerning their relevance with regard to the underlying research objective.

3.1 Relevance of Codes Based on Frequency and Appearance

In order to generate conclusions from the coded data it is helpful to compare the identified codes regarding their importance (e.g. [4]). Therefore, a value is necessary to rank each code in respect to its relevance. To calculate this index a standardized procedure is needed that identifies the importance based on the given coding of raw data. In information science the concept of information retrieval describes the process of identifying relevant information resources in collections based on given information needs [54, 55]. As a measure of relevance the two components term frequency and inverse document frequency are used [56, 57]. The first measure covers the frequency of a term by counting its occurrence within the information source. The second measure extends the term frequency by considering not a single information source but all existing sources. Therefore, the inverse document frequency counts the existence of the term in question over all considered information sources. As Robertson (2004) points out, multiplying the two constructs generates "[…] extraordinarily robust and difficult to beat […]" results [56]. Considering this, I adapted the measurement and developed an approach to evaluate the relevance of a code.

3.1.1 Concept

I modified the measurement to make it suitable in the context of qualitative data. This is possible because in the coding process each source of information (e.g. interview transcript) is screened multiple times to identify relevant information (see subsection 2.2). Hence a code can be seen as a term of question in the information retrieval domain. Transferring the mechanism, one can assume, that on the one hand a code is relevant the more often it is mentioned within a single information source. On the other hand, a code is important the more information sources contain the respective code. The following table lists the used mathematical notations and describes their meaning to measure the relevance of codes.

Table 2. Mathematical Notations and Descriptions

Notation	Description
n	Number of information sources
m	Number of codes
i	Index of information sources, $i \in \{1, \ldots, n\}$
c	Index of codes, $c \in \{1, \ldots, m\}$
f_c	Code-frequency of code *c* within all information sources, $f_c \in \mathbb{N}_0$
f_{ci}	Code-frequency of code *c* within information source *i*, $f_{ci} \in \mathbb{N}_0$
a_{ci}	Binary variable representing the appearance of code *c* in source *i*
$CARFI_c$	Index indicating the relevance of code *c*, $CARFI_c \in \mathbb{Q}^+$
AI_c	Additional information for the relevance of a code *c*, $AI_c \in \mathbb{Q}^+$
w	Wight for additive linkage of AI_c, $w \in [0,1]$
$ECARFI_c$	Extended index indicating the relevance of a code *c* based on $CARFI_c$ and AI_c, $ECARFI_c \in [0,1]$

Regarding the code-frequency (f_{ci}) each information source should have the same weighting (relative code-frequency). Therefore, the number of coded text segments for a particular code *c* in an information source *i* has to be standardized by the total number of codes mentioned in the information source *i*. With regard to Namey et al. (2008) the source-frequency is taken into account as a second measure [58]. It represents the number of information sources which mention the particular code *c* and is calculated with the binary variable a_{ci}:

$$a_{ci} = \begin{cases} 1, if\ code\ c\ is\ mentioned\ in\ source\ i \\ \\ 0, otherwise \end{cases}$$

The following formula represents the relevance of a code *c* containing the two factors relative code-frequency and source-frequency, where c' represents the total number of codes mentioned in information source *i*:

$$CARFI_c = \sum_{i=1}^{n} \frac{f_{ci}}{\sum_{c'=1}^{m} f_{c'i}} \sum_{i=1}^{n} a_{ci} \quad (1)$$

Regarding $CARFI_c$ it is assumed that codes with a high relative code-frequency combined with a high source-frequency matter more and hence have a higher relevance. This is expected because the more often a code appears in an information source relative to the other codes the more important the aspect is within this source. Because of this the relative code-frequency represents an indicator for the relevance of an aspect within an information source. But being coded more often than other codes within a source of information, does not implicate a high relevance. E. g. one must assume that a code which is only existing in one single source has less relevance than a code which is mentioned in all sources. Therefore, the source-frequency is additionally taken into consideration. It counts in how many different sources the code appears. Hence this

measure gives an insight into the topic related importance of a code regarding the context of information sources. With the multiplicative connection of the two measures codes are considered as relevant if both the source-frequency and the relative code-frequency result in high scores.

3.1.2 Extended Concept to Include Additional Information

In addition to the absolute and relative frequency other dimensions like intonations can be an indicator for relevance and hence should be considered in the approach. This additional information about the relevance of a code *c* AI_c can be included with an additive connection of constructs (see equation 2). The additive link is realized with a specific weight *w* to integrate the additional information in the relevance ranking.

$$ECARFI_c = (1 - w)\left(\frac{CARFI_c}{\sum_{c'=1}^{m} CARFI_{c'}}\right) + w\left(\frac{AI_c}{\sum_{c'=1}^{m} AI_{c'}}\right) \quad (2)$$

To ensure that the two measures in equation 2 can be combined, both terms of the sum need to be standardized (see denominators of the two terms). Additionally, the sum of weights must equal 1 which leads to the standardized relevance of a code (first term) being weighted with $1 - w$ where w stands for the weight with which the additional information is considered. $ECARFI_c$ itself results in a number between 0 and 1 due to the standardization of the related terms.

The same linkage can also be done when information about validation or rankings of codes is available from other information sources (e.g. expert interviews to validate or rank codes). If more than one additional information should be included this can be done with an adaption of weights, where each additional information is linked additive with its own weight where the sum of weights for each term equals one.

If no additional information AI_c is available, $CARFI_c$ as well as $ECARFI_c$ can be used to indicate the relevance of codes. While $CARFI_c$ results in an absolute number $ECARFI_c$ calculates to a relative number, because no additional information AI_c is available ($AI_c = 0, w = 0$). If no term of the sum is emphasized the weights for each term of the sum equal 1 divided by the total number of terms. For example, if the two terms in equation 2 should be considered equally, then $w = 1/2$.

3.1.3 Example with Textual Data from Interviews

The technique has been tested with real data from interviews with entrepreneurship experts to identify success indicators for IT-startups [59]. Eleven interview transcripts were coded on the bases of the methodology proposed from Steigleder (2008). The theory orientated content coding technique results in 22 codes with each of them representing a separate success indicator. The data set in table 3 represents the results of the coding process of textual data from eleven interviews which results in 22 codes. Each code stands for a success indicator for IT-startups.

Table 3. Data Set of Coded Qualitative Data from Interviews

		code c																				
	i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
interview i	1	2	0	4	2	2	1	3	2	2	0	2	1	1	0	0	1	0	0	0	0	0	1
	2	3	0	1	1	0	2	0	1	2	0	3	1	0	0	0	5	7	1	4	1	2	0
	3	5	0	0	1	3	0	2	0	0	1	0	1	1	7	0	0	0	0	0	3	1	0
	4	2	1	0	2	3	2	1	1	1	0	0	2	7	0	3	0	1	3	3	1	0	2
	5	0	2	2	2	1	1	0	1	0	1	2	2	0	0	1	5	4	5	3	3	3	7
	6	2	2	2	2	0	1	0	0	2	4	0	0	1	0	4	1	1	0	0	3	0	0
	7	1	1	0	2	1	1	0	0	3	0	0	0	1	0	2	1	4	0	1	0	0	1
	8	2	1	0	2	3	4	0	0	1	1	2	2	6	0	2	2	5	2	2	2	5	1
	9	3	0	0	0	1	2	1	1	3	0	1	2	0	1	0	1	1	3	3	2	3	1
	10	0	0	0	0	0	0	0	0	0	2	1	3	2	2	0	0	0	0	0	2	1	2
	11	1	0	0	2	0	0	2	3	0	0	2	0	0	0	1	1	0	1	0	1	1	1

The values given in the matrix represent the code-frequency of a particular code in an information source (f_{ci}). E. g. code 1 is coded two times in interview 1 which results in f_{11} equals two. The sum of one column results in f_c , e. g. code 1 is coded 21 times in total.

After the interview the experts have been directly asked about their top five factors for startup success. This data has been taken into account as additional information AI_c regarding the relevance for a code. In the example AI_c represents how often a success indicator, i.e. a code c, has been given as answer to this question. E. g. code 20 has six times been mentioned as top five indicator, see table 4.

By calculating the values of $ECARFI_c$ the two aspect ($CARFI_c$ and AI_c) both should be considered equally, i.e. both indicate success in the same manner. This results in $w = 0{,}5$. After applying the technique, the codes can be arranged in an ordered sequence which represents the relevance ranking for the analyzed codes and serves as a foundation for interpretation. The results are represented in table 4 and support the researcher in identifying the relevance and importance of success indicators regarding the underlying research question. Although *handling of market conditions* (rank 11, $ECARFI_{17}$ = 0,042) is the most commonly coded indicator (f_{17} = 23) it is ranked in the middle of the list. This results from a low relative code-frequency and a low value regarding the additional information ($AI_{17} = 1$). Besides that, *perseverance* (rank 8, $ECARFI_2$ = 0,054) with the lowest code-frequency (f_2 = 7) is ranked three spots above *handling of market conditions* because the additional information about the relevance (AI_2 = 5) is considered and indicates an important aspect for startup-success. This shows that experts do not accentuate the importance of being able to handle complex market conditions as much as they emphasize *perseverance* in the interviews.

Table 4. Ranking of Codes Based on *ECARFI*

Rank	Sucess indicator	c	f_c	$\sum_{i=1}^{n} a_{ci}$	$CARFI_c$	AI_c	$ECARFI_c$
1	Customer orientated problem solving	20	18	9	6,07	6	0,095
2	Sales competence and marketing power	1	21	9	6,95	4	0,082
3	Seed-customer and technology partner	16	17	8	4,27	5	0,074
4	Feedback driven product developement	21	16	7	3,56	5	0,070
5	Prototype orientated product developement	22	16	8	4,48	4	0,066
6	Team composition	12	14	8	4,21	4	0,064
7	Industry specific competence	5	14	7	3,26	4	0,058
8	Perseverance	2	7	5	1,14	5	0,054
9	Scalability and market ability	9	14	7	3,74	3	0,052
10	Staff management skills	8	9	6	2,31	3	0,043
11	Handling of market conditions	17	23	7	5,04	1	0,042
12	Industry specific financing	13	19	7	4,49	1	0,039
13	Entrepreneurship and professional experience	4	16	9	5,48	0	0,036
14	Political and regulartory business environment	18	15	6	2,62	2	0,036
15	ICT competence	6	14	8	3,44	1	0,032
16	Conversion capability and speed	7	9	5	1,97	2	0,031
17	Proof of feasability and verification	10	9	5	1,89	2	0,031
18	Industry specifc norms and requirements	19	16	6	2,82	1	0,028
19	Business model felxibility and independence	11	13	7	3,41	0	0,022
20	R&D cooperations	15	13	6	2,88	0	0,019
21	Value orientated thinking	3	9	4	1,28	1	0,018
22	Accelerator or incubator program	14	10	3	1,34	0	0,009

Besides that, one can see that codes with the same source-frequency can be brought to an order. E. g. the *sales competence and marketing power* ($ECARFI_1$ = 0,082) as well as the *entrepreneurship and professional experience* ($ECARFI_4$ = 0,036) were both mentioned from 9 different experts but are ranked on second and 13th place with a difference of 0.036. This results from the fact that experts emphasize the sales skills of the startup team within their interviews and mention it as top 4 success factor. Besides identifying the indicators of IT-startup success, the ranking makes it possible to give recommendations about the relative importance of each aspect which cannot be revealed from a list of simple code frequencies.

4 Limitations and Future Research

Although the approach generates stable results it is subject to some limitations. First, the quality of the results generated can only be as good as the quality of the underlying coding process. Therefore, using the techniques does neither understate the need for a solid data collection nor does it replace a robust coding of the qualitative data. It has to be seen as an additional measure which helps to detect hidden information in coded datasets. Second, although the approach generates neutral and objective results, the researcher still interprets them from a subjective perspective that can result in biases.

Third, it is important to understand that the results only provide an indication for relevance but cannot generate findings that are valid in general. This is due to the structure of qualitative research which aims to understand complex phenomena regarding a specific topic of interest. Fourth, frequency is just one approach to address relevance in an objective manner but does not fit for every qualitative investigation. To grant stable findings it is important to combine the proposed technique with other approaches to judge relevance (e.g. context a code comes from or is related to, verbal traces and actual speech such as laughter). In addition to that supplementary mechanisms can be applied to make the input more robust (e. g. case contrasting proposed by Flick (2014) diminishes the impact of similar sources). Fifth, the approach views each source as equal and includes each code mentioned in a source based on the total amount of codes within this source. Therefore, codes mentioned in sources with a lot of codes are less emphasized then codes mentioned in sources with few codes. This behavior must be considered when using the technique.

Future research should investigate the fit of the presented approaches for qualitative research in more detail. To check the robustness of the ranking, one could use a linear ordering technique which examines the ranking based on additional data sources indicating the ranking. Furthermore, intentionally a small data set was applied to introduce the approach in a simple and understanding manner. Although the approach has been used with large data samples in its original discipline, the next step is the investigation of the behavior with different dimensions of data size in qualitative terms.

5 Conclusion

This paper introduces a new qualitative research approach which expands the existing repertoire of qualitative data analysis methods. Therefore, I draw on an existing approach from the information science discipline to bring new insights to current qualitative methods in IS. The technique is adopted and enhanced to be used on data which results from any kind of qualitative coding process. Hence, the approach can be performed independently of the underlying methodology and does not intend to replace existing procedures but aims to extend them.
By developing the practice, I focused on answering questions about the relevance of codes. The proposed technique results in an index representing the relevance or importance of a code relative to other codes. The generated results can serve as an additional basis for analysis and interpretation of qualitative data sources. Until now analysis was only based on the system of codes and their assigned textual passages. With the additional findings from my approach a deeper insight can be generated that supports the researcher in detecting hidden information in the coded data. Regarding the intersubjective tractability and credibility the standardized measure enables replicable and transparent findings generated from qualitative data. The formalization of the approach makes it obvious how the results revealed. With this, one can counteract the quality issues tractability and credibility on the one hand and fulfill the principle of transparency on the other.

The technique addresses and complements many existing and future qualitative and mixed method studies based on coding techniques in the IS domain. Especially when it comes to large empirical data samples in terms of big data the full potential of the approach can be exploited. As the complexity increases with a growing number of sources the mechanism can support analyzing and interpreting "big qualitative data".

References

1. Sarker, S., Xiao, X., Tanya, B.: Towards an Anatomy of "Successful" Qualitiative Research Manuscripts in IS: A Critical Review and Some Recommendations. Thirty Third Int. Conf. Inf. Syst. (ICIS 2012). 1–21 (2012).
2. Chua, C.E.H., Yeow, A.Y.K.: Artifacts, actors, and interactions in the cross-project coordination practices of open-source communities. J. Assoc. Inf. Syst. 11, 838–867 (2010).
3. Davison, R.M., Vogel, D.R.: Group Support Systems in Hong Kong: An Action Research Project. Inf. Syst. J. 10, 3–20 (2000).
4. Vogelsang, K., Steinhüser, M., Hoppe, U.: A Qualitative Approach to Examine Technology Acceptance. Int. Conf. Inf. Syst. (ICIS 2013). 234–245 (2013).
5. Urquhart, C., Lehmann, H., Myers, M.D.: Putting the "theory" back into grounded theory: guidelines for grounded theory studies in information systems. Inf. Syst. J. 20, 357–381 (2010).
6. Klein, H.K., Myers, M.D.: A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems. MIS Q. 23, 67–93 (1999).
7. Schultze, U.: A Confessional Account of an Ethnography about Knowledge Work. MIS Q. 24, 3–41 (2000).
8. Conboy, K., Fitzgerald, G., Mathiassen, L.: Qualitative Methods Research in Information Systems: Motivations, Themes, and Contributions. Eur. J. Inf. Syst. 21, 113–118 (2012).
9. Marton, A.: Purposive Selection and the Quality of Qualitative IS Research. Thirty Fourth Int. Conf. Inf. Syst. 1–19 (2013).
10. Sarker, S., Xiao, X., Beaulieu, T.: Guest Editorial: Qualitative Studies in Information Systems: A Critical Review and Some Guiding Principles. Manag. Inf. Syst. Q. 37, (2013).
11. Asatiani, A.: Why Cloud? - A Review of Cloud Adoption Determinants in Organizations. Twenty Third Eur. Conf. Inf. Syst. (ECIS 2015). 1–17 (2015).
12. Lang, M., Wiesche, M., Krcmar, H.: What are the most important Criteria for Cloud Service Provider Selection? A Delphi Study. Twenty Fourth Eur. Conf. Inf. Syst. (ECIS 2016). 1–17 (2016).
13. Myers, M.D.: Interpretive Research Methods in Information Systems. In: Mingers, J. and Stowell, F. (eds.) Information Systems: An Emerging Discipline. pp. 239–266. McGraw Hill, London (1997).
14. Venkatesh, V., Brown, S.A., Bala, H.: Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems. Manag. Inf. Syst. Q. 37, 21–54 (2013).
15. Johnson, R.B., Turner, L.A.: Data Collection Strategies in Mixed Methods Research. In: Tashakkori, A. and Teddlie, C. (eds.) Handbook of Mixed

Methods in Social & Behavioral Research. pp. 297–319. Sage Publications, Thousand Oaks, CA (2003).
16. Wilde, T., Hess, T.: Forschungsmethoden der Wirtschaftsinformatik Eine empirische Untersuchung. Wirtschaftsinformatik. 49, 280–287 (2007).
17. Susman, G.I., Evered, R.D.: An Assessment of the Scientific Merits of Action Research. Adm. Sci. Q. 23, 582–603 (1978).
18. Baskerville, R.L., Wood-Harper, A.T.: Diversity in information systems action research methods. Eur. J. Inf. Syst. 7, 90–107 (1998).
19. Ytterstad, P., Akselsen, S., Svendsen, G., Watson, R.T.: Teledemocracy: Using Information Technology to Enhance Political Work. MISQ Discov. 1, (1996).
20. Yin, R.K.: The Case Study as a Serious Research Strategy. Knowledge. 3, 97–114 (1981).
21. Yin, R.K.: Case Study Research: Design and Methods. Sage Publications, Newbury Park (2003).
22. Thomas, G.: A Typology for the Case Study in Social Science Following a Review of Definition, Discourse, and Structure. Qual. Inq. 17, 511–521 (2011).
23. Alavi, M., Carlson, P.: A review of MIS research and disciplinary development. J. Manag. Inf. Syst. 8, 45–62 (1992).
24. Lewis, I.M.: Social Anthropology in Perspective. Cambridge University Press, Cambridge (1985).
25. Wynn, E.: Office conversation as an Information Medium, (1979).
26. Baskerville, R.L., Myers, M.D.: Design ethnography in information systems. Inf. Syst. J. 25, 23–46 (2015).
27. Strauss, A., Corbin, J.: Basics of Qualitative Research: Grounded Theory Procedures and Techniques. Sage Publications, Newbury Park, CA (1990).
28. Martin, P.Y., Turner, B.A.: Grounded Theory and Organizational Research. J. Appl. Behav. Sci. 22, 141–157 (1986).
29. Mayring, P.: Qualitative Inhaltsanalyse: Grundlagen und Techniken. Beltz, Weinheim (2015).
30. Steigleder, S.: Die strukturierende qualitative Inhaltsanalyse im Praxistest: eine konstruktiv kritische Studie zur Auswertungsmethodik von Philipp Mayring, (2008).
31. Saldaña, J.: The coding manual for qualitative researchers. Sage Publications, Los Angeles (2016).
32. Taylor, C.: Hermeneutics and Politics. In: Connerton, P. (ed.) Critical Sociology: Selected Readings. pp. 153–193. Penguin Books, Harmondsworth (1976).
33. Gadamer, H.-G.: Philosophical Hermeneutics. University of California Press, California (1976).
34. Krippendorff, K.: Content analysis : An introduction to its methodology. Sage Publications, Beverly Hills (1980).
35. Howcroft, D., Trauth, E.M.: Handbook of Critical Information Systems Research: Theory and Application, (2005).
36. Gläser, J., Laudel, G.: Life with and without coding: Two methods for early-stage data analysis in qualitative research aiming at causal explanations. Forum Qual. Sozialforsch. 14, (2013).
37. Kelle, U.: Theory Building in Qualitative Research and Computer Programs for

the Management of Textual Data. Sociol. Res. 2, (1997).
38. Coffey, A., Atkinson, P.: Making sense of qualitative data: Complementary research strategies. (1996).
39. Meinefeld, W.: Ex-ante-Hypothesen in der Qualitativen Sozialforschung: zwischen "fehl am Platz" und "unverzichtbar". Z. Soziol. 26, 22–34 (1997).
40. Glaser, B., Strauss, A.: The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine Publishing Company, Chicago (1967).
41. Auer-Srnka, K.J.: Hypothesen und Vorwissen in der qualitativen Marktforschung. In: Qualitative Marktforschung. pp. 159–172. Springer (2009).
42. Bogner, A., Littig, B., Menz, W.: Interviews mit Experten: Eine praxisorientierte Einführung. Springer Fachmedien, Wiesbaden (2014).
43. Miles, M.B., Huberman, A.M.: Qualitative Data Analysis: A Sourcebook of New Methods. Sage Publications, Newbury Park, CA (1984).
44. Campbell, J.L., Quincy, C., Osserman, J., Pedersen, O.K.: Coding In-depth Semistructured Interviews: Problems of Unitization and Intercoder Reliability and Agreement. Sociol. Methods Res. 42, 294–320 (2013).
45. Lincoln, Y.Y., Guba, E.G.: Naturalistic Inquiry. Sage, Beverly Hills (1985).
46. Bauer, M.W., Gaskell, G.: Qualitative researching with text, image and sound: A practical handbook for social research. Sage (2000).
47. Flick, U.: Gütekriterien qualitativer Sozialforschung. In: Baur, N. and Blasius, J. (eds.) Handbuch Methoden der empirischen Sozialforschung. pp. 411–423. Springer, Wiesbaden (2014).
48. Bryman, A.: Research Methods and Organization Studies. Unwin Hyman, London (1989).
49. Flick, U.: An introduction to qualitative research. Sage, Los Angeles (2014).
50. Kruse, J.: Qualitative Interviewforschung : Ein integrativer Ansatz, (2015).
51. Steinke, I.: Kriterien qualitativer Forschung. Juventa Verlag, Weinheim München (1999).
52. Wrona, T., Gunnesch, M.: The One Who Sees More is More Right: How Theory Enhances the "Repertoire to Interpret" in Qualitative Case Study Research. J. Bus. Econ. 1–27 (2015).
53. Zahedi, F.M., Van Pelt, W.: Web Documents ' Cultural and Femininity Masculinity. J. Manag. Inf. Syst. 23, 87–128 (2006).
54. Belkin, N.J., Croft, W.B.: Information Filtering and Information Retrieval: Two Sides of the Same Coin? Commun. ACM. 35, 29–38 (1992).
55. Salton, G., McGill, M.J.: Introduction to Modern Information Retrieval. (1986).
56. Robertson, S.: Understanding inverse document frequency: on theoretical arguments for IDF. J. Doc. 60, 503–520 (2004).
57. Wu, H.C., Luk, R.W.P., Wong, K.F., Kwok, K.L.: Interpreting TF-IDF term weights as making relevance decisions. ACM Trans. Inf. Syst. 26, 1–37 (2008).
58. Namey, E., Guest, G., Thairu, L., Johnson, L.: Data Reduction Techniques for Large Qualitative Data Sets. Handb. Team-based Qual. Res. 137–163 (2008).
59. Keller, A.: Zum Erfolg von IKT-Start-ups in der deutschen Elektrizitätswirtschaft - Eine explorative Studie auf Basis von Experteninterviews. Verlag Dr. Kovač, Hamburg (2016).