

Taxonomy Development for Business Research: A Hands-On Guideline



Dennis M. Steininger, Manuel Trenz, and Daniel J. Veit

Abstract Classification schemes are important groundwork for research on many topics of different business disciplines such as information systems (IS). They make investigating topics manageable by allowing researchers to delimit their work to certain taxa or types (e.g., of artifacts or firms) and provide a basis for generalization. Opposed to theoretically grounded typologies, taxonomies are empirically derived from entities of a phenomenon and therefore have several advantages, such as more detailed and exhaustive coverage. Initial guidelines for developing taxonomies in business have been proposed; however, research is still missing a clear set of applicable procedures to empirically build taxonomies. We tackle this topic by suggesting an inductive approach based on the procedures of content and cluster analysis. Each of the proposed six steps is amended with comprehensive state-of-the-art guidelines, suggestions, and formative measures of reliability and validity.

1 Introduction

Originally stemming from biology to differentiate animals or plants, classification schemes allow the systematic ordering or sorting of phenomena into similar groups or classes (e.g., business models, Veit et al. 2014). They are of fundamental importance for science and academic research in business (Kantor 1953; Kemeny 1959; Nickerson et al. 2013). Wolf (1926) emphasizes this importance by explaining the links and stating that verification of laws of science may only succeed after classification has been completed since it is “the first and last method employed by science.” Hence, classification schemes such as taxonomies make investigating

D. M. Steininger · D. J. Veit (✉)
University of Augsburg, Augsburg, Germany
e-mail: dennis.steininger@wiwi.uni-augsburg.de; veit@wiwi.uni-augsburg.de

M. Trenz
University of Göttingen, Göttingen, Germany
e-mail: trenz@uni-goettingen.de

phenomena manageable by allowing researchers to delimit their work to certain classes (i.e., taxa or types), such as IS technologies or firms, and also provide a basis for generalization. This allows the building of theories that apply to certain classes of developed schemes. When classifying an area of investigation, two different approaches can be used: typologies or taxonomies. Typologies are created deductively by classifying objects into predefined groups that are created based on intuition or previously existing knowledge and theory (Bailey 1994). Especially when examining an unexplored area of research, as often done with new technologies, there is a risk of researcher bias or general misconception, since existing theory is limited. Unlike theoretically grounded typologies, taxonomies are derived inductively from empirical data (i.e., entities of a phenomenon under investigation) and therefore have several advantages, such as more detailed and exhaustive coverage and mutual exclusiveness of classes. Despite some foundational work (e.g., Nickerson et al. 2013; Oberländer et al. 2019), business research is still missing a clear set of rigorous procedures on how to empirically build taxonomies of firms, artifacts, systems, user behavior, or processes. Especially in fast-moving areas such as IS, it is important to be able to describe new phenomena rigorously and quickly by applying systematic actions. Building on these thoughts, we propose the following research question for this work:

How can taxonomies be developed in business research from empirical entities using content analysis?

We tackle this question by suggesting an inductive empirical approach based on the procedures of content and cluster analysis. Content analysis allows a systematic and rigorous analysis of entities to get a first grasp of their characteristics, associated manifestations and densities. Based on these results, procedures of cluster analysis can be applied to derive final classes. The remainder of this chapter is structured as follows. In the second section we propose six steps to build taxonomies. Each of these steps is amended with state-of-the-art guidelines, alternatives, and measures of reliability and validity. Summative measures of taxonomic quality are also depicted for evaluating final taxonomic constructs. In the last section we sum up our findings, address the usefulness of taxonomic outcomes, and identify interesting topics in IS that might be investigated by using the introduced method.

2 The Process of Taxonomy Building

We introduce detailed steps and procedures to build taxonomies in IS and management-related phenomena using content and cluster analysis. The process is based on Steininger et al. (2011b), who use clustering and content analysis to inductively build a taxonomic framework of Web 2.0 characteristics. This chapter can be seen as a working example. We added inspirations from the articles of Nag et al. (2007), defining strategic management via content analysis and clustering; Al-Debei and Avison (2010), developing a business model framework through content

analysis; and the seminal work of Nickerson et al. (2013). Content analysis is a technique to gain “replicable and valid inferences from text” (Krippendorff 2004a, p. 18) and thereby find trends, characteristics, patterns, and densities. The material for analysis might include written or spoken texts as transcripts from various sources (for a list of potential sources see: Steininger 2019). The objectivity, validity, and reliability of the outcomes are obtained through rigorous rules and systematic procedures, which have been refined and adapted to the various needs of different disciplines over time (Angelmar and Stern 1978; Abbasi and Chen 2008; Steininger et al. 2011a, b) and distinguish content analysis from regular critical reading. The aforementioned potential to reliably and systematically uncover characteristics and patterns is of high relevance for constructing taxonomies. Hence, we adapt state-of-the-art procedures of inductive and deductive content analysis for major parts of the taxonomy-building process. The outline of our idea is to define a phenomenon of investigation and collect examples resembling the phenomenon as entities of investigation. We then inductively develop the characteristics of the phenomenon from these entities and deductively measure the manifestation of the characteristics for each entity.

We finally propose to cluster the entities into classes (i.e., taxa) by analyzing their manifestations and densities of characteristics. The entire process is depicted in Fig. 1, highlighted for one entity (marked with black ink). It starts with a definition of the phenomenon under investigation (e.g., electronic business models). This entails a clear statement of the research question (e.g., What classes of electronic business models do exist?). After these initial specifications, a set or population of entities and their textual descriptions resembling the phenomenon (e.g., examples of existing electronic business models) is required as a basis for analysis, which is addressed in our first suggested step on the selection and sampling of entities.

To proceed with building the taxonomy, it is necessary to analyze the manifestation of the phenomenon’s characteristics for each entity. Since we assume missing theoretical foundations on the characteristics of the phenomenon, we describe procedures on how to inductively derive raw characteristics from selected entities by using content analysis (step 2). Raw characteristics are subsequently reduced to main characteristics of the phenomenon under investigation (e.g., characteristics of electronic business models) by applying cluster analysis (step 3). These two steps might be skipped if our assumption does not hold true and there are already existing and exhaustive definitions of characteristics for the phenomenon in theory, which can be utilized for the fourth step. In this fourth step we suggest deductive content analysis procedures to measure the manifestations and densities of the characteristics for each entity (e.g., how often is a characteristic mentioned in the textual material for one entity). This can be reached through analyzing the entities by applying a coding scheme of characteristics, which might be constructed from the inductively developed (cf. steps 2/3) or aforementioned theoretically derived characteristics. The classes of similar entities for the taxonomy (e.g., virtual shopping malls) are then built by suggested procedures of cluster analysis on the resulting manifestations (step 5). We amend this penultimate step with propositions and guidelines on measures for taxonomic quality (e.g., mutual exclusiveness).

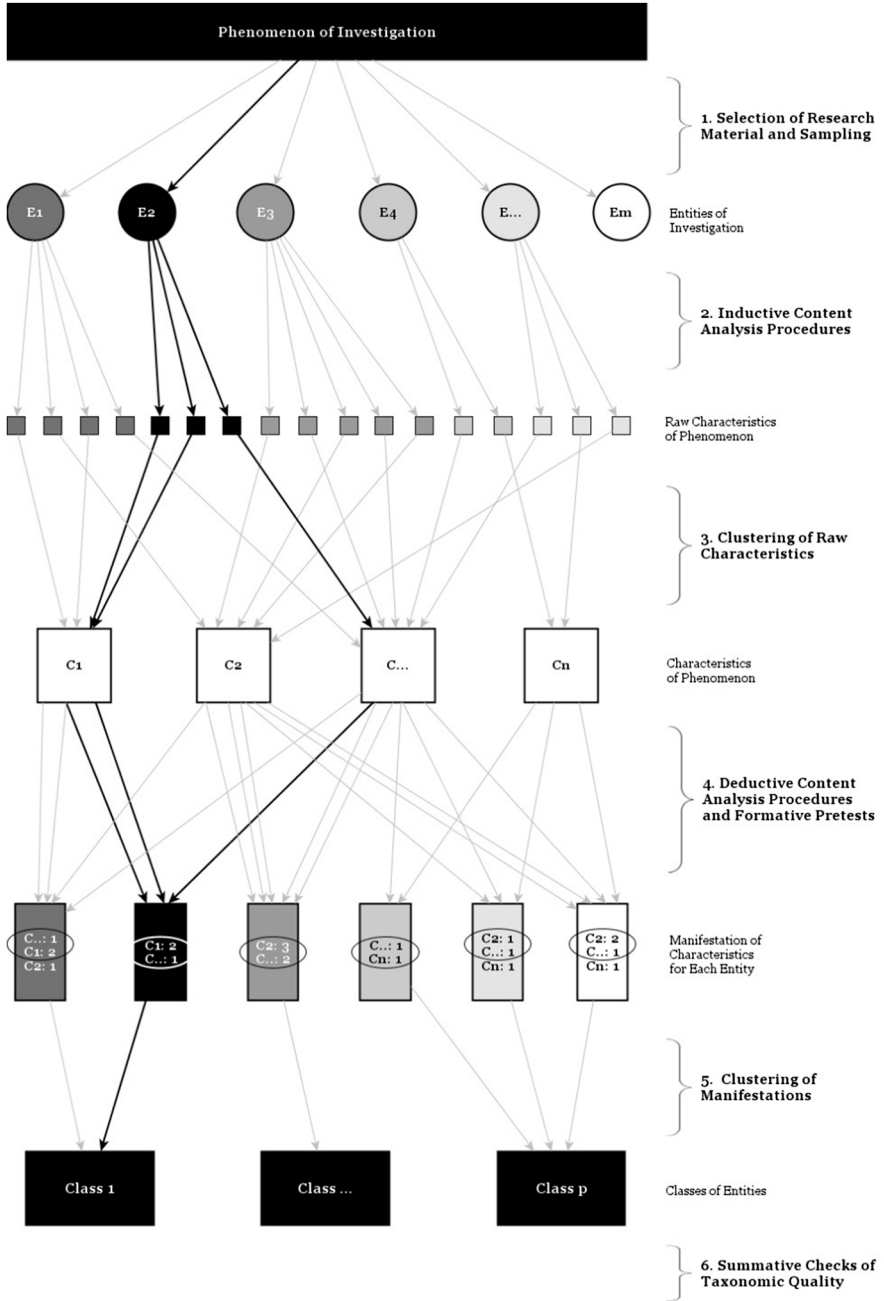


Fig. 1 Overview of the taxonomy-building process

Details and guidelines for each of our suggested steps are introduced in the following sections.

2.1 Selection of Research Material and Sampling

Entities of investigation (e.g., firms using an electronic business model) are needed as empirical research material to develop and retrieve characteristics, manifestations, and final classes (i.e., taxa) for a phenomenon. We explain procedures for selecting and sampling these entities throughout this section and amend them with hints on data sources and data collection techniques to gain rich data on the selected entities.

A representative sampling of entities might be used but, in many cases, may be neither manageable nor required. Instead, we propose to follow a theoretical sampling approach as suggested by Eisenhardt (1989). This means broadly choosing the entities of investigation for variation, heterogeneity (i.e., unique cases), or replication instead of random selection (Yin 2009). The availability of existing textual (e.g., case descriptions, annual or mission statements, product descriptions, websites, directories) or transcribable (e.g., interviews) descriptions for the entities might also be taken into account as a factor of selection during this sampling process. We suggest collecting descriptive data of the entities by following the sources of evidence given in Table 1.

It is recommended to use similar sources of evidence for all entities. Triangulation of more than one source might enrich the descriptions and lead to more robust results (cf. Yin 2009). We suggest listing derived entities in a longlist (*LL*). If entities are gained from different sources, this list should be cleaned of possible duplicates. The introduction of a selection factor (*SF*) can help to prepare the *LL* for further proceedings (Steininger et al. 2011b). This selection factor might encompass extra credit points for criteria such as an entity being a unique or extreme case, certain keywords within the name of an entity for restriction to a specific area of interest, or the availability of evidence for an entity. In a final step the *LL* has to be sorted in descending order by *SF*. Entities at the lower end of the list not reaching a certain selection factor might now be truncated, which results in shortlist (*SL*). Different approaches to gaining this shortlist might also be applied (i.e., taking a sample of entities from an existing journal paper on the phenomenon). The *SL* should be amended with an ascending research material ID (*i*) for each entity in a finalizing step.

2.2 Inductive Content Analysis Procedures

In this second step of our approach, we present a set of procedures and guidelines on how to inductively develop raw characteristics from textual descriptions of the selected entities from the preceding section.

Table 1 Data sources for entity description (Myers 2008; Yin 2009)

Name	Application	Advantages	Disadvantages
Documentation and Archival Records	Usually available in written form	Stable Unobtrusive	Bias of author unknown Retrievability and access
Web, Databases, and Social Media	Usually available in written and retrievable form	Relatively stable	Bias of author unknown Usage restrictions
Interviews	Transcription by person independent from interviewer. Final approval of transcript by interviewee	Targeted Insightful	Poor question bias Response bias Reflexivity
Fieldwork	Written memos of direct or participant observation Final check of memos by participants	Real-time coverage Contextual Insightful into behavior and motives	Time-consuming Observer bias Reflexivity
Physical Artifacts	Use of existing descriptions or composition of descriptive memos by two independent authors	Insightful into cultural features and technical operations	Selectivity Availability Access

After specification of the entities and their sampling as research material, the unit of analysis needs to be subsequently defined. This addresses the issue of “the basic unit of text to be classified” (Insch et al. 1997, p. 10) such as paragraphs or words into the categories of characteristics derived in succeeding steps. The configuration of this unit has a considerable impact on the quality and reliability of research results. Choosing a smaller unit (e.g., word) usually leads to higher reliability and possible automation but might corrode results which focus on larger meanings than transported by single words (Saris-Gallhofer et al. 1978). Following Kassarian (1977), the “theme” is usually suggested for this type of taxonomic method, ensuring the capture of word- or sentence-spanning ideas especially within the inductive phase of building raw characteristics. To stabilize the results and reliabilities, entire sentences should be used as the operationalized coding unit, which leads to solely coding a category once within one sentence (Steininger et al. 2011b). In the suggested approach, the raw characteristics should be developed inductively from the selected research material (i.e., entities of investigation). This is done to initially capture the characteristics of the phenomenon of investigation, which are needed as groundwork for further analysis.

Based on raw characteristic-building rules (Mayring 2002), the research material should be worked through consecutively, and raw characteristics are defined beginning with the first selected entity of investigation. Each occurrence of a new or additional raw characteristic-building incident should be marked and uniquely

Table 2 Units of analysis (adapted from Kassirjian 1977)

Unit	Description	Advantages	Disadvantages
Word	Analysis of single words such as key symbols or value-laden terms	Ease of coding Ease of automation Highest reliabilities	Loss of context Loss of word-spanning ideas
Sentence	Analysis of entire sentences	Relative ease of coding Clear demarcation of unit borders	Loss of sentence-spanning ideas
Theme	Analysis of single assertions about a subject	Capturing of entire subjects of investigation Very useful in most content analyses	Ambiguous unit borders Difficult coding Lower reliabilities
Item	Entire documents such as speeches, letters, manuals	Useful in classifying entire documents	Often too gross for most research
Character	Mostly used in the analysis of streaming media or commercials to analyze heroes, bad guys, etc.	Useful in the analysis of behavior or communication of actors Might be of interest to develop taxonomies of user behavior in IS	Sometimes ambiguous unit borders Context might not be captured
Space and time	Analysis by column (e.g., newspaper), line, paragraph, or minute	Useful for historical timeline analysis and longitudinal taxonomies Clear demarcation of unit borders	Loss of unit-spanning ideas and context

numbered using the research material ID i (cf. Sect. 2.1). If the marked and colored occurrence in the text defines a nonexistent characteristic, a new and unique raw characteristic ID (r) is hyphenated (e.g., $i.1-r$). If the occurrence matches an existing raw characteristic and only adds richness to the description of the characteristic, that existing characteristic number should be used instead, and the mark is suggested to be set in a different color (e.g., dark blue). All raw characteristics categories are summoned in a list (RcL). This process should be continued until saturation is reached (i.e., no new raw characteristics can be derived from the research material entities) (Mayring 2008).

2.3 Clustering of Raw Characteristics

In this section we develop a set of guidelines on how to reduce and cluster the raw characteristics developed through the procedures outlined above. The goal of this step is to reach a generalizable and manageable set of main characteristics of the phenomenon of investigation, which can be used for further analysis.

As suggested by Mayring (2008), the entire list of raw characteristics has to be iteratively reduced and qualitatively bundled until the main characteristics emerge. We depict some of the approaches available to operationalize this task in the following. A first approach is suggested by Eisenhardt and Bourgeois (1988) in iteratively comparing within-group similarities (i.e., groups of similar raw characteristics) and intergroup differences. The technique can be advanced by using matrices and introducing continuous measurement scales for comparison (Eisenhardt and Bourgeois 1988). As an alternative approach an iterative comparison of pairs can be used by listing similarities and differences for each pair (Eisenhardt 1989). Another way to operationalize the task of grouping the raw characteristics into categories of main characteristics might be based on the approach of Steininger et al. (2011b). They suggest having at least two independent researchers who are familiar with the topic judge proximities of paired raw characteristics in a matrix ranging from 100 to show perfect similarity to zero reflecting complete independence. Whichever approach is finally used, each of the resulting main characteristics should be labeled with a descriptive name, which is ideally developed inductively from associated bundles of raw characteristics (Mayring 2008). From these grouped resulting main characteristics of the phenomenon under investigation, a category or coding scheme of characteristics needs to be developed. This is reached through amending each main characteristic with explanations, “anchor examples” from the associated and coded raw characteristics and coding rules (i.e., rules regarding when an occurrence of a characteristic should be coded or excluded during analysis). For quality assurance the scheme might be tested by three or four judges following the suggestions of Moore and Benbasat (1991).

2.4 Formative Pretests and Deductive Content Analysis Procedures

In the following, we depict the deductive content analysis of the sampled entities based on the main characteristics coding scheme developed in the preceding steps. This is needed to ensure formative quality and reliability of the coding scheme and to find manifestations and densities of characteristics for each entity. A content analytical core component is the classification of the aforementioned units of analysis into the categories of characteristics by independent researchers. This process is typically referred to as “coding” (Scott 1955) and requires the category scheme of characteristics developed above. To capture word-spanning meanings and stabilize the results and reliabilities, we suggested the theme as the coding unit and entire sentences as the operationalized coding unit in this study, which leads to only coding a certain category once within one sentence (Kassarjian 1977). The finalized category scheme of characteristics (also referred to as coding scheme) is iteratively used and adjusted for an extensive training of coders. At least a second independent coder should be employed to ensure stable results and calculate intercoder reliabil-

ities (Mayring 2000). The coder(s) should be trained using research materials from *LL* with the lowest *SF*. The coding scheme and rules should be adjusted iteratively to sort out ambiguities through discussion of nonmatching codings. The procedure is repeated with different materials until the overall agreement (reliability) of all coders is calculated above 0.8 (cf. Moore 2000). This ensures intersubjectively comprehensible results and verifies the decency of the main characteristics coding scheme. Clearly distinguishable and exclusive categories of main characteristics are thereby ensured. We suggest using Krippendorff's Alpha (Hayes and Krippendorff 2007) for a sensitive and advanced measurement or the most commonly used simple "percent agreement" reliability measure of Holsti (1969). More details on possible measures, their mathematical references, advantages, and disadvantages are given in Table 3. All calculated reliabilities, discussions, and adjustments made to the coding scheme or the coding rules should be collected and given in a transparent and comprehensive manner for reproducibility (e.g., "If there are two occurrences of the same subcategory within one sentence, only the first occurrence should be

Table 3 Frequently cited measures of reliability for content analysis

Name	Advantages	Disadvantages
Krippendorff's Alpha (2007)	Allows any number of coders Takes into account agreements by chance Takes into account low coding numbers Takes into account number of categories Allows binary, nominal, ordinal, interval, ratio, polar, and circular data Allows measuring of incomplete data	Complex application Extensive details of data regarding coded occurrences needed
Holsti's Percent Agreement (1969)	Very facile and quick application Basic calculations	Does not take into account variables such as the number of categories, number of correct codings on incident, etc.
Scott's Pi (Scott 1955)	Relatively facile and quick application	Only allows nominal data Assumes same distribution of coder responses
Fleiss' Kappa (1971)	Relatively facile and quick application Extends Scott's Pi by allowing multiple coders	Only allows nominal data Assumes same distribution of coder responses
Cohen's Kappa (1960)	Takes into account agreements occurring by chance Does not assume same distribution of coder responses	Sometimes considered a too conservative measure Only allows measuring of two coders

coded, counted and marked”). Density results of the materials used for training shall be discarded after calculation of agreements and not be used for the building of final classes.

After finishing the aforementioned amendments to the coding scheme during the training session, the main coding process for all of the research material entities is initiated. This is done by analyzing all of the evidence of each entity for occurrences (i.e., manifestations) of the main characteristics categories. All manifestations should be marked and counted within the materials by category and entity. They are individually deemed as belonging to a certain category of characteristics. Finally, all manifestations in the evidence of each entity should be counted separately for every category. We suggest transforming these results into relative numbers (i.e., relative manifestations) and thereby making them comparable through dividing them by the number of averaged sentences in the sources of evidence for each entity. This number is calculated by counting the words of an entity’s sources of evidence and dividing the results by 22. The number 22 is the average of words contained within a sentence in English texts reported by Charniak (1996). For readability reasons, the averaged sentences are interchangeably referred to as “sentences” in the following. No further refinements to the coding scheme and coding rules within this main coding process should be made. Results are not to be exchanged or discussed by the coders during this main phase to avoid introducing any biases (Mayring 2000). It is suggested that coders be employed independently from the ones used for adjusting the coding scheme if possible. After finishing the coding process for all of the research material, the summative reliabilities should be calculated for the resulting manifestations. Pavlou and Dimoka (2006) suggest also calculating intra-coder reliabilities by having each coder re-code a sample after a certain time. There is no common absolute number of these agreements, which is found to be satisfactory in the academic discussion on reliabilities. This is due to large differences especially in the units of analysis and coding but also in category systems, complexity of the evaluated contents, and coder experience with the phenomenon. Nevertheless, a reliability of at least 0.7–0.85 is seen as acceptable and reachable by many authors (e.g., Mayring 2000; Krippendorff 2004a; Frueh 2007) for the “theme” as the unit of analysis that we suggest for this type of study.

2.5 Quantitative Clustering of Manifestations

Verifying the manifestations of the characteristics of each entity enables us to group the different entities. Thereby a set of classes (of entities) within the phenomenon of investigation can be identified. These classifications have usually been performed subjectively based on researchers’ ideas or intuition. Using our empirically derived and standardized densities instead leads to more objective classifications. Following the inductive procedure, again, no classes were predefined but instead derived inductively from the data sources.

Table 4 Manifestation matrix of entities and characteristics (cf. Steiningger et al. 2011b)

Characteristics entities	C ₁	C ₂	...	C _n
E ₁	x ₁₁	x ₁₂	...	x _{1n}
...
E _m	x _{m1}	x _{m2}	...	x _{mn}

Table 5 Distance matrix of entities

Entities	E ₁	E ₂	E ₃	...	E _{m-1}
E ₁	d ₂₁ = d ₁₂				
E ₂	d ₁₃	d ₂₃			
E ₃	d ₁₄	d ₂₄	d ₃₄		
...	
E _m	d _{1m}	d _{(m-1)m}

The main goal of this step is to identify classes that are mutually exclusive and collectively exhaustive. This means that there must be an appropriate class for each entity and each entity must fit into one class only (Bailey 1994). Furthermore, the classification should be generally applicable. The latter requirement is met by the extensive sampling method applied earlier, which ensures that the data used appropriately represents the phenomenon. The former two requirements are addressed by cluster analysis. Cluster analysis generally aims at finding classes such that entities within the same group are similar to each other while entities in different groups are as dissimilar as possible. The five typical steps of cluster analysis are outlined based on our problem (Aldenderfer and Blashfield 1984): (1) selection of a sample to be clustered, (2) definition of a set of variables on which to measure the entities in the sample, (3) computation of similarities among the entities, (4) use of a cluster analysis method to create groups of similar entities, and (5) validation of the resulting cluster solution.

The first step, selecting the sample, has already taken place. Regarding the selection of the cluster variables, which is usually a complex procedure (Fowlkes et al. 1988), it is again very helpful that we have already identified and reduced the relevant characteristics in the previous qualitative steps. Therefore, we can directly create the data matrix containing the densities of the characteristics that correspond to the different entities (see Table 4). In the next step, the similarity calculation takes place. Due to the standardized scale of manifestations (i.e., relative manifestations), the Minkowski distance¹ can be used to calculate these values without having to compute weights for the different characteristics (Kaufman and Rousseeuw 1990) (cf. Table 5). The elimination of potential single outliers that exhibit a large distance to all other entities should be checked manually by an in-depth analysis of the underlying data of this entity. Rash elimination of entities can lead to problems in the validity of the resulting taxonomy and should be avoided.

¹ $d(i, j) = (|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + |x_{i3} - x_{j3}|^q + \dots + |x_{in} - x_{jn}|^q)^{1/q}$, where q is a natural number larger or equal to 1, and describes the distance between the entities i and j . Most algorithms use Manhattan distances ($q = 1$) or Euclidian distances ($q = 2$).

Many different analysis techniques can be applied in order to derive clusters from this data. Generally, partitioning methods like K-Means (Howard and Harris 1966) have been shown to be superior to hierarchical methods in this case (Punj and Stewart 1983). Nevertheless, these methods need a priori information about the starting points and the number of clusters, which may not be available when investigating a new phenomenon inductively. In this case, it might be useful to apply Ward's minimum variance method (Ward 1963) to derive preliminary clusters. Their center can then be used in a partitioning algorithm like K-Means (Punj and Stewart 1983). Common software packages such as SPSS or SAS can be used to process steps 3 and 4.

Despite the importance of exhaustiveness and mutual exclusiveness, further quality indicators can be addressed. Checking the quality of classifications has been discussed in detail by Aldenderfer and Blashfield (1984). They suggest two major techniques that are relevant to our procedure: significance tests and replication. Multivariate analysis of variance (MANOVA) or discriminant analysis can be used to check the significance of the clusters. However, this method has been criticized for indicating high significance even for very bad clusters. A solution to this problem might be the inclusion of external variables, which is difficult when analyzing a new phenomenon (Aldenderfer and Blashfield 1984). The replication technique can be used to check for internal consistency of the classification. If the base of entities is large enough, the split-half method can be applied.

Two random sets of entities are clustered independently using the same clustering method. If the same classes occur across different subsets of entities, this indicates further generalizability of the classification. Another form of replication is to use different clustering methods with the same data. If the same clusters are derived, the results indicate a high validity of the classification (Aldenderfer and Blashfield 1984). After having the clusters validated, the different classes have to be interpreted. For better understanding, they should also be described verbally. This usually complex task can be accomplished using the codings and descriptions of the entities within one class. The distribution of these codings already describes the characteristics of a certain class. If the number of entities in one class is very high, the naming should be based on the characteristics of the entities in the center of the class. The clusters should then be named inductively out of the names and characteristics from their associated entities (Mayring 2008).

2.6 Summative Checks of Taxonomic Quality

As discussed earlier, checking taxonomic quality is a very challenging task. Mutual exclusiveness and collective exhaustiveness are the two major quality measures that a high-quality taxonomy has to meet (Bailey 1994). In order to increase and verify the validity of our method, we suggest performing an additional (optional) step to test discriminant validity of the classification via a sorting procedure (see: Davis 1989; Moore and Benbasat 1991). If additional entities that have not been used

for the taxonomy-building process are available, these entities should be combined with the entities from the sample into a common pool. The additional entities can be coded using the deductive procedure outlined earlier and can then be sorted into the classes mathematically to also obtain their class affiliations for subsequent comparison. Three to four judges are given the names and verbal descriptions of the classes that have been derived in the previous steps. The judges now sort all entities from the pool into the classes. Two measures can be applied to the results of this sorting process.

The first one measures the inter-judge reliability and focuses on the question of judges sorting the same entities into the same classes. We again suggest Krippendorff's Alpha (Hayes and Krippendorff 2007) or Holsti's percent agreement (Holsti 1969) to measure the level of agreement between the judges and thereby determine whether or not the descriptions precisely define the classes. Reliabilities above 0.7 can be seen as satisfactory (Krippendorff 2004a). If this level is not reached, the descriptions of the classes should be enhanced iteratively. A lack of increased inter-judge reliability even with refined descriptions indicates a general problem regarding the mutual exclusiveness or the collective exhaustiveness. Furthermore, for each class, a cumulated overall measure of correctly placed entities can be calculated.² This differs from the previous measure since it challenges the strength of the different classes separately. No description of a reasonable score for this measure is described in the literature. As a rule of thumb, the interval between 0.7 and 0.85 discussed above (Mayring 2000; Krippendorff 2004b; cf. Frueh 2007) can also be applied as a good indicator for this measure. A high value points to high construct validity and reliability of the class. This method can also be used qualitatively to identify critical class definitions and borders between two classes that should be refined.

2.7 Limitations of the Method

Potential limitations regarding the procedures introduced throughout this chapter should be taken into account. They are given below and if countermeasures do exist, they are also depicted in the following. Overall, we have tried to keep the complexity of the process low. Nevertheless, it might inhibit broader use. The process of inductively constructing raw characteristics from the entities is continued until saturation (Glaser and Strauss 1967). This allows real knowledge and deep insights to be gained in classes. Nevertheless, theoretical saturation is critical to identify. This might lead to missing definitions of characteristics threatening the collective exhaustiveness. The probability seems low since we suggested measures to objectify

²The overall measure for the quality of the class is defined as $B(i) = \frac{\#E_c}{\#E} \in (0, 1)$, where $\#E_c$ is the number of correctly selected entities into class i by all judges and $\#E$ describes the number of entities supposed to be sorted into this class.

significant saturation within the inductive process. Inductively built categories might also be biased by a coder's world views or insights on the phenomenon. The likelihood of such a bias might be lowered through introducing more than one coder to inductively build the raw characteristics. Construction of main characteristics from raw characteristics might also be subject to the coder's bias since they are qualitatively clustered. Improvement within this area might be reached by applying large proximity matrices judged by more than one person and statistical cluster analysis for their entire set.

The method of using averaged sentences for comparability reasons might lead to excessive numbers of coded sentences since figures or tables within the sources of evidence might be handled as text. This is additionally fostered by the assumption during calculations that all sentences only contain one code, which must not hold true since the rules allow coding a sentence twice with two different categories. One major critique regarding cluster analysis is that it lacks a theoretical foundation. Therefore the identified clusters may simply be statistical artifacts that capitalize on random numerical variation across entities (Thomas and Venkatraman 1988). Furthermore, cluster analysis might also find classes in situations where no clusters exist (e.g., Aldenderfer and Blashfield 1984). Our approach tries to invalidate the criticisms partly because the clusters are directly named and described based on the densities of their characteristics and are therefore not artificial constructs (Mayring 2008). Another main critique of cluster analysis is the potential multicollinearity among characteristics that may lead to overweighting of certain aspects (Ketchen and Shook 1996). Using more advanced distance measures such as the Mahalanobis distance might solve this issue (Hair et al. 2005), but this measure is supported neither by Ward's minimum variance method (Ward 1963) nor by software such as SPSS and SAS. However, our approach addresses this issue early in the research process. Since the characteristics of the topic are inductively derived from the raw categories and by controlling for weakness of the single characteristics (Frueh 2007), the risk of multicollinearity issues is reduced.

3 Conclusion

Throughout this chapter, we outlined and developed a method of building taxonomic classification schemes for business disciplines. Although the importance of such classifications is seen as very high in the research community (Wolf 1926; Kantor 1953; Kemeny 1959; Lambert 2006), these classifications have usually been performed subjectively based on researchers' ideas or intuition. The delineated approach enables researchers to derive classifications empirically, leading to more objective classifications (Bailey 1994). In essence, we proposed six subsequent steps relying on content and cluster analysis. Especially the use of content analysis in this context enhances the available set of techniques within our field. The first step begins with the sampling of entities and their sources of evidence as instantiations or examples of the topic. Since our method focuses on new and unexplored topics

of investigation, we assumed that no theoretical basis of the topic was available. Accordingly, the second and third steps proposed to develop the characteristics of the topic from selected entities by using inductive content analysis procedures.

Based on these results we proposed a fourth step of deductive content analysis to find manifestations and densities of the derived characteristics for each entity. Cluster analysis was then applied to identify specific classes in the research material, leading to a taxonomic classification scheme. Formative state-of-the-art procedures for quality assurance were suggested throughout all steps of the method. Additionally, summative measures of taxonomic quality for the resulting constructs were outlined. We hope that our results will help academics to develop empirically grounded rigorous taxonomies in their fields of research by applying our suggestions, guidelines, and depicted alternatives. Taxonomies are important vehicles in IS and management research since they allow investigations on a topic to be limited to certain subclasses or taxa, which makes research projects more manageable. Lastly, they are of high value for intra- and interclass generalization, enabling the development of theories through analysis of these classes and their generalizations. There are innumerable applications of our method in the field of business and technology research. New and upcoming phenomena such as cloud computing applications and crowdsourcing services might require taxonomic classification, but also long-standing nonempirically grounded typologies in areas such as outsourcing, operational application software systems, or electronic business model research might be revisited and updated by applying our method to the topic.

References

- Abbasi A, Chen H (2008) Cybergate: A Design Framework and System for Text Analysis of Computer-Mediated Communication. *MIS Q* 32:811–837
- Al-Debei MM, Avison D (2010) Developing a Unified Framework of the Business Model Concept. *Eur J Inf Syst* 19:359–376
- Aldenderfer M, Blashfield R (1984) Cluster Analysis. Beverly Hills Sage Univ Pap
- Angelmar R, Stern LW (1978) Development of a Content Analytic System for Analysis of Bargaining Communication in Marketing. *J Mark Res* 15:93–102
- Bailey KD (1994) Typologies and Taxonomies: An Introduction to Classification Techniques. Sage Publications, Thousand Oaks
- Charniak E (1996) Tree-Bank Grammars. In: Proceedings of the National Conference on Artificial Intelligence. pp 1031–1036
- Cohen J (1960) A Coefficient of Agreement for Nominal Scales. *Educ Psychol Meas* 20:37–46. <https://doi.org/10.1177/001316446002000104>
- Davis FD (1989) Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q* 13:319–340
- Eisenhardt KM (1989) Building Theories from Case Study Research. *Acad Manage Rev* 14:532–550
- Eisenhardt KM, Bourgeois LJ (1988) Politics of Strategic Decision Making in High-Velocity Environments: Toward a Midrange Theory. *Acad Manage J* 31:737–770. <https://doi.org/10.2307/256337>

- Fleiss JL (1971) Measuring Nominal Scale Agreement Among Many Raters. *Psychol Bull* 76:378–382. <https://doi.org/10.1037/h0031619>
- Fowlkes EB, Gnanadesikan R, Kettnering JR (1988) Variable Selection in Clustering. *J Classif* 5:205–228
- Frueh W (2007) *Inhaltsanalyse: Theorie und Praxis* (in German), 6th edn. UVK Verlagsgesellschaft, Konstanz
- Glaser BG, Strauss AL (1967) *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine, New York
- Hair JF, Black B, Babin B (2005) *Multivariate Data Analysis*, 6th ed. Prentice Hall
- Hayes AF, Krippendorff K (2007) Answering the Call for a Standard Reliability Measure for Coding Data. *Commun Methods Meas* 1:77–89
- Holsti OR (1969) *Content Analysis for the Social Sciences and Humanities*. Addison-Wesley, Reading
- Howard N, Harris B (1966) A Hierarchical Grouping Routine, IBM 360/65 Fortran IV program. Univ Pa Comput Cent
- Insch GS, Moore JE, Murphy LD (1997) Content Analysis in Leadership Research: Examples, Procedures, and Suggestions for Future Use. *Leadersh Q* 8:1–25
- Kantor JR (1953) *The Logic of Modern Science*. University of Akron Press, Akron
- Kassarjian H (1977) Content Analysis in Consumer Research. *J Consum Res* 4:8–18
- Kaufman L, Rousseeuw PJ (1990) *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley Online Library
- Kemeny JG (1959) *A Philosopher Looks at Science*. Van Nostrand, Princeton
- Ketchen DJ, Shook CL (1996) The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique. *Strateg Manag J* 17:441–458. [https://doi.org/10.1002/\(SICI\)1097-0266\(199606\)17:6<441::AID-SMJ819>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G)
- Krippendorff K (2004a) *Content Analysis: An Introduction to its Methodology*, 2nd edn. Sage Publications, Thousand Oaks
- Krippendorff K (2004b) Reliability in Content Analysis: Some Common Misconceptions and Recommendations. *Hum Commun Res* 30:411–433
- Lambert S (2006) Do We Need a “Real” Taxonomy of e-Business Models? In: *ACIS 2006 Proceedings. 17th Australasian Conference on Information Systems*
- Mayring P (2000) Qualitative Content Analysis. *Forum Qual Soc Res* 1:1–10
- Mayring P (2002) *Einführung in die qualitative Sozialforschung: Eine Anleitung zu qualitativem Denken* (in German), 5th edn. Beltz Verlag, Weinheim
- Mayring P (2008) *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (in German), 10th edn. Beltz, Weinheim
- Moore JE (2000) One Road to Turnover: An Examination of Work Exhaustion in Technology Professionals. *MIS Q* 24:141–168
- Moore GC, Benbasat I (1991) Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation. *Inf Syst Res* 2:192–222
- Myers MD (2008) *Qualitative Research in Business & Management*, illustrated edition. Sage Publications Ltd.
- Nag R, Hambrick DC, Chen M-J (2007) What is Strategic Management, Really? Inductive Derivation of a Consensus Definition of the Field. *Strateg Manag J* 28:935–955. <https://doi.org/10.1002/smj.615>
- Nickerson RC, Varshney U, Muntermann J (2013) A Method for Taxonomy Development and Its Application in Information Systems. *Eur J Inf Syst* 22:336–359. <https://doi.org/10.1057/ejis.2012.26>
- Oberländer AM, Lösser B, Rau D (2019) TAXONOMY RESEARCH IN INFORMATION SYSTEMS: A SYSTEMATIC ASSESSMENT. Res Pap
- Pavlou PA, Dimoka A (2006) The Nature and Role of Feedback Text Comments in Online Marketplaces: Implications for Trust Building, Price Premiums, and Seller Differentiation. *Inf Syst Res* 17:392–414. <https://doi.org/10.1287/isre.1060.0106>

- Punj G, Stewart DW (1983) Cluster Analysis in Marketing Research: Review and Suggestions for Application. *J Mark Res* 20:134–148. <https://doi.org/10.2307/3151680>
- Saris-Gallhofer IN, Saris WE, Morton EL (1978) A Validation Study of Holsti's Content Analysis Procedure. *Qual Quant* 12:131–145. <https://doi.org/10.1007/BF00144066>
- Scott W (1955) Reliability of Content Analysis: The Case of Nominal Scale Coding. *Public Opin Q* 19:321–325. <https://doi.org/10.1086/266577>
- Steininger DM (2019) Linking Information Systems and Entrepreneurship: A Review and Agenda for IT-Associated and Digital Entrepreneurship Research. *Inf Syst J* 29:363–407
- Steininger DM, Huntgeburth JC, Veit D (2011a) Conceptualizing Business Models for Competitive Advantage Research by Integrating the Resource and Market-Based Views. In: *AMCIS 2011 Proceedings*. pp 1–12
- Steininger DM, Huntgeburth JC, Veit DJ (2011b) A Systemizing Research Framework for Web 2.0. In: *ECIS 2011 Proceedings*. AISel, Helsinki, Finland, pp 1–13
- Thomas H, Venkatraman N (1988) Research on Strategic Groups: Progress and Prognosis. *J Manag Stud* 25:537–555. <https://doi.org/10.1111/j.1467-6486.1988.tb00046.x>
- Veit D, Clemons E, Benlian A, et al (2014) Business Models: An Information Systems Research Agenda. *Bus Inf Syst Eng* 6:45–53. <https://doi.org/10.1007/s12599-013-0308-y>
- Ward JH (1963) Hierarchical Grouping to Optimize an Objective Function. *J Am Stat Assoc* 58:236–244
- Wolf A (1926) *Essentials of Scientific Method*. Macmillan, New York
- Yin RK (2009) *Case Study Research: Design and Methods*, 4th edn. Sage Publications, Inc, Thousand Oaks

Dennis M. Steininger is an assistant professor of Information Systems and Digital Entrepreneurship at the Faculty of Business and Economics of the University of Augsburg, Germany. In addition, he is co-managing director of the Augsburg Center for Entrepreneurship as well as a research fellow and guest lecturer in the Institute for SME and Entrepreneurship Research at the University of Mannheim, Germany. He received his PhD in Business Information Systems from the University of Mannheim (2016), Germany, supervised by Professor Dr. Daniel J. Veit. His research focuses on the success factors and impacts of digital entrepreneurship and related business models. His work has been published in the *Journal of Management Information Systems*, *Information Systems Journal*, *Electronic Markets*, and *Technological Forecasting and Social Change*, among others. Dennis is currently serving as a guest editor of a special issue on digital entrepreneurship at *Business & Information Systems Engineering* and as track chair and associate editor at a number of leading information systems conferences.

Manuel Trenz is Chair of Interorganizational Information Systems at the Faculty of Business and Economics of the Georg-August-University of Göttingen, Germany. He received his PhD in Business Information Systems from the University of Mannheim (2014), Germany, and his habilitation from the University of Augsburg (2019), Germany, both supervised by Professor Dr. Daniel J. Veit. His research addresses the targeted use of information systems and digital innovations and their influence on organizations and individuals. In particular, he focuses on interactions with and the impact of digital services and platforms, as well as the intersections between physical and digital worlds. His work has been published in journals such as *Management Information Systems Quarterly*, the *Journal of Management Information Systems*, the *European Journal of Information Systems*, *Information & Management*, and *Business & Information Systems Engineering*. Manuel is currently serving as senior editor for *Internet Research*, as associate editor for the *Information Systems Journal*, and as a member of the editorial board of *Electronic Markets*. In addition, he frequently serves as a track and workshop chair at leading information systems conferences.

Daniel J. Veit currently a full professor and Chair of Information Systems and Management at the Department of Business Administration of the Faculty of Business and Economics of the University of Augsburg, Germany. He holds a PhD in Economics and Business Engineering (2002) as well as a habilitation degree in Business Administration (2006), both from the University of Karlsruhe (TH), Germany, supervised by Professor Dr. Christof Weinhardt. He is also a visiting professor of Information Systems at the Department of Digitalization at Copenhagen Business School, Denmark. His research focuses on the transformational effects of information systems and digitalization in society, with a specific focus on sustainability. His publications have appeared in outlets such as the *Management Information Systems Quarterly*, the *Journal of Management Information Systems*, the *European Journal of Information Systems*, the *Journal of Service Research*, *Information & Management*, *Internet Research*, *MIS Quarterly Executive*, the *Journal of Business Economics*, and *Business & Information Systems Engineering*. He serves as an Editorial Board Associate Editor for the *Journal of the Association for Information Systems*, as well as associate editor for the *Information Systems Journal*. In addition, he serves as a member of the distinguished board of advisors for *Internet Research*. In 2019 he received the “Outstanding Reviewer Award” of the *Journal of the Association for Information Systems*. He has frequently served as a track chair and associate editor for leading conferences in the information systems field. In 2014 he received the conference best paper award at the European Conference on Information Systems in Tel Aviv, Israel. In 2016 he served as a program co-chair of the European Conference on Information Systems in Istanbul, Turkey. He is the principal investigator of a German Federal Ministry of Education and Research grant awarded to study the impact of the sharing economy on German society. Earlier in his career he was admitted to the young researchers’ promotion program of the Volkswagen Foundation. During the past 14 years he has served as Associate Dean for international affairs and Academic Director of the ESSEC & Mannheim Executive MBA Program at Mannheim Business School, Germany, among other positions.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

