UNIVERSITÄT AUGSBURG
Universitätsbibliothek

OPEN ACCESS
OPUS AUGSBURG

Building taxonomies in IS and management: a systematic approach based on content analysis

Dennis M. Steininger, Manuel Trenz, Daniel Veit

Angaben zur Veröffentlichung / Publication details:

Steininger, Dennis M., Manuel Trenz, and Daniel Veit. 2013. "Building taxonomies in IS and management: a systematic approach based on content analysis." In *Proceedings of the 11th International Conference on Wirtschaftsinformatik (WI2013), 27.02. - 01.03.2013, Universität Leipzig, Germany*, 90. New York, NY: AISeL. http://aisel.aisnet.org/wi2013/90/.

Nutzungsbedingungen / Terms of use:

licgercopyright

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:
Deutsches Urheberrecht
Weitere Informationen finden Sie unter: / For more information see:
https://www.uni-augsburg.de/de/organisation/bibliothek/publizieren-zitieren-archivieren/publiz/

Building Taxonomies in IS and Management – A Systematic Approach Based on Content Analysis

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

University of Mannheim, Business School, Mannheim, Germany
{steininger,trenz,veit}@bwl.uni-mannheim.de

Abstract. Classification schemes such as taxonomies are important groundwork for research on many topics in Information Systems (IS) and Management. They make investigating topics manageable by allowing researchers to delimit their work to certain taxa or types and provide a basis for generalization. Opposed to theoretically grounded typologies, taxonomies are empirically derived from entities of a phenomenon under investigation and therefore have several advantages such as more detailed and exhaustive coverage. Nevertheless, research is still missing a clear set of procedures on how 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, alternatives and formative measures of reliability and validity.

Keywords: Taxonomy, Classification, Typology, Content analysis, Method

1 Introduction

Classification schemes allow the systematic ordering or sorting of phenomena into similar groups or classes. They are of fundamental importance for science and academic research in disciplines such as Information Systems (IS) and Management [1-2]. Wolf [3] 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" [3]. Hence, classification schemes such as taxonomies make investigating 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 building theories that apply to certain classes of the developed schemes. When classifying an area of investigation, two different approaches can be used: typologies or taxonomies. Typologies are created deductively by classifying the objects into predefined groups that are created based on intuition or previously existing knowledge and theory [4]. Especially when examining an unexplored area of research, the threat of researcher bias or general misconception is very high, since existing theory is limited. Opposed to theoretically grounded typologies, taxonomies are derived inductively from empirical data (i.e. entities of a phenomenon under investigation) and therefore have several

11th International Conference on Wirtschaftsinformatik,
27th February – 01st March 2013, Leipzig, Germany

advantages such as more detailed and exhaustive coverage and mutual exclusiveness of classes. Nevertheless, IS research is still missing a clear set of rigorous procedures on how to empirically build taxonomies of 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 outlined above we propose the following research question for this article:

How and by which procedures can mutually exclusive and collectively exhaustive taxonomies be built rigorously in the IS and Management disciplines from empirical entities?

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 under investigation to get a first grasp on 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 paper 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 formative 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-related phenomena using content and cluster analysis. The process is based on Steininger et al. [5] who use clustering and mainly content analysis to inductively build a taxonomic framework of Web 2.0 characteristics. This article can be seen as a working example. We added inspirations from the articles of Nag et al. [6], defining Strategic Management via content analysis and clustering and Al-Debei and Avison [7] developing a business model framework through content analysis. Content analysis is a scientific research technique to gain "replicable and valid inferences from text" [8] and thereby find trends, characteristics, patterns or densities. Materials for analysis might include written or spoken texts as transcripts. 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 [5], [9-11] and distinguish content analysis from regular critical reading. The aforementioned potential in rigorously and reliably uncovering 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 suggested in the remainder of this paper. 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.

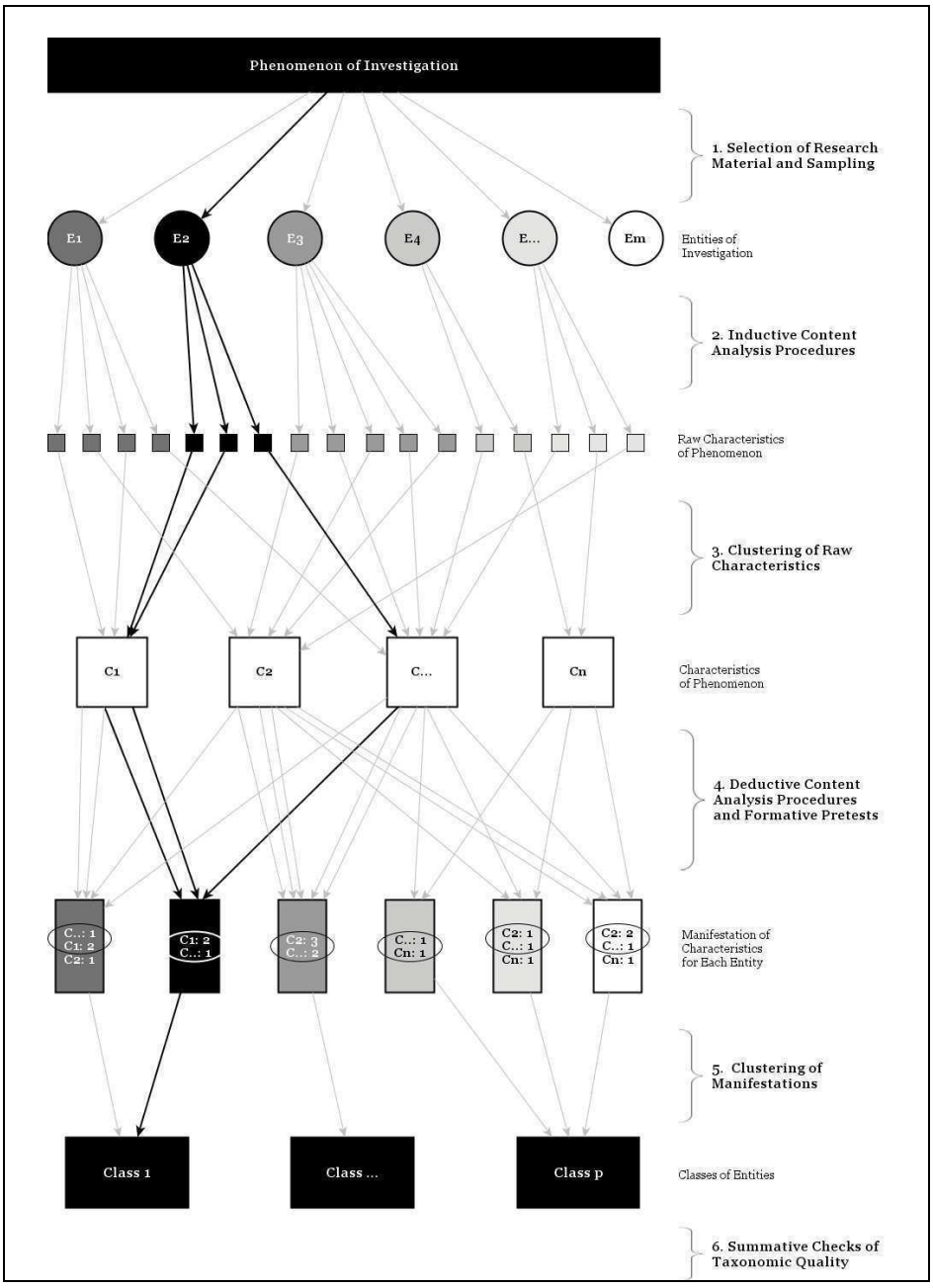


Fig. 1. Detailed Overview on the Taxonomy-Building Process

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 Figure 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) are required as a basis for analysis, which is addressed in our first suggested step on the selection and sampling of entities. Analyzing the manifestation of the phenomenon's characteristics for each entity is needed to proceed in building the taxonomy. 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 suggested to be 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 forth 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 by propositions and guidelines on measures for taxonomic quality (e.g. mutual exclusiveness). Details and guidelines on each of our suggested steps are given in the sections below.

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 neither be manageable nor required. Instead, we propose to follow a theoretical sampling approach as suggested by Eisenhardt [12]. This means broadly choosing the entities of investigation for variation, heterogeneity (i.e. unique cases) or replication instead of random selection [13]. 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.

Table 1. Possible Sources of Descriptions for Entities [13-14]

Name	Application	Advantages	Disadvantages
Documentation and Archival Records	Usually available in written form.	• Stable • Unobtrusive	• Bias of author unknown • Retrievability and Access
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

It is recommendable 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. [13]. We suggest listing derived entities in a longlist (*LL*). If entities are gained from different sources, this list needs to be cleaned from possible duplicates. The introduction of a selection factor (*SF*) can help to prepare the *LL* for further proceedings [5]. 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 gain this shortlist might also be applied (i.e. taking a sample of entities from an existing journal paper on the phenomenon). The SL is to 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 suggested outline, 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.

After specification of the entities of investigation and their sampling as research material, the unit of analysis needs to be defined subsequently. This addresses the issue of "the basic unit of text [e.g. word or paragraph] to be classified" [15] into the categories of characteristics derived in succeeding steps. The configuration of this unit has a considerable impact on 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 [16]. Following Kassarjian [17], the 'theme' is usually suggested for this type of taxonomic method ensuring the capturing of word or sentence-spanning ideas es-

pecially within the inductive phase of building raw characteristics. To stabilize the results and reliabilities, entire sentences are to be used as the operationalized coding unit, which leads to solely coding a category once within one sentence [5]. In the suggested approach the raw characteristics are to 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.

Table 2. Units of Analysis and Implications (adapted from [17])

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

Based on raw characteristics building rules [18], the research material needs to be worked through consecutively and raw characteristics are to be defined beginning with the first selected entity of investigation. Each occurrence of a new or additional raw characteristic-building incident needs to be marked and uniquely numbered using the research material ID *i* (cf. Section 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 with an existing raw characteristic and only adds richness to the description of the characteristic, that existing characteristic number is to be used instead, and the mark is suggested to be set in a different color (e.g. dark blue). All raw characteristics categories are suggested to be summoned in a list (RcL). This process is to be continued until saturation is reached (i.e. no new raw characteristics can be derived from the research material entities) [19].

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 on the phenomenon of investigation, which can be used for further analysis.

As suggested by Mayring [19], the entire list of raw characteristics has to be iteratively reduced and qualitatively bundled until 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 [20] 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 [20]. As an alternative approach an iterative comparison of pairs can be used by listing similarities and differences for each pair [12]. 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. [5]. 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 is to be provided with a descriptive name, which is ideally developed inductively from associated bundles of raw characteristics [19]. 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 on when an occurrence of a characteristic needs to 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 [21].

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 aforementioned units of analysis into the categories of characteristics by independent researchers. This process is typically referred to as 'coding' [22] 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 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 [17]. The finalized category scheme of characteristics (also referred to as coding scheme) is iteratively to be used and adjusted for an extensive training of coders. At least a second independent coder needs to be employed to ensure stable results and calculate intercoder reliabilities [23]. The coder(s) need(s) to be

trained in a one day workshop using research materials from LL with the lowest SF. The coding scheme and rules have to be adjusted iteratively to sort out ambiguities through discussion of non-matching codings. The procedure is repeated with different materials until the overall agreement (reliability) of all coders is calculated above 0.8, cf. [24]. 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 Krippendorf's Alpha for a sensitive and advanced measurement or the most commonly used simple 'percent agreement' reliability measure of Holsti [25]. 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 need to 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 is to be 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 the entire research material entities is started. This is done by analyzing the entire evidence of each entity for occurrences (i.e. manifestations) of the main characteristics categories. All manifestations are to 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 are to 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 of 22 is the average of words contained within a sentence in English texts reported by Charniak [26]. 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 are to be made. Results are not to be exchanged or discussed by the coders during this main phase [23]. It is suggested to employ coders independently from the ones used for adjusting the coding scheme if possible. After finishing the coding process of the entire research material, the summative reliabilities need to be calculated for the resulting manifestations. Pavlou and Demoka [27] suggest also calculating intracoder 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 on the phenomenon. Nevertheless, a reliability of at least 0.7 to 0.85 is seen as acceptable and reachable by many authors (e.g. [8], [23], [28]) for the 'theme' as the unit of analysis that we suggest for this type of study.

Table 3. Frequently Cited Measures of Intercoder Reliability for Content Analysis

Name	Advantages	Disadvantages
Krippendorf's Alpha [29]	• 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 [25]	• Very facile and quick application • Basic calculations	Does not take into account variables such as the number of categories, correct codings on incident etc.
Scott's Pi [30]	• Relatively facile and quick application	• Only allows nominal data • Assumes same distribution of coder responses
Fleiss' Kappa [31]	• Relatively facile and quick application • Extents Scott's Pi by allowing multiple coders	• Only allows nominal data • Assumes same distribution of coder responses
Cohen's Kappa [32]	• 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

2.5 Quantitative Clustering of Manifestations

Having verified 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.

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 [4]. Furthermore the classification should be generally applicable. The latter requirement is met by the extensive sampling method applied before 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 [33]: (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, (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 usually is a very complicated procedure [34], 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 (cf. Table 4). In the next step, the similarity calculation takes place. Due to the standardized scale of manifestations (i.e. relative manifestations), the Minkowski distance[1] can be used to calculate these values without having to compute weights for the different characteristics [35] (cf. Table 5). The elimination of potential single outliers that have a high distance to all other entities has to 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.

Many different cluster methods can be applied in order to derive clusters from this data. Generally, partitioning methods like K-Means [36] have been shown to be superior to hierarchical methods in this case [37]. 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 a apply Ward's minimum variance method [38] to derive preliminary clusters. Their center can then be used in a partitioning algorithm like K-Means [37]. Common software package like SPSS or SAS can be used to process steps 3 and 4.

Table 4. The Manifestation Matrix of Entities and Characteristics (cf. [5])

Characteristics / **Entities**	$\mathbf{C_1}$	$\mathbf{C_2}$		$\mathbf{C_n}$
$\mathbf{E_1}$	x_{11}	x_{12}	…	x_{1n}
…	…	…	…	…
$\mathbf{E_m}$	x_{m1}	x_{m2}	…	x_{mn}

Table 5. The Distance Matrix of Entities

Entities	$\mathbf{E_1}$	$\mathbf{E_2}$	$\mathbf{E_3}$	**…**	$\mathbf{E_{m-1}}$
$\mathbf{E_1}$	$d_{21} = d_{12}$				
$\mathbf{E_2}$	d_{13}	d_{23}			
$\mathbf{E_3}$	d_{14}	d_{24}	d_{34}		
…	...	…	…	…	
$\mathbf{E_m}$	d_{1m}	…	…	…	$d_{(m-1)m}$

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 [33]. 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 for this problem might be

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

the inclusion of external variables which is difficult when analyzing a new phenomenon [33]. 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 parts of the same 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 way 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 [33]. 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 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 [19].

2.6 Summative Checks of Taxonomic Quality

As discussed before, 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 [4]. 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 (based on [21], [39]). If additional entities that have not been used for the taxonomy building 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 before and then be sorted into the classes mathematically to also have 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 Krippendorf's Alpha [29] or Holsti's percent agreement [25] 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 [8]. If this level is not reached the descriptions of the classes need to 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

[2] The 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.

described in literature. As a rule of thumb, the interval between 0.7 and 0.85 discussed above [23], cf. [28], [40] 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 rather 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 article should be taken into account. They are given below and if counter measures 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. This allows gaining real knowledge and deep insights on 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 significant saturation within the inductive process. Inductively built categories might also be biased by a coder's world views or insights on the phenomenon. Lowering the likelihood of such a bias might be reached through introducing more than one coder for inductively building the raw characteristics. Construction of main characteristics from raw characteristics might also be subject to 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 theoretical foundation. Therefore the identified clusters may simply be statistical artifacts that capitalize on random numerical variation across entities [41]. Furthermore, cluster analysis might also find classes in situations where no clusters exist, e.g. [33]. 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 no artificial constructs [19]. Another main critique of cluster analysis is the potential multicollinearity among characteristics that may lead to overweighting of certain aspects [42]. Using more advanced distance measures like the Mahalanobis distance might solve this issue [43], but this measure is supported neither by Ward's minimum variance method [38] nor by software like SPSS and SAS. However, our approach addresses this issue early in the research process. Since the characteristics of the topic are derived from the raw categories inductively and by controlling for weakness of the single characteristics [28], the risk of multicollinearity issues is reduced.

3 Conclusion

We outlined and developed a method of building taxonomic classification schemes for the IS and Management disciplines throughout this papers. Although the importance of such classifications is seen as very high in the research community [1-3], [44], 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 [4]. 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 is focusing on new and unexplored topics of investigation, we assumed no theoretical basis of the topic to be available. Accordingly, the second and third step 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 is 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 are outlined. We conclude with an extensive discussion of potential limitations of our method. We believe 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 limiting investigations on a topic to certain subclasses or taxa, which makes research projects more manageable. Lastly, they are of high value for intra- and inter-class 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 IS research. New and upcoming phenomena such as cloud computing applications, crowdsourcing services might need taxonomic classification, but also long standing non-empirically 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

1. Kantor, J.R.: The Logic of Modern Science. University of Akron Press, Akron (1953)
2. Kemeny, J.G.: A Philosopher Looks at Science. Van Nostrand, Princeton (1959)
3. Wolf, A.: Essentials of Scientific Method. Macmillan, New York (1926)
4. Bailey, K.D.: Typologies and Taxonomies: An Introduction to Classification Techniques. Sage Publications, Thousand Oaks (1994)
5. Steininger, D.M., Huntgeburth, J.C., Veit, D.J.: A Systemizing Research Framework for Web 2.0. In: ECIS 2011 Proceedings, pp. 1–13. AISeL, Helsinki, Finland (2011)

6. Nag, R., Hambrick, D.C., Chen, M.-J.: What is Strategic Management, Really? Inductive Derivation of a Consensus Definition of the Field. Strat. Mgmt. J. 28, 935–955 (2007)
7. Al-Debei, M.M., Avison, D.: Developing a Unified Framework of the Business Model Concept. European Journal of Information Systems 19, 359–376 (2010)
8. Krippendorff, K.: Content Analysis: An Introduction to its Methodology. Sage Publications, Thousand Oaks (2004)
9. Abbasi, A., Chen, H.: CyberGate: A Design Framework and System for Text Analysis of Computer-mediated Communication. MIS Quarterly 32, 811–837 (2008)
10. Angelmar, R., Stern, L.W.: Development of a Content Analytic System for Analysis of Bargaining Communication in Marketing. Journal of Marketing Research 15, 93–102 (1978)
11. Steininger, D.M., Huntgeburth, J.C., Veit, D.J.: Conceptualizing Business Models for Competitive Advantage Research by Integrating the Resource and Market-Based Views. In: AMCIS 2011 Proceedings, Detroit, USA (2011)
12. Eisenhardt, K.M.: Building Theories from Case Study Research. Academy of Management Review 14, 532–550 (1989)
13. Yin, R.K.: Case Study Research: Design and Methods. Sage Publications, Inc, Thousand Oaks (2009)
14. Myers, M.D.: Qualitative Research in Business & Management. Sage Publications Ltd. (2008)
15. Insch, G.S., Moore, J.E., Murphy, L.D.: Content Analysis in Leadership Research: Examples, Procedures, and Suggestions for Future Use. The Leadership Quarterly 8 (1), 1–25 (1997)
16. Saris-Gallhofer, I.N., Saris, W.E., Morton, E.L.: A Validation Study of Holsti's Content Analysis Procedure. Quality and Quantity 12 (2), 131–145 (1978)
17. Kassarjian, H.: Content Analysis in Consumer Research. Journal of Consumer Research 4, 8–18 (1977)
18. Mayring, P.: Einführung in die qualitative Sozialforschung: Eine Anleitung zu qualitativem Denken (in German). Beltz, Weinheim (2002)
19. Mayring, P.: Qualitative Inhaltsanalyse: Grundlagen und Techniken (in German). Beltz, Weinheim (2008)
20. Eisenhardt, K.M., Bourgeois, L.J.: Politics of Strategic Decision Making in High-Velocity Environments: Toward a Midrange Theory. Academy of Management Journal 31, 737–770 (1988)
21. Moore, G.C., Benbasat, I.: Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation. Information Systems Research 2, 192–222 (1991)
22. Scott, W.: Reliability of Content Analysis: The Case of Nominal Scale Coding. Public Opinion Quarterly 19, 321–325 (1955)
23. Mayring, P.: Qualitative Content Analysis. Forum: Qualitative Social Research 1, 1–10 (2000)
24. Moore, J.E.: One Road to Turnover: An Examination of Work Exhaustion in Technology Professionals. MIS Quarterly 24, 141–168 (2000)
25. Holsti, O.R.: Content Analysis for the Social Sciences and Humanities. Addison-Wesley, Reading (1969)
26. Charniak, E.: Tree-Bank Grammars. In: Proceedings of the National Conference on Artificial Intelligence, pp. 1031–1036 (1996)

27. Pavlou, P.A., Dimoka, A.: The Nature and Role of Feedback Text Comments in Online Marketplaces: Implications for Trust Building, Price Premiums, and Seller Differentiation. Information Systems Research 17, 392–414 (2006)
28. Frueh, W.: Inhaltsanalyse: Theorie und Praxis (in German). UVK, Konstanz (2007)
29. Hayes, A.F., Krippendorff, K.: Answering the Call for a Standard Reliability Measure for Coding Data. Communication Methods and Measures 1, 77–89 (2007)
30. Scott, W.: Reliability of Content Analysis: The Case of Nominal Scale Coding. Public Opinion Quarterly 19, 321–325 (1955)
31. Fleiss, J.L.: Measuring Nominal Scale Agreement Among Many Raters. Psychological Bulletin 76, 378–382 (1971)
32. Cohen, J.: A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 20, 37–46 (1960)
33. Aldenderfer, M., Blashfield, R.: Cluster Analysis. Beverly Hills: Sage University Paper (1984)
34. Fowlkes, E.B., Gnanadesikan, R., Kettenring, J.R.: Variable Selection in Clustering. Journal of Classification 5, 205–228 (1988)
35. Kaufman, L., Rousseeuw, P.J.: Finding Groups in Data: An Introduction to Cluster Analysis. Wiley Online Library (1990)
36. Howard, N., Harris, B.: A Hierarchical Grouping Routine, IBM 360/65 Fortran IV program. University of Pennsylvania Computer Center (1966)
37. Punj, G., Stewart, D.W.: Cluster Analysis in Marketing Research: Review and Suggestions for Application. Journal of Marketing Research 20, 134–148 (1983)
38. Ward, J.H.: Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association 58, 236–244 (1963)
39. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 13, 319–340 (1989)
40. Krippendorff, K.: Reliability in Content Analysis: Some Common Misconceptions and Recommendations. Human Communication Research 30, 411–433 (2004)
41. Thomas, H., Venkatraman, N.: Research on Strategic Groups: Progress and Prognosis. Journal of Management Studies 25, 537–555 (1988)
42. Ketchen, D.J., Shook, C.L.: The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique. Strategic Management Journal 17, 441–458 (1996)
43. Hair, J.F., Black, B., Babin, B.: Multivariate Data Analysis. Prentice Hall (2005)
44. Lambert, S.: Do We Need a “Real” Taxonomy of e-Business Models?. In: ACIS 2006 Proceedings, 17th Australasian Conference on Information Systems (2006)