

# Reviews of Cultural Artefacts: Towards a Schema for their Annotation

**Kristin Kutzner, Anna Moskvina, Kristina Petzold, Claudia Roßkopf, Ulrich Heid, Ralf Knackstedt**

Institute for Business Administration and Business Information Systems, Institute for Information Science and Language Technologies, Institute for Creative Writing and Literary Studies, Institute for Cultural Politics  
University of Hildesheim, Universitätsplatz 1, 31141 Hildesheim, Germany  
{kutznerk, moskvina, petzold, rosskopf, heidul, ralf.knackstedt}@uni-hildesheim.de

## Abstract

Digital transformation allows new forms of discussion on cultural and aesthetic practices. Several stakeholders are able to comment and discuss cultural artefacts (books, museums and their exhibitions). Today, a variety of platforms exists: from general platforms (e.g., Amazon, Tripadvisor) to specialized ones (e.g., LovelyBooks, Behance). So far, there has not been any analysis of reviews of cultural artefacts across platforms. Accordingly, this study identifies and classifies text components of reviews about cultural artefacts. Based on the coding paradigm in the sense of Grounded Theory, (1) the components are empirically identified and (2) structured, resulting in a first category system. After evaluating and modifying the system (3), a multi-layered category system resulted. Thereafter, a group of students applied the system to provide insights into the understandability of the system. By providing such a categorisation, we intend to contribute to further research in analysing the contents and text structure of online reviews.

**Keywords:** cultural artefacts, reviews, multi-layered category system

## 1. Introduction

Today, a variety of online platforms including e-commerce, social media and specialized rating platforms, allows consumers to rate and communicate their opinion about products or services. According to industry research reports, purchasing decisions of consumers are highly influenced by online reviews (Deloitte, 2007; Duan et al., 2008). Further, a Nielsen (2012) report, surveying more than 28.000 internet users worldwide, found that such online customer reviews are the second most trusted source of brand information.

Furthermore, digital transformation supports new or changed forms of collaboration (Kutzner et al., 2018) and participation (O'Reilly, 2005), including discussion on cultural and aesthetic practices. Through the interactive mode of digital media for instance the formerly clear lines between different groups of stakeholders are blurring, e.g. between producer and consumer (see the notion of 'Prosumer' in Toffler, 1980) and between laymen and professionals. Nearly every cultural artefact can get a review: a movie (e.g., imdb.com), a book (e.g., goodreads.com) or a new mobile application. In this study we focus on artistic artefacts (museums and their exhibitions) and books as cultural artefacts. Writing a review about such an artefact, a reviewer can choose among a variety of platforms, from general, commercial platforms (e.g., Amazon, Tripadvisor) and specialized platforms and community-based platforms (e.g., LovelyBooks, Behance) to more text oriented platforms (e.g., Sobooks, Mojoreads, Lector). Some of these platforms not only present the related cultural artefacts, but also support the interaction between several reviewers (comments function), the rating of cultural artefacts and its participative further development (co-creation). In this field, reviews about cultural artefacts can be seen as textual materializations of cultural practices and of their perception.

Earlier studies have tended to address heterogeneous and mostly isolated aspects concerning reviews. Most of them analysed reviews in English. For instance, based on product ratings, e-commerce platforms, especially Amazon (e.g.,

McAuley and Leskovec, 2013), analysed online consumer reviews. A few studies analysed reviews in German (e.g., Mehling et al., 2018). Computational and linguistic processing focused on techniques like opinion mining (e.g., Pang and Lee, 2008) or sentiment analysis (e.g., Wiegand and Ruppenhofer, 2015; Klinger et al., 2016) to investigate how reviews can be automatically classified into being positive, negative or neutral. However, to the best of our knowledge, there is no research approach analysing reviews, especially about cultural artefacts, across platforms. Therefore, this study focuses on reviews of cultural artefacts (books, museums and their exhibitions) in German from different types of platforms. Accordingly, we aim to answer the following research question as a basis for further research:

- *What kind of components are contained in reviews of cultural artefacts?*

Our contribution is a multi-layered category system for characterising components of reviews of cultural artefacts (e.g., contents and communicative acts). The category system contributes to our ongoing research on reviews in the digital world and aims at characterising and analysing reviews of cultural artefacts. Therefore, as a next step after building the category system, we hope to be able to find patterns of components within reviews. In a first step, we briefly outline the background of reviews, cultural artefacts and the digital world in which they are written (Section 2). Based on our research design (Section 3), we iteratively built the multi-layered category system (Section 4) and evaluated it, applying it several times (Section 5). We then discuss the results and future research directions (Section 6) and conclude with our main findings (Section 7).

## 2. Background

In this section, we specify the terms related to our research question. Therefore, we introduce our definition of a cultural artefact, the concept of a review and its components.

In this study, we define cultural artefacts as artistic artefacts (museums and their exhibitions) and books. This focus on the cultural field also defines the sort of review we are

looking for. In this context the concept of a review is characterized by its strong relation to the tradition of art and literary criticism as a professional journalistic form of discussing and reviewing newly published cultural artefacts.

In general, scholars and practitioners do not agree on a universal definition of the term review and of its components. Therefore, a review could be analysed and understood from different point of views. Some of the most prominent positions include the following statements:

- Review as a product of evaluation of a cultural artefact: a review is considered as an article published by a journalist, that describes, explains, interprets and/or evaluates a cultural artefact. Most characteristic parts of a review include recommendation or dissuasion as well as statements on the originality and entertaining qualities of an artefact (Stegert, 1997).
- Review as the central text type of literary criticism: a review is understood as the critical discussion of a new publication; the most common and most important type of text in literary criticism (Pfohlmann, 2005).
- Review as an expert expression of opinion: a review is the opinion-expressing form of literary and art criticism. Book reviews, film criticism, a judgmental report on a painting exhibition or an expert journalistic expression are examples (La Roche et al., 2013).
- Review means asking for terms of arts, their functions and their origin. It further involves judging, making oneself unpopular and not being afraid of misunderstandings (Rauterberg, 2007).

As the statements indicate, a review may contain several components, addressing several aspects of a cultural artefact. For instance, descriptions, explanations, interpretations of an artefact as well as its critical discussion and expressions of the reviewer’s opinion are components of a review. Consequently, a review is characterised by several textual components and, therefore, can be analysed from different perspectives. Furthermore, reviews and their components might vary depending on the type of platform (e.g., general platforms like Amazon and Tripadvisor, specialized and community-based platforms like LovelyBooks or Behance) and on the addressed cultural artefact (e.g., books, museums and their exhibitions). However, existing research mostly tended to address isolated aspects of reviews. Therefore, there still seems to be a need to investigate multiple perspectives on reviews of cultural artefacts. Accordingly, in this study we aim to provide a multi-layered category system that covers several perspectives on a review, as a basis for further research in this field.

### 3. Research Design

In order to identify components of reviews we conducted an iterative three-stage research design, containing build and evaluate activities that characterise the Design Science Research in Information Systems (e.g., March and Smith, 1995; Peffers et al., 2007). It consists of (Stage 1) the

identification of components and (Stage 2) the development of a category system (build activities). To leverage rigorousness and to demonstrate the utility, quality and efficacy of the category system, we evaluated the system (Stage 3, evaluate activities), applying it several times (Figure 1).

		Inputs	Methods/Steps	Outputs
Build	Stage 1: Identify components	• Sample reviews	• Perform open coding iteratively	• 130 types of review components
	Stage 2: Develop category system	• 130 types of review components	• Perform axial coding iteratively	• Multi-layered category system
Evaluate	Stage 3: Evaluate category system	• Sample reviews • Multi-layered category system	• Apply multi-layered category system iteratively	• Adapted multi-layered category system

Figure 1: Research design.

**Stage 1: Identify components.** First, we empirically derived components of reviews. To start this process, we selected ten sample reviews from different types of platforms (e.g., social media platforms, blogs, other rating and exchange platforms), addressing artistic artefacts and books. To contribute to the robustness, nine researchers independently analysed the reviews. Each researcher named segments of the reviews with short labels that characterise the components of the reviews. In the sense of Grounded Theory this procedure is called open coding (Glaser, 1978). It is a first step towards making analytic interpretations of the reviews (Charmaz, 2006). In a workshop, the researchers consolidated their components. In total, 130 different types of review components have been identified.

**Stage 2: Develop category system.** Second, the researchers independently structured and reassembled the identified components of reviews in a new way that is called axial coding in the sense of Grounded Theory (e.g., Strauss and Corbin, 1990, 1998), resulting in a multi-layered category system. The results were consolidated in a subsequent workshop. The category system has been enriched by addressing both review components of artistic artefacts and books. In addition, similar components have been merged.

**Stage 3: Evaluate category system.** Selecting further sample reviews of artistic artefacts and books, the researchers independently applied the category system and annotated the reviews. Again, the researchers consolidated their results and experiences while annotating the reviews. As a result, the category system has been modified. These steps have been repeated several times until no more changes occurred. In May 2018, 24 students applied the resulting category system, annotating about 430 randomly selected Amazon book reviews (McAuley et al., 2015; He and McAuley, 2016). As a step towards evaluation of the category system, we chose to analyse reviews of books from Amazon, due to its availability for scientific research, its large database, and the great variety of artefacts (different authors, different books) and reviewers. As the category system focuses on both reviews of artistic artefacts and books, it has to be applied to reviews of artistic artefacts as well. Therefore, the evaluation of May 2018 can be seen as a first step within an evaluation activity in progress. In the May 2018 exerciser, several reviews

have been independently analysed by three students. Subsequently, we measured the nominal scale agreement (Fleiss kappa, e.g., Fleiss, 1971; Fleiss and Cohen, 1973) and the absolute observed agreement of each category, to measure the agreement between the three raters. We further asked the students for their experience which categories they found easier to identify than others.

#### 4. Category System of Review Components

Following the research design (Section 3), we built a multi-layered category system, containing multiple components of a review (Stage 1 and 2). In this section, we describe the category system in general and, for reasons of space limitations, we only present selected components of Layer 1 and 2 in more detail.

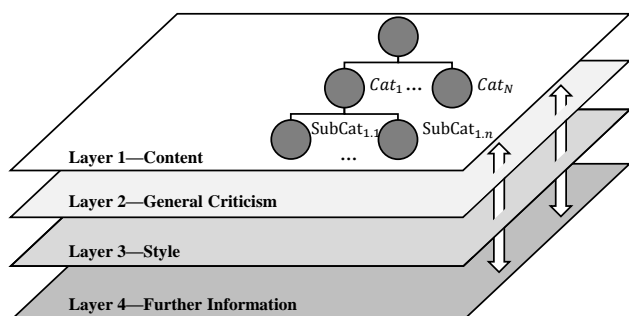


Figure 2: Multi-layered category system.

**Overview of the system.** The category system is divided into four different layers that distinguish several components of reviews: *Layer 1—Content* addresses various themes that are directly related to the cultural artefact (e.g., story of the artefact, the reviewer’s emotions and her or his assessment of the artefact). Furthermore, aspects of the background and of the reception context of the artefact are annotated as part of this layer (e.g., location where the artefact has been perceived, biography of the author, reflection about the structure and objective of the review). When a reviewer addresses the aspects of the artefact or its context by means of the text components classified on Layer 1, he or she may have certain intentions and/or comes up with criticism. To indicate these intentions *Layer 2—General Criticism* contains different communicative acts (e.g., summarise, recommend, discourage, thank, ask questions) that are always related to the components of Layer 1. In writing a review, the reviewers sometimes use particular language styles (e.g., use of rhetorical means like irony, hyperboles, and metaphors). *Layer 3—Style* contains these stylistic components. Moreover, sometimes multimedia content (e.g., pictures, links, emojis) is used in reviews. In addition, some sort of metadata (e.g., structure of the review) can be captured. *Layer 4—Further Information* addresses these aspects of a review.

The category system is used to annotate text strings in the reviews. Further, the categories are organized in a hierarchical structure, where each category has a numerical and a textual label used in the practical annotation work. For instance, Layer 1 contains category  $Cat_1$  that is divided into several subcategories from  $SubCat_{1,1}$  to  $SubCat_{1,n}$ . A subcategory can be divided into further subcategories and so on (Figure 2).

As a compact guideline for annotators, we built a tabular overview that complements the hierarchically ordered, labelled components by operationalised explanations and examples (codebook).

**Presentation of selected categories.** As described above, each layer consists of several hierarchically ordered categories. In total, the category system contains 108 different categories. More than half of the categories are assigned to Layer 1. On the topmost level, Layer 1 contains eight categories (Figure 3). For reasons of space limitations, we will only present selected subcategories of category 1.1 in more detail (Figure 4).

Layer 1—Content	
1.1	Reviewer-Artefact-Relation: Focused
1.2	Reviewer-Artefact-Relation: Extended
1.3	Artefact-Author-Relation and its Environment
1.4	Artefact in Medium 1-Artefact in Medium 2-Relation (Intermediality of the Artefacts)
1.5	Relation between Artefacts
1.6	Reflexions on the Review(ing Process)
1.7	Reviewers’ Self-thematisation
1.8	Relation between Reviews

Figure 3: Topmost categories of Layer 1.

Layer 1—Content	
1.1	Reviewer-Artefact-Relation: Focused
1.1.1	Individual aspects
1.1.1.1	Citation
1.1.1.2	Translation
1.1.1.3	Author
1.1.1.4	Content
1.1.1.5	Title
1.1.1.6	Physical Properties of the Artefact
1.1.1.7	Outer Appearance
1.1.1.8	Language Style
1.1.2	View of the Artefact as a Whole
...	...
Layer 2—General Criticism	
2	Communicative Acts in the Reviews
2.1	Summary
...	...
2.9	Mention without Assessment
2.10	Assessment
2.10.1	Positive/Agreement
2.10.2	Negative/Disagreement
2.10.3	Ambivalent
...	...

Figure 4: Selected categories of Layer 1 and 2.

Layer 1—Category 1.1. This category contains various themes addressed by the reviewers that are directly related to the cultural artefact. First, the reviewer might discuss individual aspects (category 1.1.1) related to the artefact. For instance, he or she might quote some content, address the translation, the author, detailed content or the title of the artefact. Also physical properties such as the quality of paper or illustrations, the outer appearance of the reviewed book (e.g., cover or used condition) and the language style of the artefact can be addressed by the reviewer. Alternatively or in addition, not only individual aspects, but also the artefact as a whole can be discussed (Figure 4).

Layer 2—Category 2. Addressing the categories of Layer 1, the reviewer always pursues some intention. Therefore, Layer 2 offers different communicative acts in the reviews for annotating these intentions (Figure 4).

## 5. Application and Evaluation

As a first test of the category system, 24 students were asked to apply it in the manual annotation of 430 randomly selected Amazon reviews (Stage 3 of the research design). We now present an example of an annotated text passage which illustrates how labels from our category system are attached to text components of a review. In addition, we present selected results of the evaluation of the category system.

**Annotation example.** Annotators of reviews might usefully read a given review several times, from different perspectives. As a result, each text passage of the review is annotated with categories of both, Layer 1 and 2. If it involves a particular language style or multimedia content, the categories of Layer 3 and 4 can be additionally annotated.

Original, German Text Passage	Free Translation
Ein sehr <u>charmant witziges</u> <b>1.1.2, 2.10.1</b> und <u>unterhaltsames Buch</u> <b>1.1.2, 2.10.1</b> ganz im <u>typischen Stil</u> <b>1.1.1.8, 2.9</b> Ellen DeGeneres. <b>1.1.1.3, 2.9</b>	A <u>very charming</u> <b>1.1.2, 2.10.1</b> and <u>entertaining book</u> <b>1.2, 2.10.1</b> in the <u>typical style of</u> <b>1.1.1.8, 2.9</b> Ellen DeGeneres. <b>1.1.1.3, 2.9</b>

Figure 5: Annotation example from a review text passage. For instance, we read a text passage from a sample review (for reasons of presentation, we translated it freely from German to English): «A very charming and entertaining book in the typical style of Ellen DeGeneres» (Amazon). First, we analyse the text passage from the perspective of Layer 1. The reviewer addresses the book in general (category 1.1.2), stating that it is «very charming and entertaining». Thereafter, the reviewer addresses the issue of the language style «the typical style» (category 1.1.1.8) and the author «Ellen DeGeneres» (category 1.1.1.3). In a second round, we read the text passage from the perspective of Layer 2. Writing that the book is very charming and entertaining, the reviewer wants to assess the artefact in a positive way (category 2.10.1). Both the «typical style» and «Ellen DeGeneres» are simply

mentioned by the reviewer (category 2.9). Third, the annotator should take the perspectives of Layer 3 and 4. However, as the sample text passage does not address these categories, no further annotation is required (Figure 5).

**Evaluation.** Applying the category system and annotating 430 reviews, the students had to annotate sentences or indicator expressions (such as those underlined in Figure 5) with the labels from our category system. In sum, 14.235 text passages have been identified and annotated with categories by the students. Some categories have been more frequently used than others. For instance, the content of an artefact (category 1.1.1.4) is most commonly recognized by the students (2603 times). The summary of some content is the second most common category (category 2.1) that is recognized by the students (2403 times, Figure 6).

Top 10—Frequencies of Categories (Total of 430 Reviews)		
#	No.	Category
2603	1.1.1.4	Content
2403	2.1	Summary
1369	2.10.1	Positive/Agreement
1297	1.1.2	View of the Artefact as a Whole
542	1.1.1.8	Language Style
510	1.1.4	Own Emotions
355	1.1.1.6	Physical Properties of the Artefact
319	1.1.1.3	Author
316	2.10.2	Negative/Disagreement
282	2.9	Mention without Assessment

Figure 6: Top 10—Frequencies of categories, total. Expecting 30 students to annotate the reviews, we divided the reviews equally among the students. To be able to measure the nominal scale agreement (Fleiss kappa) among the different raters (e.g., Fleiss, 1971; Fleiss and Cohen, 1973), each review has been assigned to three different students. However, only 24 students in the end worked on the task. As a consequence, not all reviews have been annotated three times.

Top 10—Frequencies of Categories (Triple of 139 Reviews)		
#	No.	Category
952	1.1.1.4	Content
862	2.1	Summary
516	2.10.1	Positive/Agreement
405	1.1.2	View of the Artefact as a Whole
206	1.1.1.8	Language Style
185	1.1.4	Own Emotions
123	1.1.1.6	Physical Properties of the Artefact
119	2.10.2	Negative/Disagreement
105	2.9	Mention without Assessment
91	1.1.1.7	Outer Appearance

Figure 7: Top 10—Frequencies of categories, threefold.

For analysing the agreement between three raters regarding each category, we reduced the data set to reviews that have been annotated threefold (Triple), resulting in 139 reviews and different overall frequency figures for the categories (Figure 7).

A calculation of the inter-annotator agreement on the task is not trivial and yet provides only limited insight. On the one hand, not all text passages have consistently been annotated by all annotators at both levels (Layers 1 and 2); this leads, for example, to low values of Fleiss' kappa. On the other hand, the category system contains a large number of categories, and many of them have rarely been used, so that the calculation does not provide adequately interpretable results. As a consequence, we get rather low kappa values (the same applies for Krippendorff's alpha for most categories, Krippendorff, 2011) which at first sight suggest little agreement between the annotators.

Nevertheless, we can interpret some of the data, calculating the absolute agreement for each category. For instance, regarding the most frequently used category, Content (1.1.1.4), in 48% of the cases the raters use this category unanimously. Furthermore, in 24% of the annotations, two raters agree and one rater disagrees (i.e., he or she does not use the category Content for annotating the same sentence). Only in 28% of the cases, two raters disagree and one agrees. Thus, in 72% of the cases the majority of the raters agrees on the use of the category Content (Figure 8). As described in Section 3, we also asked for the students' personal opinion about accurately assignable (vs. problematic) categories. The most commonly mentioned category is category Content: this confirms the calculated result. Thus, we can conclude that this category is understandable for the raters and easy to identify in the reviews. A similar interpretation applies to the use of category 1.1.2 (view of the artefact as a whole). Regarding the absolute agreement of this category, in 66% of the cases only one rater uses category 1.1.2. In only 7% of the cases, three raters agree on the use of this category (Figure 8). In the student's comments, category 1.1.2 is mentioned as not accurately assignable by the majority. Thus, both results indicate that category 1.1.2 is not understandable enough and hard to identify in the review texts and thus has to be improved.

		Relation of raters who agree and disagree (agree : disagree)		
No.	Category	3:0	2:1	1:2
1.1.1.4	Content	48%	24%	28%
1.1.2	View of the Artefact as a Whole	7%	27%	66%

Figure 8: Relation of raters who agree/disagree.

## 6. Discussion and Future Research Directions

As described in Section 5, the absolute observed agreement per category provides first insights into the comprehensibility of the categories. As a consequence, we are able to revise the category system. From the first annotation experiment we learn that positive evaluative statements are much more frequent in our sample than negative ones; many reviews make reference to the style or language of the reviewed book; and the physical appearance of the book is mentioned quite prominently,

likely because of the special situation of Amazon delivering the book. Finally, we note that a considerable number of annotated text passages make reference to the emotions of the reviewers during the reception process.

However, because of several circumstances (e.g., variety of incomplete annotations) the measurement of the nominal scale agreement (Fleiss kappa) could not provide sufficient results. As described above (Section 3), the category system is meant to be platform-independent and generic enough to be applicable to both, reviews of artistic artefacts and books. Thus, the described evaluation can only be seen as a partial evaluation activity in progress. The enhanced category system will be applied, in more controlled experiments, for reviews of artistic artefacts as well, resulting in new annotations of reviews and additional sample annotations to measure the agreement between raters.

Moreover, the category system represents an idealized schema for annotating reviews. It serves as a starting point to manually analyse the components of reviews. However, as a next step, we want to identify the review components automatically with methods of machine learning. For instance, the use of support vector machines (e.g., Cortes and Vapnik 1995; Boser et al., 1992) or decision trees (e.g., Rasoul and Landgrebe, 1991; Cho et al. 2002) can support the automatic classification of the components. The annotations that are manually identified by the students (Section 3) and/or in new annotation rounds, may serve as training data to learn a model and to predict components of reviews.

Furthermore, in the medium term, we not only want to identify components of reviews, but also patterns of and relationships between components (e.g., relationship between the view of the artefact as a whole and its assessment). Therefore, the training data can be used as an input for further machine learning algorithms, like cluster analysis (e.g., Hartigan and Wong, 1979; Elkan, 2003) or sequential pattern-mining algorithms like Apriori-based algorithms (e.g., Slimani and Lazzez, 2013). Moreover, to support the visualisation and analysis of reviews and their components, domain-specific modelling approaches (e.g., Guizzardi et al., 2002; Kishore and Sharman, 2004) can be valuable further research directions.

As a result, this study provides a double contribution to the field of Digital Humanities: First, the category system and the first sample annotations constitute a starting point for a classification of textual components of reviews, and for its subsequent automation by means of machine learning methods. Second, this study and the ongoing research project which it is a part of go beyond the methodological questions and address—by examining digital practices of cultural participation through reviews—the understanding of digital culture and society on a more fundamental level which is, eventually, a central question of Digital Humanities.

## 7. Conclusion

In order to identify *components that are contained in reviews of cultural artefacts*, we used the coding paradigm in the sense of Grounded Theory (e.g., Glaser, 1978; Strauss and Corbin, 1990, 1998). We thus empirically

derived components by analysing and annotating selected reviews. These components then have been structured, resulting in a multi-layered category system. To evaluate the system, we applied the categories, annotating several reviews and we modified the system up to a stable version. As a result, the system contains Layer 1—Content, Layer 2—General Criticism, Layer 3—Style and Layer 4—Further Information. Each layer consists of several categories that are organized in tree-like hierarchies. To further evaluate the category system, students applied the resulting system, annotating about 430 Amazon book reviews. The measurement of the absolute agreement of each category provides first insights into the comprehensibility of the categories. Thus, it supports revising the category system.

Overall, our findings contribute to the ongoing research on reviews in the digital world and to the analysis and characterisation of reviews of cultural artefacts. Based on the category system and on sample annotations, researchers are further able to identify the review components automatically with methods of machine learning and to discover patterns of review components.

## **8. Acknowledgements**

This research was conducted in the scope of the research project “Rez@Kultur” (01JKD1703), which is funded by the German Federal Ministry of Education and Research (BMBF). We would like to thank them for their support.

## 9. Bibliographical References

- Boser, B. E., Guyon, I. M. and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In Proceedings of the fifth annual workshop on Computational learning theory. ACM, pages 144–152.
- Charmaz, K. (2006). Constructing Grounded Theory. A Practical Guide Through Qualitative Analysis. SAGE.
- Cho, Y. H., Jae, K. K. and Soung, H. K. (2002). A personalized recommender system based on web usage mining and decision tree induction. *Expert systems with Applications*, (23:3):329–342.
- Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, (20:3):273–297.
- Deloitte (2007). New Deloitte Study Shows Inflection Point for Consumer Products Industry; Companies Must Learn to Compete in a More Transparent Age. Press Release, Deloitte Services LP, New York, October 1.
- Duan, W., Gu, B., Whinston, A. B. (2008). Do online reviews matter? An empirical investigation of panel data. In *Decision Support Systems*. 45., pages 1007–1016.
- Elkan, C. (2003). Using the Triangle Inequality to Accelerate k-Means. In Proceedings of the International Conference on Machine Learning, Washington DC, USA.
- Fleiss, J. L. (1971). Measuring Nominal Scale Agreement among many Raters. *Psychological Bulletin*, 76(5): 378–382.
- Fleiss, J. L. and Cohen, J. (1973). The Equivalence of weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. *Educational and Psychological Measurement*, 33:613–619.
- Glaser, B. G. (1978). Theoretical sensitivity. Mill Valley, CA: The Sociology Press.
- Guizzardi, G., Ferreira Pires, L. and Van Sinderen, M. J. (2002). On the role of domain ontologies in the design of domain-specific visual modeling languages. In Proceedings of the ACM OOPSLA.
- Hartigan, J. A. and Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society, Series C (Applied Statistics)* 28 (1):100–108.
- He, R. and McAuley, J. (2016). Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In Proceedings of the 25th international conference on world wide web, pages 507–517, InternationalWorldWideWeb Conferences Steering Committee.
- Kishore, R. and Scharman, R. (2004). Computational Ontologies and Information Systems. *Foundations. Communications of the Association for Information Systems*, 14 (8):158–183.
- Klinger, R., Suliya, S. S. and Reiter, N. (2016). Automatic Emotion Detection for Quantitative Literary Studies -- A case study based on Franz Kafka's „Das Schloss“ und „Amerika“. In *Digital Humanities*, Kraków, Poland.
- Krippendorff, K. (2011). Agreement and Information in the Reliability of Coding. *Communication Methods and Measures*, 5 (2):1–20.
- Kutzner, K., Schoormann, T. and Knackstedt, R. (2018). Digital Transformation in Information Systems Research: A Taxonomy-based Approach to Structure the Field. In *European Conference on Information Systems*, Portsmouth, UK.
- La Roche, W., Hooffacker, G. and Meier, K. (2013). Einführung in den praktischen Journalismus: Mit genauer Beschreibung aller Ausbildungswege Deutschland Österreich Schweiz. Springer-Verlag.
- March, S. T., Smith, G. (1995). Design and Natural Science Research on Information Technology. *Decision Support Systems*, 15 (4):251–266.
- McAuley, J. and Leskovec, J. (2013). Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text. In Proceedings of the 7th ACM Conference on Recommender Systems, pages 165–172, New York NY, USA, ACM.
- McAuley, J., Targett, C., Shi, Q. and Van Den Hengel, A. (2015). Image-based recommendations on styles and substitutes. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 43–52, ACM.
- Mehling, G., Kellermann, A., Kellermann, H., Rehfeldt, M. (2018). Leserrezensionen auf amazon.de : Eine teilautomatisierte inhaltsanalytische Studie. Bamberg, Bamberg University Press.
- Nielsen (2012). Nielsen's latest Global Trust in Advertising report. <https://retelur.files.wordpress.com/2007/10/global-trust-in-advertising-2012.pdf> (downloaded 2018-04-30).
- O'Reilly, T. (2005). What is the Web 2.0?, URL: <http://www.oreilly.com/pub/a/web2/archive/whatis-web-20.html> (downloaded 2018-06-15).
- Pföhlmann, O. (2005). Kleines Lexikon der Literaturkritik. Verlag LiteraturWissenschaft.de.
- Pang, B. and Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1–2):1–135.
- Peppers, K., Tuunanen, T., Rothenberger, M. and Chatterjee, S. (2007/2008). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24 (3):45–77.
- Rasoul, S. S. and Landgrebe, D. (1991). A survey of decision tree classifier methodology. In *IEEE transactions on systems, man, and cybernetics*, (21:3):660–674.
- Rauterberg, H. (2007). Und das ist Kunst?!: eine Qualitätsprüfung. Frankfurt am Main, S. Fischer Verlag.
- Slimani, T. and Lazzez A. (2013). Sequential Mining: Patterns and Algorithms Analysis. *International Journal of Computer and Electronics Research*, 2(5):639–647.
- Stegert, G. (1997). Die Rezension: Zur Beschreibung einer komplexen Textsorte. *Beiträge zur Fremdsprachenvermittlung*, 31:89–110.
- Strauss, A. and Corbin, J. (1990). Basics of qualitative research: Grounded theory procedures and techniques. Newbury Park, CA, Sage.
- Strauss, A. and Corbin, J. (1998). Basics of qualitative research: Grounded theory procedures and techniques. Thousand Oaks, CA, Sage, 2<sup>nd</sup> edition.
- Toffler, A. (1980). The Third Wave: The Revolution That Will Change Our Lives. London/New York, Collins.
- Wiegand, M. and Ruppenhofer, J. (2015). Opinion Holder and Target Extraction based on the Induction of Verbal Categories. In Proceedings of the 19th Conference on Computational Language Learning, pages 215–225, Beijing, China, Association for Computational Linguistics.