5-11-2023

A taxonomy for deriving business insights from user-generated content

Pablo Weingart
University of Kassel, weingart@uni-kassel.de

Thiemo Wambsganss
Swiss Federal Institute of Technology in Lausanne, thiemo.wambsganss@epfl.ch

Matthias Soellner
University of Kassel, soellner@uni-kassel.de

A TAXONOMY FOR DERIVING BUSINESS INSIGHTS FROM USER-GENERATED CONTENT

Research Paper

Pablo, Weingart, University of Kassel, Germany, weingart@uni-kassel.de.

Thiemo, Wambsganss, Swiss Federal Institute of Technology in Lausanne (EPFL), Switzerland, thiemo.wambsganss@epfl.ch.

Matthias, Söllner, University of Kassel, Germany, soellner@uni-kassel.de.

Abstract

Deriving business insights from user-generated content (UGC) is a widely investigated phenomenon in information systems (IS) research. Due to its unstructured nature and technical constraints, UGC is still underutilized as a data source in research and practice. Using recent advancements in machine learning research, especially large language models (LLMs), IS researchers can possibly derive these insights more effectively. To guide and further understand the usage of these techniques, we develop a taxonomy that provides an overview of business insights derived from UGC. The taxonomy helps both practitioners and researchers identify, design, compare and evaluate the use of UGC in this IS context. Finally, we showcase an LLM-supported demo application that derives novel business insights and apply the taxonomy to it. In doing so, we show exemplary how LLMs can be used to develop new or extend existing natural language processing (NLP) applications in the realm of IS.

Keywords: user-generated content, business insights, large language models, natural language processing.

1 Introduction

User-generated content (UGC) remains a widely investigated phenomenon in a variety of research disciplines such as information systems (IS, Lukyanenko et al., 2019), marketing research, political or social sciences (Timoshenko and Hauser, 2019). UGC can be generally described as information generated by end users outside of professional routines (Lukyanenko et al., 2019). It often contains important insights such as opinions, rankings, suggestions, personal experiences, or client requirements. Most of the time, this data is openly accessible online since it is included in social media posts, blog articles, tweets, product reviews, or survey replies that include images, audio, video, or text (Krumm et al. 2008). In business or organizational contexts, UGC provides an asset to develop competitive advantages. UGC is therefore analyzed and used to capture and explain a multitude of important business indicators and research constructs such as customer needs (Timoshenko and Hauser, 2019), purchase intentions (Agarwal, 2020) or user innovations (Safi and Yu, 2017). In the remainder of the work, we refer to these as “business insights”.

However, due to its unstructured nature, technical constraints and decentralized occurrence, UGC is still underutilized as a source of data for research and practice. The data is stored across multiple web platforms and comes in a variety of data formats (text, image, audio and video), with variable data quality, e.g., containing spam, spelling errors, unusual spellings, or multiple languages (Lukyanenko et al., 2019, 2017; Moens et al., 2014). Moreover, while UGC has been extensively studied in the past, these studies often do not use contemporary state-of-the-art natural language processing (NLP) techniques. For instance, dictionary-based sentiment analysis or unsupervised topic modeling approaches (e.g., latent dirichlet allocation) are used for data analysis, while newer transformer-based models for sentiment, emotion detection or topic classification are available. Even in recent This might lead to limitations in the solution and problem scope that can be investigated in UGC. For example, in earlier studies, as shown in a meta-analysis by Rietsche et al. (2019), researchers predict review helpfulness based on simple metrics such as number of words, sentences, or arguments.

Due to the rapid nature of improvements in the area of NLP and machine learning, studies can possibly derive insights more effectively by using more novel NLP techniques such as pre-trained transformer models, and, specifically, large language models (LLMs), e.g., OpenAI's GPT-3 (generative pre-trained transformer) and Codex or Google's LaMDA. For instance, Bražinskas et al. (2021) identify informative reviews for a specific product, summarize these reviews and then highlight the key benefits and drawbacks formulated in these reviews. With few-shot learning, dataset sizes can be kept relatively small, avoiding the costly annotation of large datasets. LLMs differ from pre-trained transformer models by their significantly larger model size and scale of the data that they were trained on. A major advantage of using LLMs is that researchers and practitioners can immediately test and build applications through prompt engineering without creating large-scale datasets in advance (Ye et al., 2021). Research on the use of LLMs in interdisciplinary research is scarce as many of these models are closed source, are often still in beta, and have only recently begun to be openly accessible via an API layer. Recently, open-source alternatives have been released. Combined with the API access, researchers can now use a variety of LLMs to test hypotheses and design IS with relative ease. Most studies applying these techniques to interdisciplinary problems are found in technical disciplines (e.g., Chintagunta et al., 2021). These papers are usually motivated by technical NLP and computational linguistic perspectives, missing out on behavioral or design-driven research aspects and seldom considering a socio-technical perspective on the research phenomena, especially the social dimensions (e.g., the user or structural context) of applying a certain technology (Bostrom and Heinen, 1977). The large amount of individual IS studies on the automated extraction of UGC provide this socio-technical perspective, but, importantly, also provide a benchmark to which studies using these newer techniques can be compared to. By aggregating them, they provide an overview of insights in the form of important qualitative and quantitative attributes of UGC that define and motivate the usage of UGC within an IS context (e.g., data sources as well as input and output variables). In this regard, IS research can provide a promising perspective to apply this lens and classify the usage and extraction of UGC into its relevant dimensions and characteristics. For example, applying the socio-technical system lens allows to focus not only on the technical embedding (e.g., the NLP technology applied in a certain task), but also on the context the users operated in when creating the UGC. This nuanced perspective on dimensions and characteristics can ultimately yield different configurations of use cases and outcomes for different stakeholders (Bostrom and Heinen, 1977). Also, the mentioned shortcomings of current UGC analysis approaches, compared to the illustrated potentials, highlight the necessity to better understand the application types, capabilities, and potential implications of these techniques for IS research. To the best of our knowledge, no such structure on UGC exists for IS research yet. Therefore, a consistent knowledge aggregation and literature overview on insights, which novel techniques like LLMs can potentially derive, predict or aggregate, will help both researchers and practitioners to systematically design, compare, and evaluate the use and impact of LLMs for new or existing NLP applications or IS research streams. Consequently, our paper contributes to the fields of IS and NLP by answering the following research question (RQ):

RQ: *What are dimensions and characteristics for the automated extraction of business insights from user-generated content in an IS context?*

In order to contribute to our research question, we conduct a systematic literature review as proposed by Webster and Watson (2002) and vom Brocke et al. (2015). Based on this sample of 215 publications, we develop a taxonomy that helps both practitioners and researchers identify, design, compare and evaluate the use of textual UGC within an IS to derive business insights. For example, researchers could use this taxonomy to evaluate the current technological state of an IS and the varieties of UGC it uses with respect to deriving business insights. The taxonomy was developed in an iterative taxonomy development process according to Nickerson et al. (2013) and resulted in 6 distinct dimensions and 30 characteristics. Finally, we provide an example of an LLM-supported IS which derives business insights in the form of product suggestions from Amazon and Google Play reviews. By providing this example, we show how the taxonomy can be applied by fellow researchers and practitioners. Furthermore, we provide an early showcase how LLMs can be used to develop new or extend existing NLP applications or research streams in the realm of IS.

2 Conceptual background and related work

2.1 Advanced deep learning techniques

In this section we briefly go over NLP in general to provide a common understanding. Afterwards, we describe the current state-of-the-art deep learning techniques that motivate our work.

NLP is an interdisciplinary field of computer science, machine learning, artificial intelligence, and computational linguistics that deals with making human language accessible to computers. NLP also plays a role in emerging fields like digital humanities and computational social sciences (Eisenstein, 2019). NLP intersects with the IS field by analyzing the interplay between technology and social science (Bostrom and Heinen, 1977). Text mining is a related term that is typically less concerned with linguistic structure and more interested in fast algorithms compared to NLP (Eisenstein, 2019). NLP is relevant for the analysis of UGC because data from UGC is inherently unstructured as well as large in quantity, making manual analysis unfeasible. In IS research as well as other interdisciplinary fields, common NLP tasks to make sense of UGC include sentiment analysis, information retrieval and semantic annotation (Kang et al., 2020; Liu et al., 2017). To perform these tasks, statistic algorithms are used. There are multiple NLP reviews which indicate that machine learning algorithms have been increasingly in use, while deep learning approaches were not yet present (Kang et al., 2020; Liu et al., 2017). More recent reviews on the use of NLP do however indicate an uptick in use of deep learning techniques, especially language models (Li et al., 2022). Language models are statistical models that are used to predict the probability of a sequence of words. They are commonly used to analyze natural language text. Language models are usually pre-trained on a large, general dataset and then fine-tuned to perform downstream tasks such as autocompleting code (Xu et al., 2022). Language models are starting to be used in IS as well: for example, Au et al. (2022) use transformer models to perform an analysis on an organization's strategic strengths, weaknesses, opportunities and threats (SWOT) based on data from a text corpus. Wambsganss et al. (2021) designed an IS that provides feedback on argumentation in writing exercises for students. By using models pretrained with few-shot learning in mind, researchers can finetune the models on smaller datasets, lowering the amount of annotation work that is needed (Ye et al., 2021). The ongoing trend is that language models are growing ever larger. The model size can be measured by the number of parameters a model has. In 2018, BERT-Large was introduced with 340M parameters. In 2020, the model GPT-3 was released to selected researchers with a size of 175B parameters. Current models can scale up to a trillion parameters with mixture-of-experts models (Fedus et al., 2021). As outlined in the introduction, we refer to such large models as LLMs. Research on the use of LLMs is rare as these models were not easily accessible to researchers in the past. Recently however, more and more open source LLMs have been released to the public. Furthermore, proprietary models like GPT-3 have lifted restrictions and are now publicly available through an API. Since larger models often perform better and can solve more complex tasks, compared to smaller ones, we want to motivate their use to the IS community by showing potential applications on UGC based on IS research as well as showcase an IS prototype that derives novel insights from such UGC.

2.2 UGC and its existing classifications

UGC can be generally described as information produced by members of the general public (Lukyanenko et al., 2019). According to Naab and Sehl (2017) there are three criteria that differentiate UGC from other types of content. First, UGC is characterized by the contribution of the content through the user. Simply forwarding or copying existing content is not sufficient. The contribution can be made within the context of an existing framework such as news article comments or creating individual content such as writing a blog article on a specific topic. Second, the content must be published in order to be accessible to the public. Finally, UGC is created outside of a professional context and is thus usually unpaid and unrelated to a specific organization (Lukyanenko et al., 2019).

UGC encompasses text, video, image and audio content. In the past, researchers have mainly focused on text content due to technical constraints. With the advancement of deep learning techniques other types of content can be analyzed better. For example, with the proliferation of speech-to-text software like Whisper by OpenAI, audio content can be transcribed to text reliably and make audio content accessible to NLP algorithms relying on text. In consequence, our work also uses the NLP lens that is evident in research. Some types of UGC are further differentiated by researchers. Lukyanenko et al. (2019) describe crowdsourced UGC which is gathered from custom IS that collect specific data from contributors external to the organization, e.g., reporting local issues like cars occupying bike lanes. This also includes content from online user innovation communities, e.g., Salesforce's IdeaExchange.

In IS research, literature reviews and other grounding papers on UGC have focused on specific types or aspects of UGC. For example, one widely investigated type of UGC are online reviews from platforms such as Amazon. Donthu et al. (2021) define the primary topics of research on social networks and online reviews within the context of electronic word-of-mouth. Other studies take a more detailed approach like Rietsche et al. (2019) who aggregated studies to present key determinants of online review helpfulness. Lukyanenko et al. (2019) highlight the effects of data collection design choices on the quality of crowdsourced UGC. Other IS studies look at the use of NLP to extract information from textual data generally and cover studies on UGC. For example, Liu et al. (2017) and Kang et al. (2020) provide a general overview on the use of NLP in IS and management research. In a similar vein, Landolt et al. (2021) take a more technical perspective and create a taxonomy on deep learning techniques. Further, Weingart et al. (2022) analyze the potentials of leveraging textual data, including UGC, to automatically extract arguments with NLP techniques. With their lack of focus on UGC, these studies do not yield sufficient insights regarding the analysis of the content to generate (business) insights. Going beyond IS, there exist several studies analyzing the use of UGC for a specific research discipline, including journalism (Jönsson and Örnebring, 2011; Naab and Sehl, 2017), marketing (Nguyen and Tripathi, 2019; Soylemez, 2021) or e-commerce (Li et al., 2022). For their respective domain, these studies identify UGC dimensions and characteristics on the techniques used to analyze UGC, the use cases UGC enables and its data sources (e.g., Amazon) and types, but do not aggregate them in an integrative way. While useful guiding lenses for our taxonomy, these studies do not cover the mentioned socio-technical lens unique to IS. Furthermore, as our systematic literature review indicated, IS researchers have developed unique use cases that the studies do not cover. Although there is a large body of IS research on applications building upon UGC, we have found no recent study that took a holistic perspective on UGC during our extensive systematic literature review. To the best of our knowledge, there exists no review on UGC that covers the novel lens we want to apply: evaluating the use of textual UGC within IS to derive business insights to help both researchers and practitioners identify, design, compare and evaluate the use of UGC for their own projects. Notably, recent technological advancements surrounding LLMs, specifically GPT-models, enable the development of novel use cases for deriving business insights from textual UGC for both researchers and practitioners. For this purpose, IS researchers have already developed many studies that provide a solid foundation to build upon. Taxonomies facilitate knowledge sharing and provide insight into the relationships between objects, which can help with decision-making (Bailey 1994; Usman et al. 2017) on the use of textual UGC. In consequence, our taxonomy of these IS studies can offer a unique perspective on how UGC has been utilized to generate business insights in the past, identify corresponding gaps and help compare newer LLM-driven studies to contemporary IS literature, based on the dimensions and characteristics we derived from the literature. Finally, with our exemplary application of the taxonomy, we want to approach people designing data-driven products or projects that aim to develop business insights on a specific topic and show them the potential of combining UGC with LLMs.

3 Research Methodology

Our research methodology is split into four steps. First, we performed a systematic literature review based on Webster and Watson (2002) and vom Brocke et al. (2015). Based on the empirical sample we then derived the taxonomy's dimensions and characteristics aligned with the research question. The creation of the taxonomy was guided by the method by Nickerson et al. (2013). In the third step, we evaluated the taxonomy and its quality through interviews with two senior researchers and one practitioner. Finally, to guide the usage of the taxonomy for research and practice, we apply the taxonomy to an IS that we developed to generate business insights from UGC. Next, we describe each step in greater detail.

3.1 Systematic literature review

In order to gather a representative coverage of literature on text-based use cases in IS literature, we conduct a systematic literature review according to the principles of Webster and Watson (2002) and vom Brocke et al. (2015). We align the search process to the process, source, coverage, and technique dimensions. We perform the search process sequentially using the databases of the AIS Senior Scholars' Basket of Eight as the source for dates up to September 2022. Furthermore, we add the ICIS and ECIS conference proceedings to include more novel research. To reach a representative coverage we employ TIKEAB (title, keyword, and abstract) search and comprehensive forward and backward search as techniques. In the following, we describe the steps we took to come to the final set of relevant literature. Based on existing literature reviews (Li et al., 2022; Liu et al., 2017) we identified different keywords and synonyms that researchers use when presenting work on generating business insights from UGC. We iteratively refined our search string during the research when performing early analysis of the sampled studies. This resulted in the following search string: ("natural language" or "NLP" or "text mining" or "user-generated content" or "UGC" or "online review" or "blog" or "ewom" or "electronic word of mouth" or "social media" or "social network"). We used all variations of the keywords such as singular, plural, with, or without a hyphen to generate further input. Although the search string is broad, through employing TIKEAB and outlet filters the amount of hits was constrained to a reasonable amount. Table 1 summarizes the database hits and number of hits we selected for the review.

Database	AisEL	Ebsco	ProQuest	ScienceDirect	Total
Total hits	1261	425	38	37	1761
Relevant after screening abstract	284	74	3	3	364
Relevant after detailed analysis	132	41	1	1	175
Forward/backward search results	35	5	0	0	40
All results	167	46	1	1	215

Table 1. Database hit summarization.

Selection of papers: The outlet-based search yielded 1761 hits. We screened the papers based on their abstracts and, if unclear, text. Aligned to our research question, the primary inclusion criterion was that some sort of NLP algorithm was used to generate (business) insights based on UGC. This criterion is necessary because AM strictly operates on argumentative text. Hence, we only included papers that aim at solving a practical text-related task, e.g., predicting a review's helpfulness based on characteristics of the text. We found 175 relevant papers according to our criteria. Performing forward and backward search, we went through all the references of each paper. We again filtered the references by database, as specified in the previous step, and identified 40 additional papers. This is shown in Table 1 for each database. If a paper was found multiple times during forward and backward search, only the first occurrence was counted.

Analysis of papers: The 215 relevant papers were analyzed from a concept-centered perspective. We aimed to identify IS that use NLP to derive business insights from UGC. The concept matrix was initially based on similar classifications from existing research. The concepts were usually directly mentioned in the study, e.g., which NLP method was used. We also created various insight concepts since they were not evident from existing literature. The codings were performed by two researchers independently in multiple iterations. To solve differences in coding, the two coders met with a third researcher to discuss the codings and come to a final agreement.

3.2 Taxonomy development

The aim of the paper is to create a novel artifact that helps both practitioners and researchers identify, design, compare and evaluate the use of UGC within an IS to generate business insights. For this purpose, we developed the taxonomy according to the method outlined by Nickerson et al. (2013) because it is successfully used and well-understood within the field of IS. The process is split into different steps. With the first step, the meta-characteristic is defined. The meta-characteristic guides the creation of the taxonomy in general and its dimensions and characteristics specifically. The characteristics describe the objects in a specific domain of interest. To emphasize the focus on IS research, our objects are either the IS itself or the components of an IS. This distinction is necessary because some papers in our sample used multiple data sources to develop insights, breaking the mutual exclusiveness rule of the taxonomy characteristics. The domain of interest is then the use of UGC in IS to generate business insights. For the purpose of our work, a component refers either to an IS as a whole, if possible, or a part of the IS that leverages an NLP technique to derive business insights based on UGC.

For example, Hu et al. (2018) model brand personality based on multiple data sources including employee reviews and brand-related tweets. Each of these represent a unique component that is classified on its own. We specify the meta-characteristic of our taxonomy as the following: “we aim to create a novel artifact that helps both practitioners and researchers identify, design, compare and evaluate the use of textual UGC within an IS to generate business insights.” We especially want to support practitioners developing data-driven projects that aim to develop business insights. The taxonomy focuses on textual UGC since text is the primary object of investigation in the literature we covered. Also, spoken dialogue can be transcribed to text and can then be analyzed with NLP.

There are two approaches that describe an iteration within the taxonomy development process. In the empirical-to-conceptual approach, a subset of relevant objects is identified in line with the meta-characteristic. These objects are then grouped to characteristics and finally dimensions. In the conceptual-to-empirical approach, the researcher’s existing knowledge and intuition is used to conceptualize similarities between potential objects. To determine when to end the iterative taxonomy development process, Nickerson et al. (2013) define subjective and objective ending conditions. We tracked that the following ending conditions were fulfilled:

1. All objects or a representative sample of objects have been examined.
2. No new dimensions or characteristics were added, split or merged in the last iteration.
3. Every dimension and characteristic is unique within its dimension.
4. At least one object is classified under every characteristic of every dimension.

In Table 2, we visualize the iterations that we’ve gone through before finishing in iteration 4 when all ending conditions were fulfilled. In iteration 1 we have mostly used the empirical-to-conceptual approach because there is a rich literature base on UGC available, presenting objects that can be classified as well as multiple classifications that already outlined suitable characteristics, at least to some extent. With our focus on generating business insights as well as the introduction of large language models, we also used the conceptual-to-empirical approach in iteration 1 to provide the unique lens (that addresses the research gap we outlined) which is not already found in the literature. Proceeding to iteration 4, we used the empirical approach and were able to classify all relevant IS components within the taxonomy. We finished with 6 dimensions: “insight level”, “insight measure”, “user type”, “data source”, “content type” and “NLP technique”. The dimensions are further specified in a later section.

Iteration	Approach	Taxonomy	EC met
1	Both	T_1 = {**research metric** (review quality, product demand, ...), **data source** (e-commerce, social network, QA platforms, innovation communities, …, blogs), **user type** (learner, discussion participant, analyst, blogger, customer, employee, crowdworker), **data type** (online review, social media post, discussion post, article), **NLP technique** (topic modelling, linear regression, …, non-transformer neural network models, transformer models, large language models)}	4
2	Empirical-to-conceptual	T_2 = {**insight area** (innovation, knowledge, company, competitor, brand, user, post, review, service, product, …, other), **insight measure** (sentiment, quality, persuasiveness, demand, …, other), **user type** (discusser, analyst, blogger, customer, employee, crowdworker, other), **data source** (review platform, social network, discussion community, crowdsourcing platform, other), **content type** (review, post, article, other), **NLP technique** (traditional statistics, non-transformer neural network models, transformer models, large language models, other)}	4
3	Empirical-to-conceptual	T_3 = {**insight level** (organization, brand, product, user, content), **insight measure** (quality, sentiment, popularity, similarity, personality, textual), **user type** (expert, consumer, crowdworker, employee, other), **data source** (*same*), **content type** (*same*), **NLP technique** (*same*)}	3, 4
4	Empirical-to-conceptual	T_4 = {**insight level** (*same*), **insight measure** (*same*), **user type** (*same*), **data source** (*same*), **content type** (*same*), **NLP technique** (*same*)}	1, 2, 3, 4

Table 2. Summary of taxonomy iterations.

Notably, we've added an *other* characteristic for three of our dimensions. As explained, the taxonomy is also aimed at practitioners which may not be familiar with IS. In the work, we exclusively focused on empirical studies in IS. Naturally, some issues which may be present in practice have not been covered in IS research yet. Furthermore, we have only looked at a subset of IS literature with the restriction to the mentioned outlets. Without an *other* characteristic it is not feasible to cover all these cases that might occur and invalidate the use of the taxonomy otherwise. In fact, the other characteristic can be used to continuously iterate on the taxonomy. We do want to emphasize however that, for the paper sample we analyzed, the *other* characteristics can be omitted.

While we do not present extended detail to the taxonomy development throughout its iterations due to space constraints, we want to highlight the key steps of deriving the initial dimensions and characteristics of the taxonomy by means of an exemplary paper by Hu et al. (2018). To derive business insights that are analyzed in the sample, we extracted the metrics that were stated in the research design, in the exemplary case brand personality. Analyzing the description of the data used in the study, we can quickly derive that profile self-descriptions, and more generally, social media posts (data type) from Twitter are analyzed, a social network (data source). These posts are written by customers (user type) of a brand. The authors further extend their data with employee (user type) reviews (data type) from GlassDoor which was also classified as a social network based on the definition in the paper (data source). As outlined previously in this section, since the paper uses two separate instances of UGC, they are separately classified in our taxonomy. Finally, the study leverages elastic-net regression which we initially classified under the linear regression data type characteristic.

3.3 Taxonomy evaluation

With the (preliminary) taxonomy being completed, the evaluation of the taxonomy is necessary to assess and ensure its usefulness. Nickerson et al. (2013) define five attributes for a useful taxonomy: conciseness, robustness, comprehensibility, and explanatory power. To evaluate our taxonomy according to these attributes, we performed semi-structured interviews. To cover both research and practical perspectives, we interviewed two senior researchers with a background in both NLP and UGC research in the context of IS as well as one practitioner who has expertise in the use of NLP and UGC to derive business insights. Our interview guideline consisted of fourteen open-ended questions aligned to the mentioned five attributes. The guideline was based on the taxonomy evaluation suggestions of Szopinski et al. (2019). We sent the taxonomy as well as the meta-characteristic to the interviewees for iterations two and three. Furthermore, to gain more insights on actual use, we asked them to apply the taxonomy to IS they have created in the past, if applicable. We also asked for general comments and improvements based on their experience with applying the taxonomy. The responses were generally favorable, and the taxonomy was simple to understand and comprehensive enough to apply to their own IS. However, the researchers noted the existence of the *other* characteristics that we included to cover diverse practitioner use that may not be present in the sampled IS studies. With the researcher and practitioner feedback, we were able to streamline the dimensions and characteristics to strike a reasonable balance with regards to the abstraction level, according to the interviews. Incorporating the feedback for example, we provided less detail on older NLP techniques for analyzing UGC which researchers deemed less important with the paradigm shift to transformer models and LLMs. This was confirmed through related research literature (e.g., Weingart et al., 2022). Also, they expressed less familiarity with some parts of the taxonomy, especially large language models and noted that, due to the rapid development of the technology, a special emphasis on the extensibility of the taxonomy may be necessary. With the feedback we incorporated, they noted the suitable abstraction level of the taxonomy that should be able to cover many of these extensions. Finally, they also suggested changes regarding imprecise wording which we subsequently iterated upon.

3.4 Taxonomy application

The usefulness of a taxonomy is determined by its actual use by the targeted users. While the interviews during the taxonomy evaluation were useful for making sure the taxonomy fulfills the quality criteria by Nickerson et al. (2013), the usefulness remains vague as its use has not been studied. We want to guide this usage by looking at a specific example and use case that illustrates how the taxonomy can be applied to an IS in section 6. This IS represents a technical demo that uses UGC in the form of online reviews to derive pain points and suggestions to improve a product and mobile app. To reduce bias, we opted to showcase a demo that was designed by one of the authors before formulating the taxonomy. Thus, the demo was not customized to fit the taxonomy in any way.

4 Taxonomy on business insights from UGC

In this section, we present the current state of the taxonomy after performing four iterations and its revisal with the feedback we received during its evaluation. With the taxonomy, we aim to create a novel artifact that helps both practitioners and researchers to identify, design, compare and evaluate the use of UGC within an IS to generate business insights. In total, we classified IS components from 215 papers to the taxonomy. The taxonomy focuses on text UGC since text is the primary object of investigation in the literature we covered. Also, spoken dialogue can be transcribed to text and can then be analyzed with NLP. The resulting dimensions and characteristics can be seen in Figure 1.

As already specified in the section on taxonomy development, a component refers either to an IS as a whole or a part of the IS that leverages an NLP technique to derive business insights based on UGC. For example, Hu et al. (2018) model brand personality based on data sources including employee reviews and brand-related tweets. Each of these represent a unique component that is individually classified.

Next, we describe each dimension and describe its characteristics. We put special emphasis on the dimensions insight level and insight measure as they are the most relevant to our research question on deriving business insights. The remaining dimensions user type, content type, data source and NLP technique were less emphasized because they have been based partly on existing UGC literature and are thus less novel. Besides being necessary for comprehensiveness, we have also applied the unique lens on deriving business insights with contemporary NLP techniques and data sources. This adds necessary context. As outlined in the section on taxonomy development, we have added an “other” characteristic for most of the dimensions to reflect taxonomy objects that have not been part of the literature sample, but may be relevant in practice. Since we cannot provide sufficient clarity on such cases, we do not cover these characteristics further in the remainder of the section.

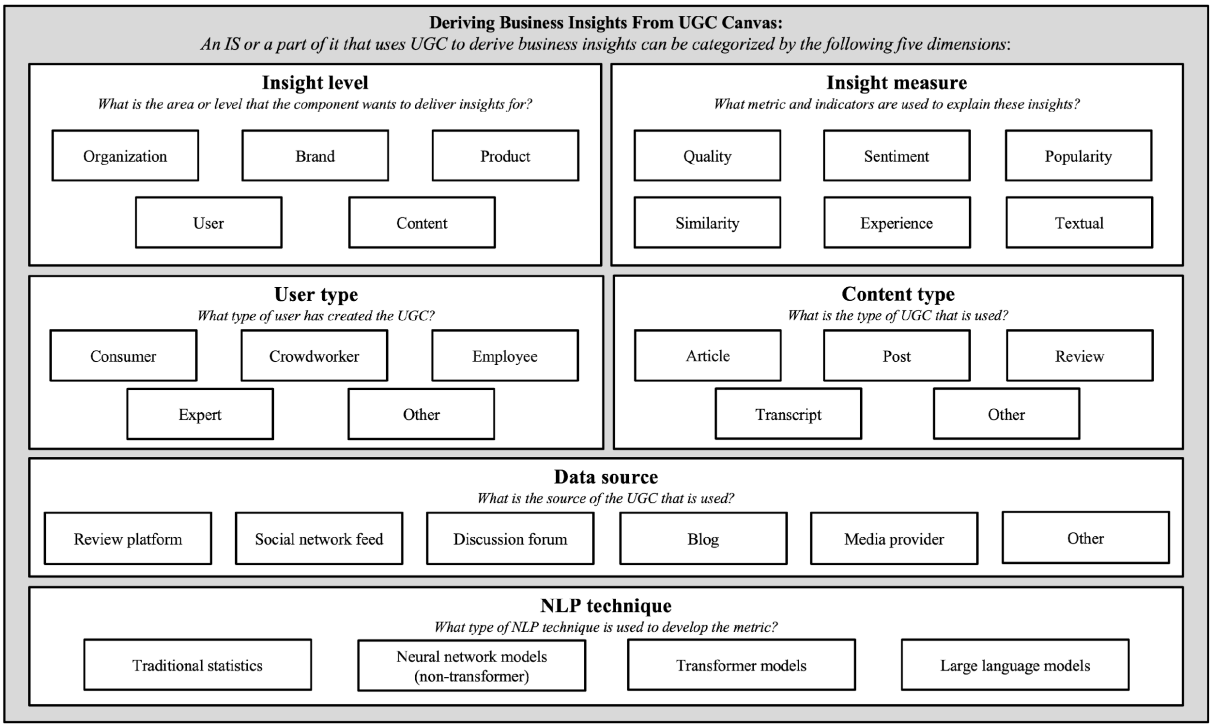

Figure 1. Deriving Business Insights from UGC Canvas.

Insight measure: The insight measure describes the metric, variable, or indicator that IS researchers apply to UGC to explain and quantify business insights. Although defining the insight measure is not the first logical step when applying our taxonomy, we put this description first as example measures provide needed context for the next paragraph on the insight level dimension. These measures relate to quality, sentiment, popularity, similarity, and experience. The first dimension refers to insight measures that assess the quality of the analyzed UGC, e.g., review helpfulness, idea quality, or post persuasiveness. With sentiment measures, insights refer to emotions related to the insight level, e.g., product features. Popularity measures provide information on how prominently and often an insight level is featured within the analyzed UGC, e.g., consumer interest regarding products or brands or sales performance. Similarity measures enable the comparison between different elements within an insight level, e.g., substitution products. Experience refers to insights on the behavior of users with content of the respective insight level. For example, purchase decisions with respect to a product or retention behavior on social networks. Finally, the *textual* measure refers to all qualitative insights that are extracted from UGC. In contrast to the other measures, these are insights that are available in text form, e.g., review suggestions aggregated from online reviews, like how a human would present insights. These measures represent the paradigm shift that we outlined in the introduction with greater capacity for advanced deep learning techniques. Currently, such approaches have not been well studied in research. With further work, we expect to be able to differentiate and structure this measure better.

Insight level: the insight level represents the area and purpose for which business insights shall be derived. The dominant levels found in the literature are organization, brand, product, user, and content. The levels are descendingly ordered in a way indicating the breadth of the unit of analysis and the necessity to aggregate UGC to come to useful conclusions. The product, brand, and organization levels broadly align to research on business intelligence. These levels were not merged to allow a more granular use of our taxonomy, especially by practitioners to reflect the differentiation of these three levels that is evident in IS literature. Within the organization level, business insights are derived that relate to the organization in the context of the market it operates. It includes metrics related to stock performance (Zhang et al., 2021) and the organization's strategic positioning, e.g., its strengths and weaknesses (Au et al., 2022). The brand level refers to a firm's brand and represents metrics like brand engagement, brand reputation, or brand personality (Hu et al., 2018). Many studies in IS literature also analyzed competitor organizations in varying capacities. The product level represents business insights related to an organization's product or service. In the literature, we found metrics like product sentiments, product popularity, or product demand (Karanam et al., 2021). Furthermore, we also included insights on innovation that are used to drive the development of said products or services, e.g., idea quality (Ye et al., 2016). These are classified into these three levels based on their focus on product, brand, or organization. The user level represents insights related to a user or a group of users. This includes more general measures such as the information disclosure level (Wei et al., 2020) but also customer-related ones such as customer agility (Zhou et al., 2018) or social proof (Khurana et al., 2019). For the latter, it is necessary for users of the taxonomy to clearly differentiate between the user and other levels. For example, we found a measure concerning customer needs (Stahlmann et al., 2022) which are strongly associated with the product and therefore better classified there. Finally, the content level describes insights that are related to the written content, often on an individual level not immediately related to the organization. Example measures commonly found in the literature are a review's helpfulness (Wang et al., 2011), the quality of a post (Singh et al., 2014), or post persuasiveness (Shin et al., 2020). While this level does not intrinsically contain business insights, content approaches may still provide important explanations for the insights and measures of the aforementioned levels.

User type: this dimension defines the type of user whose content is being analyzed to derive business insights on a certain insight level. We specified five dominant user types from IS research: consumer, crowdworker, employee and expert. The choice of user type often imposes constraints on available content types that should be analyzed. For instance, consumer UGC is mostly available in the form of online reviews and social network posts. Consumers are users that use products or services of a company and may have arbitrary experiences regarding the use. Furthermore, this type also refers to users that interact with other, mostly user-generated, content. Crowdworkers are users that provide UGC in crowdsourcing communities, especially in (product) innovation formats. Employees usually do not create UGC per the definition of Naab and Sehl (2017) when they are creating content for which they are paid. However, for IS studies that leverage data from internal IS of an organization, whose use is voluntary, the concept of UGC has been used. Furthermore, employees create UGC when providing feedback on their employer on websites like Glassdoor. Finally, experts are users with expertise about certain topics that usually have a following of consumers on their respective content platforms. We divide experts into two subtypes: analysts and influencers. Analysts are users that provide longer-form (text) content on specific topics, usually in article form on blogs. As expected, influencers generate content that may promote certain products or services. We also include content creators with this subtype. Unlike the analyst subtype, they focus on providing video and audio content, e.g., podcasts whose transcripts can be analyzed. We chose to name this subtype influencer instead of the more intuitive content creator to avoid naming confusion.

Content type: with this dimension, types of text-related UGC are differentiated that are used to derive business insights. These types illustrate in what form the content is structured and presented through the user. The (online) review type is well known in IS research and represents any UGC that has a reviewing character, most commonly assessing a product or service on review platforms such as Amazon, Google Maps, Yelp, etc. The post type represents UGC texts that are embedded within an IS and often used to interact with other users and content. This includes questions and answers, posts on social networks (e.g., Tweets on Twitter), comments on other people's posts, articles or reviews, and more. In contrast, the article type refers to longer and free-form text that is similar to essays and covers a wide variety of topics. These articles may be embedded in IS such as blogs or discussion websites (e.g., SeekingAlpha) but are not strictly dependent on the use of a specific IS. For example, one article can be presented on social networks, newsletters, print media, etc. Finally, transcripts describe text that has been transcribed from spoken language. Transcripts include monologues on a specific topic of interest as well as interviews or dialogues between multiple people. This content type is relatively novel due to the recent advancement of speech-to-text technologies and is not well understood in literature yet. Hence, further adjustments to differentiate this transcript are necessary in the future.

Data source: the data source dimension represents commonly used (text) sources of data to which NLP techniques are applied. It is necessary to take a granular approach when classifying UGC data sources because platforms sometimes provide diverse functionality, resulting in unclear classification boundaries. For example, Facebook is mainly a social network that also provides the ability to review companies on Facebook Page Ratings. Furthermore, with Facebook's Groups feature, users can discuss various topics of interest. Hence, to ensure a clear boundary between characteristics, each of these components would be put into its own respective characteristic according to the definition of this dimension. A review platform is a platform on which review content is provided. Popular review platforms among our analyzed sample include sections on Amazon, Yelp and TripAdvisor as well as its Chinese counterparts. Social network feeds provide posts and its comments traditionally associated with social networks. Most commonly feeds from Twitter and Facebook are analyzed. Discussion forums include websites and IS on which users discuss a certain topic relevant to its community, ask and answer questions, etc. This also includes crowdsourcing communities commonly investigated in IS on which users provide product feedback or feature requests in a structured and company-specific context, e.g., Salesforce's IdeaExchange. Finally, the media provider data source relates to web platforms that provide media, especially video and audio, which can be analyzed with NLP through transcription. Examples include podcast aggregators or video streaming websites.

NLP technique: this dimension describes the NLP technique or method that is used by researchers in IS to understand text is another commonly analyzed dimension. For example, Liu et al. (2017) provide a structure for common NLP tasks within IS such as sentiment analysis or semantic annotation. With our work, we shift our focus more to state-of-the-art techniques because we explicitly want to outline the capacity to develop these IS use cases further. The characteristics here are mainly used to track progress. While a more powerful statistical model is not necessarily equivalent to better outcomes, it does show where research has been performed to better understand the use of these techniques. Since we have already elaborated on these techniques in the previous section on advanced deep learning techniques, we do not elaborate on them here. It is important to highlight that LLMs usually leverage the transformer architecture. Hence, for the purpose of the visualization, the "transformer models" characteristic refers to transformer models that are not covered by the LLM definition. We have chosen this wording for now because transformer models are already a more established term in IS research and are distinct from LLMs.

5 Exemplary application of the taxonomy

As outlined, usefulness is central to determine the success of a taxonomy. With this section, we want to guide its usage by looking at a specific example that illustrates how the taxonomy can be applied to an IS or its component. To reflect the advanced deep learning techniques that motivate this paper, and the creation of the taxonomy as well as offering the taxonomy to practitioners, we have opted to showcase an LLM-supported IS that is currently being developed by the authors. Hence, our approach is more novel, but it is important to note that the IS has not been validated externally yet. In the following we briefly describe the use case and its classification within the taxonomy.

The demo IS is a web application (see Figure 2) that analyzes an arbitrary set of online reviews specific to a product on Amazon or an app on the Google Play Store. The insight level is therefore determined to be *product*. The aim was to derive qualitative suggestions from user reviews of the respective products. The insight measure is therefore *textual* and the user type *consumer*. The content type is *review*. We used an Amazon review dataset (Bražinskas et al., 2021) and scraped app reviews from the Play Store for this purpose. The source type is therefore *review platforms*.

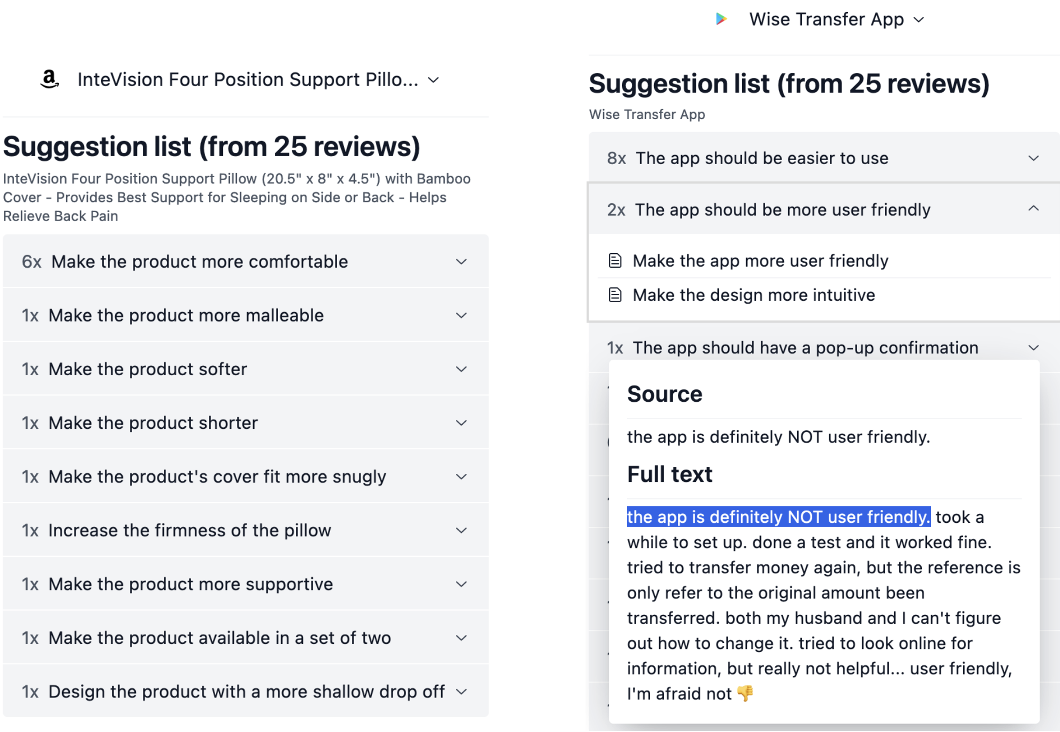

Figure 2. Product suggestions identified from online reviews.

As written in the method, to apply the taxonomy, the IS may have to be split into its components if a diverse NLP pipeline is used. This is necessary because an IS can feature any combination of data sources, content, types, etc. Otherwise, mutual exclusiveness between taxonomy characteristics may not be given. This split may be omitted by practitioners, if mutual exclusiveness is not needed to understand the analyzed IS. As seen, this split was not necessary for our IS. However, if for example diverse source types are used, e.g., on top of review our IS also analyzed preferences stated in social network posts, two separate classifications should be performed. Beyond classification, the taxonomy can also be used to guide the creation of a new design IS artifact that uses NLP to derive insights from UGC through following the questions posed in the canvas for each dimension.

6 Discussion and contribution to theory and practice

With this paper, we created a novel taxonomy that helps both practitioners and researchers identify, design, compare and evaluate the use of UGC within an IS to derive business insights. In the taxonomy, we examine and aggregate the various ways in which UGC is used in IS and suggest a novel design space and structure of dimensions as well as the corresponding characteristics. For better comprehensibility, we illustrate our resulting structure as an intuitive feature canvas. Regarding practice, we especially address people designing data-driven products or projects that aim to develop business insights on a specific topic. Researchers can use this taxonomy to evaluate the current technological state of an IS and the varieties of UGC it uses with respect to deriving business insights. Furthermore, the classification can be used to compare the IS to others with the unique lens we applied. The taxonomy puts special emphasis on integrating novel NLP approaches, especially LLMs. With this emphasis, we hope to trigger future research on the use of LLMs in IS research in general and for UGC specifically. To display the unique capacity of LLMs to researchers in the IS community, we presented a prototype that derived product suggestions from an arbitrary review dataset. In doing so, we also show how LLMs can be used to develop new or extend existing NLP applications or research streams in the realm of IS. As outlined in section 2.2, IS researchers have analyzed different types of UGC, such as online reviews, for a large variety of use cases as well as purposes and may therefore benefit from the use of LLMs.

From a theory perspective, we extend this literature by combining existing literature to create a new and up-to-date taxonomy from an integrative IS research point-of-view that also goes beyond the existing classifications of UGC in a single domain. With this integrative IS research point-of-view, we aim to support the decision-making of researchers and practitioners (Bailey, 1994; Usman et al., 2017). Our taxonomy opens possibilities for further study and a more thorough understanding of how to build, assess, or compare UGC extraction pipelines to derive business insights (ideally) based on LLM. We believe that our research findings can offer a valuable starting point for further developing guidelines or frameworks for IS for analyzing UGC, especially for special use cases where LLM might play a crucial role for successful deployment. Given the immense growth of the use of UGC in research and practice, as well as the interest in LLM, further research on this topic is warranted. Such research will require a structural foundation and a theoretical understanding that embeds LLM into IS research on UGC. In this paper, we offer such an understanding by presenting our taxonomy, its evaluation and exemplary application. In this regard, we offer researchers to apply a more nuanced perspective when leveraging LLM or other NLP models based on our canvas (Figure 1).

From a practical perspective, a common understanding of relevant characteristics also fosters several insights for the fields of marketing research, computer linguistics or political sciences. Our systematic classification enables researchers and practitioners to design, evaluate, compare, and predict how different component combinations impact the predictive outcome in our domain more effectively. For instance, at a basic level, our taxonomy determines the minimum number of design decisions an IS researcher must take when designing an IS that derives insights from UGC.

Of course, this research has several limitations which provide avenues for future research. First, the literature review is limited to the journals from the AIS Senior Scholars' Basket of Journals as well as the ICIS and ECIS conferences. Extending the review and taxonomy to literature from other research disciplines or additional IS outlets could improve and alter the taxonomy. Furthermore, we only looked at empirical studies. Practical designs are often more cutting-edge and provide additional insights and practical context. With the rapid advancements of NLP technology in mind, parts of the study may become outdated quickly and require follow-up. This is especially true for our exemplary application. Beyond the prototype we showcased, further usage demonstrations of the taxonomy are needed. However, encouraged by the evaluation with experts, and outlined in this section previously, we believe that our taxonomy represents a valuable state-of-the-art tool to analyze the use of UGC within IS to derive business insights for current IS research and practice.

7 Conclusion

To conclude, our taxonomy helps both practitioners and researchers identify, design, compare and evaluate the use of UGC within an IS with the purpose of deriving business insights. The taxonomy is based on a sample of 215 IS papers. Our work provides a framework for the use of LLMs in such a setting and aims to motivate researchers to study these models further for their own studies and advance the usage of UGC within IS. Finally, we presented an IS prototype that derives product suggestions, displaying the novel capacity of LLMs and applied our taxonomy to it. In doing so, we showed how LLMs can be used to develop new or extend existing NLP applications or research streams in the realm of IS.

References

Agarwal, M. (2020). Importance of User Generated Content as a part of Social Media Marketing that drives Customer’s Brand Awareness and Purchase Intentions. The International Journal of Analytical and Experimental Modal Analysis 12, 3071.

Au, C., Winkler, T.J., Paul, H. (2022). Towards a Generation of Artificially Intelligent Strategy Tools: The SWOT Bot.

Bailey, K.D. (1994). Typologies and Taxonomies: An Introduction to Classification Techniques. Sage, Newbury Park, CA.

Bostrom, R.P., Heinen, J.S. (1977). MIS Problems and Failures: A Socio-Technical Perspective. Part I: The Causes. MIS Quarterly 1, 17–32. https://doi.org/10.2307/248710

Bražinskas, A., Lapata, M., Titov, I. (2021). Learning Opinion Summarizers by Selecting Informative Reviews. https://doi.org/10.48550/arXiv.2109.04325

Chintagunta, B., Katariya, N., Amatriain, X., Kannan, A. (2021). Medically aware gpt-3 as a data generator for medical dialogue summarization, in: Machine Learning for Healthcare Conference. PMLR, pp. 354–372.

Donthu, N., Kumar, S., Pandey, Neeraj, Pandey, Nitesh, Mishra, A. (2021). Mapping the electronic word-of-mouth (eWOM) research: A systematic review and bibliometric analysis. Journal of Business Research 135, 758–773. https://doi.org/10.1016/j.jbusres.2021.07.015

Eisenstein, J. (2019). Introduction to Natural Language Processing, Adaptive Computation and Machine Learning series. MIT Press, Cambridge, MA, USA.

Fedus, W., Zoph, B., Shazeer, N. (2021). Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity.

Hu, Y., Gal, D., Hong, Y. (2018). Modeling brand personality with business value of social media analytics: Predicting brand personality with user-generated content and firm-generated content, in: 39th International Conference on Information Systems, ICIS 2018. Association for Information Systems.

Ishii, Y., Madotto, A., Fung, P. (2022). Survey of Hallucination in Natural Language Generation. ACM Comput. Surv 1.

Jönsson, A.M., Örnebring, H. (2011). User-generated content and the news: Empowerment of citizens or interactive illusion? Journalism Practice 5, 127–144.

Kang, Y., Cai, Z., Tan, C.-W., Huang, Q., Liu, H. (2020). Natural language processing (NLP) in management research: A literature review. Journal of Management Analytics 7, 139–172. https://doi.org/10.1080/23270012.2020.1756939

Karanam, A., Agarwal, A., Barua, A. (2021). ML-Based Product Design: The Case of Mobile Apps.

Khurana, S., Qiu, L., Kumar, S., 2019. When a doctor knows, it shows: An empirical analysis of doctors' responses in a Q&A forum of an online healthcare portal. Information Systems Research 30, 872–891.

Landolt, S., Wambsganss, T. and Söllner, M. (2021). A taxonomy for deep learning in natural language processing. Hawaii International Conference on System Sciences.

Li, S., Liu, F., Zhang, Y., Zhu, B., Zhu, H., Yu, Z. (2022). Text Mining of User-Generated Content (UGC) for Business Applications in E-Commerce: A Systematic Review. Mathematics 10, 3554.

Liu, D., Li, Y., Thomas, M.A., (2017). A Roadmap for Natural Language Processing Research in Information Systems, in: HICSS. https://doi.org/10.24251/HICSS.2017.132

Lukyanenko, R., Parsons, J., Wiersma, Y., Wachinger, G., Huber, B., Meldt, R. (2017). Representing crowd knowledge: Guidelines for conceptual modeling of user-generated content. Journal of the Association for Information Systems 18, 2.

Lukyanenko, R., Parsons, J., Wiersma, Y.F., Maddah, M. (2019). Expecting the Unexpected: Effects of Data Collection Design Choices on the Quality of Crowdsourced User-Generated Content. MIS Q 43, 623–647. https://doi.org/10.25300/MISQ/2019/14439

Moens, M.-F., Li, J., Chua, T.-S. (2014). Mining user generated content. CRC press.

Naab, T.K., Sehl, A. (2017). Studies of user-generated content: A systematic review. Journalism 18, 1256–1273.

Nguyen, T.T., Tripathi, A. (2019). A Taxonomy of User-Generated Content (UGC) Applications, in: Workshop on E-Business. Springer, pp. 169–182.

Nickerson, R.C., Varshney, U., Muntermann, J. (2013). A method for taxonomy development and its application in information systems. European Journal of Information Systems 22, 336–359. https://doi.org/10.1057/ejis.2012.26

Rietsche, R., Frei, D., Stoeckli, E., Söllner, M. (2019). Not all Reviews are Equal - A Literature Review on Online Review Helpfulness. Research Papers.

Safi, R., Yu, Y. (2017). Online product review as an indicator of users' degree of innovativeness and product adoption time: a longitudinal analysis of text reviews. European Journal of Information Systems 26, 414–431.

Shin, D., He, S., Lee, G.M., Whinston, A., Cetintas, S., Lee, K.-C. (2020). Enhancing Social Media Analysis with Visual Data Analytics: A Deep Learning Approach. MIS Quarterly 44, 1459–1492. https://doi.org/10.25300/MISQ/2020/14870

Singh, P.V., Sahoo, N., Mukhopadhyay, T. (2014). How to Attract and Retain Readers in Enterprise Blogging? Information Systems Research 25, 35–52. https://doi.org/10.1287/isre.2013.0509

Soylemez, K.C. (2021). 4W of user-generated content: why who we are and where we post influence what we post. Journal of Research in Interactive Marketing.

Stahlmann, S., Ettrich, O., Schoder, D. (2022). Deep Learning Enabled Consumer Research for Product Development.

Szopinski, D., Schoormann, T., Kundisch, D. (2019). Because Your Taxonomy is Worth IT: towards a Framework for Taxonomy Evaluation., in: ECIS.

Timoshenko, A., Hauser, J.R. (2019). Identifying customer needs from user-generated content. Marketing Science 38, 1–20.

Usman, M., Britto, R., Börstler, J., Mendes, E. (2017). Taxonomies in software engineering: A Systematic mapping study and a revised taxonomy development method. Information and Software Technology. https://doi.org/10.1016/j.infsof.2017.01.006

vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., Cleven, A. (2015). Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research. Communications of the Association for Information Systems 37. https://doi.org/10.17705/1CAIS.03709

Wambsganss, T., Kueng, T., Söllner, M., Leimeister, J.M. (2021). ArgueTutor: An Adaptive Dialog-Based Learning System for Argumentation Skills, in: Kitamura, Y., Quigley, A., Isbister, K., Igarashi, T., Bjørn, P., Drucker, S. (Eds.), Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 1–13. https://doi.org/10.1145/3411764.3445781

Wang, G., Liu, X., Fan, W. (2011). A Knowledge Adoption Model Based Framework for Finding Helpful User-Generated Contents in Online Communities. Presented at the International Conference on Information Systems 2011, ICIS 2011.

Webster, J., Watson, R.T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly 26, xiii–xxiii.

Wei, X., He, W., Zhang, X., Zhao, C., Zhao, H. (2020). A Machine Learning Method for Measuring Information Disclosure in Sharing Economy Platforms., in: ICIS.

Weingart, P., Wambsganss, T., Söllner, M. (2022). Unleashing the Potential of Argument Mining for IS Research: A Systematic Review and Research Agenda. Presented at the ICIS 2022 Proceedings.

Xu, F.F., Alon, U., Neubig, G., Hellendoorn, V.J. (2022). A systematic evaluation of large language models of code, in: Proceedings of the 6th ACM SIGPLAN International Symposium on Machine Programming. pp. 1–10.

Ye, J., Blohm, I., Breschneider, U., Goswami, S., Leimeister, J.M., Krcmar, H. (2016). Promoting the quality of user generated ideas in online innovation communities: A knowledge collaboration perspective. Association for Information Systems.

Ye, Q., Lin, B.Y., Ren, X. (2021). Crossfit: A few-shot learning challenge for cross-task generalization in nlp. arXiv preprint arXiv:2104.08835.

Zhang, X., Du, Q., Zhang, Z. (2021). What Types of Crowds Generate More Valuable Content? Evidence from Cross-Platform Posting. Evidence from Cross-Platform Posting (July 7, 2021).

Zhou, S., Qiao, Z., Du, Q., Wang, G.A., Fan, W., Yan, X. (2018). Measuring Customer Agility from Online Reviews Using Big Data Text Analytics. Journal of Management Information Systems 35, 510–539. https://doi.org/10.1080/07421222.2018.1451956