**Wolfram Höpken / Matthias Fuchs / Maria Lexhagen**

# Analyzing Tourism Online Reviews: An Extended Approach to Hierarchical Topic Detection Using Keyword Clustering

## Abstract

Tourism managers are increasingly turning to the online sphere to gain relevant customer insights. However, current approaches to analyzing vast and rapidly changing user-generated content (UGC) face several limitations. Supervised approaches require significant effort to provide pre-tagged training data and cannot dynamically identify topics mentioned in UGC. On the other hand, unsupervised approaches typically do not support different abstraction levels or enable a successive refinement of analysis in a drill-down manner, which is often expected as a practical requirement of tourism and destination management. Our research objective is, therefore, to extend current supervised approaches for identifying predefined topics by adopting unsupervised approaches using cluster analysis. The results emphasize that unsupervised approaches can (1) detect non-predefined topics dynamically with an accuracy similar to supervised approaches, thus demonstrating the potential to replace them and avoid the necessity of providing pre-tagged training data. (2) To build a topic hierarchy, unsupervised approaches sense more fine-grained topics as an enhancement of predefined topics on a lower level of abstraction, enabling more powerful drill-down-like analyses. Overall, the proposed extended approach to topic detection promises to support tourism management by meaningfully analyzing the increasing mass of visitors' online feedback.

***Keywords***: topic detection, topic hierarchy, keyword clustering, user-generated content, tourism online reviews

## 1. Introduction

In the tourism domain, structured and unstructured data from the omnipresent digital environment offer vast opportunities for gaining managerial insights, such as information on tourists' experiences, feelings, interests, and opinions; the analysis of tourist behavior; and the prediction of tourism trends (Höpken et al., 2015; Höpken et al., 2020; Li et al., 2016). Li et al.'s (2018) seminal analytical framework for classifying big data in tourism suggests categories of such data based on their sources. Notably, in one of the three types of sources, it is the user who, using social media, creates online textual data (e.g., reviews and blogs) and visual data, the so-called UGC. Hence, typical data sources for UGC studies in tourism are consumer review websites, blogs, media-sharing websites, social networks, and virtual communities (Höpken et al., 2017).

A well-known automated method of analyzing UGC in the tourism domain is sentiment analysis, which detects consumer reviews' topics and sentiments (Aurchana et al., 2014). On the one hand, topic detection automatically identifies pre-defined topics mentioned within an online review (Menner et al., 2016) or extracts sets of linguistic terms with a coherent meaning from a review as so-called latent topics (Rossetti et al., 2016). On the other hand, sentiment detection refers to extracting subjectivity in a text in the form of objective or

**Wolfram Höpken**, PhD, Corresponding Author, Professor, Institute for Digital Transformation, Ravensburg-Weingarten University of Applied Sciences, Weingarten, Germany; ORCID ID: https://orcid.org/0000-0002-4175-1295; e-mail: wolfram.hoepken@rwu.de

**Matthias Fuchs**, PhD, Professor, Department of Business Economics, Law, Geography and Tourism/ETOUR, Mid-Sweden University, Östersund, Sweden; ORCID ID: https://orcid.org/0000-0003-3964-2716; e-mail: matthias.fuchs@miun.se

**Maria Lexhagen**, PhD, Professor, Department of Business Economics, Law, Geography and Tourism/ETOUR, Mid-Sweden University, Östersund, Sweden; ORCID ID: https://orcid.org/0000-0002-6610-9303; e-mail: maria.lexhagen@miun.se

TOURISM
An International Interdisciplinary Journal
Original Research Article
Vol. 72/ No. 1/ 2024/ 7 - 19
UDC: 338.48: [655.55:004.031.4; https://doi.org/10.37741/t.72.1.1

subjective statements and positive or negative polarity (Neidhardt et al., 2017). The use of sentiment analysis has gained tremendous popularity in recent years since UGC-based data's volume, variety, and velocity (i.e., the real-time character) make it a desirable source for opinion mining and sentiment analysis.

The specific task of topic detection, which is the focus of this study, has been addressed through different approaches in the literature (Mehraliyev et al., 2022). Relatively simple approaches identify predefined topics based on wordlists or lexicons. In such cases, the list of topics is fixed, and the listed words must typically be predefined manually. Consequently, more robust approaches have emerged in the form of supervised and unsupervised machine learning techniques. Supervised approaches automatically learn relevant words to identify a specific topic based on a corpus of training data (i.e., reviews). However, they continue to suffer from two drawbacks: (1) topics must be predefined, so a dynamic identification of topics and their changes over time (i.e., a topic drift) are not possible, and (2) manually pre-tagged comprehensive training data must be provided, constituting a work-intensive task leading to subjective and, thus, potentially biased results. By contrast, in the case of unsupervised machine learning approaches (e.g., keyword clustering or topic modelling), topics are not predefined, may vary over time (i.e., topic drift), and can be identified on a more fine-grained level (i.e., long-tail topic detection). Supervised and unsupervised approaches to topic detection have been used intensively in tourism science. Nonetheless, to the best of our knowledge, it has not been analyzed (1) whether unsupervised approaches can complement or even wholly replace supervised approaches and (2) whether it makes sense and is possible to hierarchically identify topics at finer levels of granularity, thus enabling drill-down-like analyses.

The current study, therefore, presents a new approach. It aims to extend supervised topic-detection approaches and identify predefined topics in the tourism domain by adopting unsupervised topic detection using cluster analysis (i.e., keyword clustering). Unsupervised topic detection enables the dynamic identification of subjects without requiring predefined topics and, thus, can dynamically identify changes and trends concerning the strength and sentiment of cases (Liu, 2011). Completely replacing supervised topic detection with unsupervised approaches may even make manually classifying a significant amount of consumer reviews as training input to a supervised machine learning process unnecessary. Thus, in the current study, we will answer the following research question:

RQ1: Are unsupervised approaches capable of detecting topics dynamically with an accuracy like supervised methods, thus demonstrating their potential to replace directed processes of topic detection in tourism?

Additionally, we anticipate that unsupervised topic detection may be capable of identifying topics at a finer level of granularity. This feature of topic detection would facilitate long-tail topic detection (Anderson, 2004). Although most sales in internet-based tourism today are made with niche products, the capacity of unsupervised topic detection to identify topics on a finer level of granularity has not been tested in the tourism literature thus far. From a practical perspective, narrowing down a topic analysis in a top-down manner is fascinating. Such a process begins with high-level topics and then drills down to more fine-grained issues hierarchically, enabling less fragmented and more comprehensive results. This leads us to pose research question 2:

RQ2: Are unsupervised topic-detection approaches capable of detecting more fine-grained tourism topics that can be hierarchically assigned to predefined issues as a valuable extension at a lower level of abstraction?

To respond to research questions 1 and 2, an unsupervised keyword-clustering approach will be conceptualized and executed. First, the approach will be applied on the same level of granularity as classical supervised topic-detection approaches to validate their replacement by unsupervised approaches and, second, on a more fine-grained level to build a topic hierarchy. All presented and discussed approaches will be cross-validated by comparison with classical supervised methods of topic detection, such as naïve Bayes or support vector machine (SVM)-based text classification.

The current study will be conducted on hotel reviews extracted from TripAdvisor's online review platform for the Swedish destination of Halland, a tourism region on the southwest coast of Sweden.

## 2. Literature review

The analysis of user-generated content (UGC) typically addresses topic detection, which aims to identify the topic (or aspect) the user discusses, and sentiment detection, which seeks to determine the polarity of the statement. Recently, Mehraliyev et al. (2022) executed a literature review of 70 papers from the hospitality and tourism domain and discovered that the analysis of big data is scant and that testing the performance of comparative approaches continues to be uncommon, especially for the task of feature extraction (i.e., topic detection). The specific task of topic detection, which we will focus on in this study, has been addressed through different approaches in the literature, as discussed below.

### 2.1. Wordlist, or lexicon-based approaches

Liu (2011) thoroughly overviews the general approach of wordlist- or lexicon-based sentiment and topic detection. Predefined (e.g., tourism-relevant) topics (e.g., room, service and personal, and food and beverage in the case of hotel reviews) are described by a wordlist or lexicon, defined manually, or extracted from pre-classified documents through text mining techniques. The review text, which is to be classified into one of the predefined topics, is typically divided into single sentences or statements, as a full review usually addresses many different topics. Each sentence or statement is then assigned (i.e., classified) to one of the predefined topics based on the number of words of the dominant topic occurring within the statement. Following the well-known vector space model (Liu, 2011), the review text is treated as a "bag of words" (i.e., word vector), thus ignoring the order of words and sentence structure. Often, POS (part of speech) tagging is used to filter certain word types, like nouns or verbs, and to reduce the word vector size to improve performance and precision.

Höpken et al. (2017) and Fuchs et al. (2017) extracted customer reviews from platforms like TripAdvisor and Booking.com to compare dictionary-based and machine-learning algorithms for identifying topics and their sentiment. Accuracy tests revealed that dictionary-based techniques could outperform some machine learning algorithms, such as k-nearest neighbour and the Naïve-Bayes method. Mariani and Borghi (2021) used lexicon-based approaches to detect subjectivity, polarity (sentiment), and environmental presence and depth (based on an ecological dictionary) in hotel online reviews from Booking.com and TripAdvisor. They aimed to analyze travelers' environmental discourse in hotel online reviews specifically.

Referring to the examples highlighted here, wordlist-based approaches suffer from several drawbacks. Identifying topics must be defined a-priori and, thus, cannot change over time. Consequently, a potential topic drift cannot be identified. Additionally, wordlists must be manually defined for each topic. Therefore, the wordlists' expressiveness determines the approach's precision, which makes this task particularly difficult and cumbersome.

### 2.2. Supervised machine learning techniques

In the case of word-list-based text classification, the knowledge covered by words capable of identifying a particular topic class is part of the word list and is typically provided manually. Supervised learning approaches, by contrast, use machine learning techniques to automatically identify correlations between words and certain classes from pre-classified training data. The learned classification model can then be used to classify new unseen data and assign the most probable class (i.e., topic) based on the occurrences of certain words or word combinations. Topics must be predefined, and a text corpus consisting of many product reviews must be pre-classified as training data (Liu, 2012). Technically, these approaches build on the bag of words concept and pre-processing steps described above, especially filtering words based on appropriate POS tags (e.g., nouns and verbs) as input to topic detection. The most common classification techniques for topic detection are *k*-nearest neighbor (k-NN), naïve Bayes, SVM, or artificial neural networks (e.g., deep neural networks).

*K*-nearest neighbor is a lazy algorithm that simply stores all training data as a classification model. When applied to new data, the algorithm calculates the most likely class as the majority class of the *k*-nearest, thus the most similar training data elements (Liu, 2011). Naïve Bayes is a purely probabilistic approach for calculating the class probabilities of a component (e.g., a sentence) based on single and independent element characteristics, such as word occurrences. Thus, like linear regression, the naïve Bayes algorithm cannot handle interdependencies of input attributes. Nevertheless, it is a well-suited algorithm, especially for high-dimensional input spaces typical for text classification. Support vector machines transform a binary classification task into a problem of linear programming, which can be solved with well-established algorithms such as Simplex. A linear boundary (i.e., a hyperplane in an n-dimensional space) separates the two classes and, if they are not linearly separable, transforms the input space into a typically higher-dimensional feature space by using a kernel function. In the case of text classification, the *kernel* function is generally omitted due to the high dimensionality of the original input space. Like naïve Bayes, support vector machines are well-suited for text classification (Liu, 2012).

Various tourism scholars, including Xiang et al. (2017), Martinez-Torres and Toral (2019), and Luo et al. (2021), applied supervised machine learning methods such as naïve Bayes, k-NN, or SVM to the overall task of sentiment detection. For instance, Höpken et al. (2017) and Fuchs et al. (2017) used different machine-learning algorithms for topic detection. They showed that support vector machines (SVM) outperformed the k-nearest neighbor and the Naïve-Bayes method.

Like wordlist-based approaches, supervised machine learning approaches suffer when used for topic detection because topics must be predefined and cannot change over time. Additionally, a text corpus of sufficient size must be pre-classified and provided as training data, constituting a work-intensive task. Finally, the number of topics to be detected is limited, as classification techniques typically perform worse if too many different classes should be identified. Thus, supervised topic detection is mainly limited to high-level topic detection.

## 2.3. Unsupervised machine learning techniques

In the case of unsupervised machine learning techniques, topics are not predefined but deduced from words cooccurring in documents (i.e., total reviews or single review statements or sentences). Typical approaches used in past and current research studies in tourism are (1) the *identification of frequent words*; (2) *keyword clustering*, i.e., using cluster analysis to identify latent topics based on words often co-occurring in review texts or statements; (3) *factor analysis (latent semantic indexing)* using the term-document matrix as a basis for a factor analysis to identify latent topics; and (4) *topic modelling* (e.g., *Latent Dirichlet Allocation* [LDA]), generating a topic model, defining each document as a distribution of topics and each topic as a distribution of words characterizing the topic.

Several studies can be found in the literature. Menner et al. (2016) completed a comparative study of different unsupervised machine-learning approaches: identifying frequent words, keyword clustering, and latent semantic indexing. They showed that keyword clustering outperformed all other approaches. Xiang et al. (2017) analyzed the relationship between linguistic characteristics, semantic features, sentiment, rating, and usefulness of hotel reviews across different social media platforms. Within their comparative study, the authors executed an unsupervised topic detection with a Latent Dirichlet Allocation (LDA) and a sentiment detection using a word-list-based and naïve Bayes text classification approach. Qi et al. (2018) constructed a typology of cultural tourists based on TripAdvisor reviews. Visitation pattern characteristics for five typological categories were deduced from the most frequent (i.e., top 100) unique terms. Moreover, Latent Dirichlet Allocation (LDA) was applied by Serrano et al. (2021), Luo et al. (2021), and Ahani et al. (2021). More precisely, Ahani et al. (2021) applied LDA to discover main topics (i.e., satisfaction dimensions) in crawled data from medical tourism websites and a subsequent clustering based on the expectation maximization approach. Geetha et al. (2017) applied a hierarchical clustering technique to explain the expressed sentiments by identifying terms used within customer reviews; thus, in their study, clustering was not used for topic detection but for grouping terms in general.

As outlined, unsupervised approaches to topic detection offer the advantage that topics do not have to be predefined but can be found dynamically and change over time (i.e., topic drift). Additionally, there is no need to provide pre-classified training data.

## 2.4. Current limitations and the proposed approach

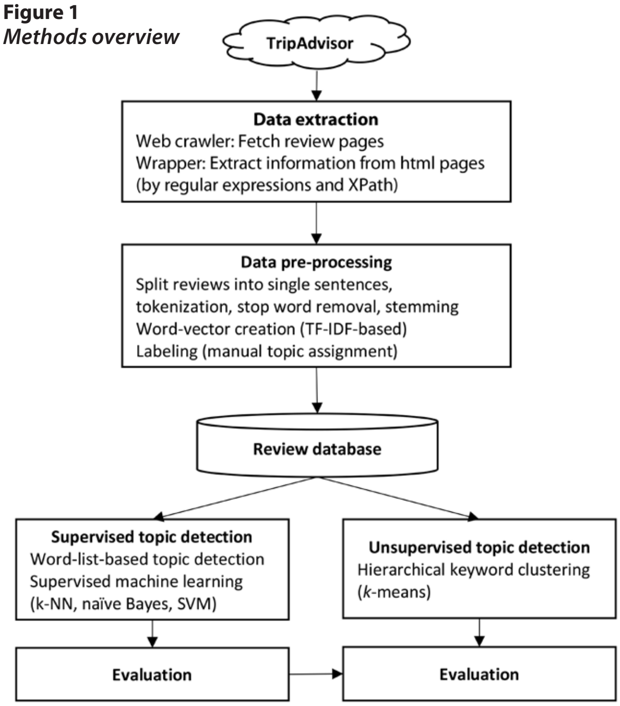
As discussed, supervised and unsupervised topic-detection approaches are widely used in tourism. However, the idea that unsupervised approaches can cover predefined topics (i.e., identify topics) that semantically overlap with predefined ones, therefore replacing supervised approaches without the burden of providing pre-classified training data, has never been validated. The current study will close this gap and answer the research question of whether unsupervised approaches are adept at dynamically detecting predefined topics with an accuracy similar to supervised approaches (RQ1).

Moreover, unsupervised topic-detection approaches can identify topics at any level of granularity and, thus, more fine-grained topics than supervised approaches. Nonetheless, they have never been used to identify issues on different levels of granularity in the tourism domain and build a topic hierarchy to refine topics drill-down successively. The current study will close this gap and answer the research question of whether unsupervised topic-detection approaches can detect more fine-grained issues that can be assigned appropriately to predefined topics as valuable extensions on a lower level of abstraction (RQ2).

# 3. Methods

The figure below (Figure 1) provides an overview of the methods used in the present study. It illustrates the tasks of data extraction, pre-processing, and supervised and unsupervised topic detection, described in the following sections.

**Figure 1**
*Methods overview*



## 3.1. Data extraction and pre-processing

There are different approaches to extracting relevant data in the form of user reviews from online platforms, such as social media or review sites. While an increasing number of platforms offer access through specific interfaces (e.g., application programming interfaces [APIs]), a more flexible approach extracts data directly from the website (i.e., single *HTML* pages), which is then feasible for any kind of online platform. Therefore, this study used a *web crawler* to search different pages of TripAdvisor's online platform and a consecutive *wrapper* to extract relevant information from *HTML* pages using regular expressions and XPath statements (Liu, 2011).

Completing reviews typically address many topics, so assigning single sentences to a concrete case is more appropriate. Thus, extracted review texts are first split into single sentences. Then, we execute further steps of pre-processing using specific text-processing techniques. First, *tokenization* splits a natural language text into single tokens (i.e., words) and removes all non-letter characters. Second, *stop words* (i.e., common words), which carry no specific meaning and comprise up to 20–30% of all terms in a natural language text (Liu, 2012), are removed to increase the expressiveness of the remaining words. Third, *stemming*, a technique to reduce inflected words to their word stem, an analogue to stop word removal, can potentially decrease the number of different words by up to 40–50% (Liu, 2011). Fourth, *part-of-speech (POS) tagging* assigns each word in a sentence to a word category (e.g., noun or verb) and is used in our study to filter words with specific POS tags, making use of the Penn Treebank POS tag set, which specifies 36 different POS tags (Mitchell et al., 1994). Fifth, *N-grams* are contiguous sequences of items, words in our case, treated as compound words (Liu, 2011). In this study, we make use of bi- and tri-grams. Natural language processing (NLP) approaches to topic detection typically build on a word-vector model. Thus, the original text (i.e., sentence) is transformed into a word vector consisting of a binary word or term occurrences, term frequencies, or, in our case, term frequency–inverse document frequency (TF-IDF) values, which specify the representativeness of a term for a sentence.

## 3.2. Supervised topic detection

Supervised topic detection here denotes identifying a limited number of predefined and fixed topics. Specifically, each review sentence or statement is assigned to a predefined topic. As highlighted above, in this study, we use wordlist-based approaches and supervised machine learning techniques, which serve as a baseline to validate the suitability of unsupervised methods for topic detection and answer research question 1.

### *3.2.1. Word-list-based topic detection*

Based on a list of representative words for each class (i.e., topic), the word-list-based approach simply counts word occurrences for each category. It identifies the majority class (cf. section 2.1). In the case of a tie, we chose the most frequent and, thus, dominant type. Following previous research (Höpken et al., 2017), we manually defined wordlists for the topics room, food and beverages, location, standard facilities, staff, destination, guests, and property, containing between 3 and 13 words.

### *3.2.2. Supervised learning approaches*

This part of our study aims to develop a solid foundation for answering research question 1: whether the proposed unsupervised topic detection approach can accurately identify predefined topics like supervised learning approaches. Thus, we used three of the most famous machine learning techniques for text classification, namely k-nearest neighbors (k-NN), naïve Bayes, and support vector machines (SVM) to perform the classification task of topic detection (Liu, 2012). For all three techniques, we tested different options for filtering words based on their POS tag (i.e., other forms of nouns) and bi- and trigrams. For k-NN, values of k up to 200 were tested. SVM has been executed with the dot-product as a kernel function. A 10-fold cross-validation validated all approaches.

## 3.3. Unsupervised topic detection by keyword clustering

In contrast to the supervised learning approaches mentioned above, unsupervised topic detection does not require pre-classified training data or predefined topics. Instead, the approach deduces topics directly from the data (Liu, 2012). Like supervised topic detection, in addition to other pre-processing steps, POS tagging is used in this study to filter nouns and verbs. Previous studies have compared different approaches to unsupervised topic detection, namely *identification of frequent nouns and verbs, keyword clustering, and latent semantic indexing*, a factor analytical approach (Höpken et al., 2017; Menner et al., 2016). In this study, we focus on keyword clustering to validate its suitability for identifying predefined topics with accuracy, like supervised approaches (research question 1) and identifying fine-grained topics hierarchically (research question 2).

Keyword clustering groups sentences into clusters comprising the exact words, and the collections represent latent topics. In our study, the predominant words within the penalties of a set, mirrored by high TF-IDF values (called *keywords*), can then be used to describe, or characterize the topic. In this study, the *k*-means clustering algorithm was applied. The *k*-means algorithm has been parametrized by the number of clusters *k* to be found and aims to achieve even partitioning. Thus, the algorithm is well suited to identifying groups on different abstraction levels by specifying different values for *k*.

When executing keyword clustering on different abstraction levels, each cluster on one level of abstraction is assigned to the most overlapping cluster on the next highest abstraction level by calculating the number of sentences in both clusters. Using this approach, we can iteratively refine topics on a lower abstraction level and build a topic hierarchy with consecutive topic refinements on different abstraction levels (research question 2).

To answer research question 1, this iterative process is started on the same abstraction level, thus with the same number of topics as the predefined topics in the case of supervised topic detection. We test whether unsupervised learning can substitute directed learning approaches by calculating the percentage of overlapping sentences and assigning each topic to the most overlapping predefined topic.

# 4. Findings

## 4.1. Data extraction

In this study, we extracted all tourism reviews (622) from the social media platform TripAdvisor (www.tripadvisor.com) for 24 hotels in the Swedish destination region Halland during a funded research project in the pre-COVID-19 period between 2015 and 2016. Of these reviews, 73% were provided by Swedish customers, 6% by tourists from other Scandinavian countries, and 21% by tourists from the rest of the world, which corresponds well to the statistical guest mix of Sweden (www.scb.de). As a relatively small tourism region in an early destination lifecycle stage (Eber et al., 2018), Halland offers a small but homogeneous corpus of tourism reviews, constituting an appropriate basis to validate our approach.

All reviews were split into single sentences or statements, and sentences with less than three words were removed, resulting in 4,761 single sentences. Sentences have been manually classified into the predominant topics (i.e., Room: 899; Food and Beverages: 795; Location: 524; Common Facilities: 602; Staff: 489; Destination: 130; Guests: 68; Property: 44; and N/A: 1210), serving separately as training and test data for the supervised and unsupervised approaches, respectively, as described above.

## 4.2. Supervised topic detection

In this subsection, we present the results of all approaches aiming to classify review sentences into predefined topics, namely a word-list-based approach and the supervised learning techniques k-nearest neighbor (k-NN), support vector machines (SVM) and naïve Bayes. Interestingly, despite its simplicity, the word-list-based approach reached an average accuracy over all topics of 71.0% and a kappa of 0.642, constituting competitive results compared to the other algorithms applied (cf. Table 1). Cohen's kappa measures accuracy compared to the precision reached by chance and takes values between 0 (corresponding to the accuracy of the baseline) and 1 (corresponding to 100% accuracy). The k-NN algorithm reached an accuracy of 71.9% (kappa 0.654) with k=120 and filtering nouns by the POS tag "N.*" (a regular expression summarizing the POS tags NN [Noun, singular or mass], NNS [Noun, plural], NNP [Proper noun, singular], NNPS [Proper noun, plural]). Interestingly, N-grams could not improve our findings. SVM achieved the best results of all supervised learning algorithms with an accuracy of 75.5% (kappa 0.699) when using bi-grams and no POS tagging. Finally, although naïve Bayes is known to be an appropriate algorithm for text classification, our study produced poor results with an accuracy of only 51.5% (kappa 0.427) when using bi-grams and no POS tagging.

**Table 1**
***Results of supervised topic detection***

| Algorithm | Accuracy | kappa |
|---|---|---|
| Word-list-based | 71.0% | 0.642 |
| *k*-NN POS=NN.* n-grams=no | 71.9% | 0.654 |
| SVM POS=no n-grams=2 | 75.5% | 0.699 |
| Naïve Bayes POS=no n-grams=2 | 51.5% | 0.427 |

The results of supervised learning approaches for topic detection in Table 1 constitute the baseline of our study to validate whether unsupervised approaches can identify predefined topics with similar accuracy, as presented in the next section.

## 4.3. Unsupervised topic detection using keyword clustering

As the unsupervised topic detection approach, we executed k-means-based keyword clustering on different levels of granularity. First, such clustering was applied to the same level as the predefined topics, thus with k = 10, to answer research question 1. All clusters were adequately assigned to one of the predefined topics, underpinned by the keywords through which each cluster can be described (e.g., the predefined topic *common facilities* by the keywords a*rea, pool*, and *spa* or the topic *location* by the keywords *city*, *Halmstad*, *station*, *town*, *train*, and *walk* (column 2 of Table 2). When checked on the level of single sentences, the assignment reaches satisfactory accuracy levels per cluster at slightly above or near 70% for all but two clusters (column 5 of Table 2), demonstrating the possibility of replacing supervised topic detection with unsupervised topic detection in the tourism domain (research question 1). Thus, using unsupervised topic detection approaches, the work-intensive task of manually classifying training data may be omitted, and topics can be identified dynamically.

Second, we executed the clustering approach for higher values of *k* (*k*=25) to identify more fine-grained topics and build a topic hierarchy (i.e., to answer research question 2). Table 2 shows the corresponding results. For example, food and beverages are divided into five more granular topics: *breakfast*, *restaurant*, *dinner*, *drinks*, and *food*, with an assignment accuracy of 61.8%. The topic *location* is divided into four subtopics: *beach*, *centrality*, *town*, and *parking*, with an assignment accuracy of 68.5%. Finally, the topic *room* is divided into four subtopics: *bathroom*, *view*, *bed*, and *room costs*, with an assignment accuracy of 73.7%. Notably, all results in Table 2 demonstrate that fine-grained clusters (i.e., topics) constitute a meaningful extension of predefined or higher-level issues and, thus, that the proposed approach is suitable for building a meaningful topic hierarchy (research question 2). More precisely, results in Table 2 demonstrate that topics identified on different abstraction levels support a stepwise refinement of the level of UGC analysis and a drill-down-like approach when examining topics customers are discussing and concerned with. Refining the analysis from reviews addressing, for example, food and beverage to those specifically addressing topics like breakfast or dinner supports a more systematic and less fragmented analysis path when assessing fine-grained issues exclusively.

**Table 2**
*Results of unsupervised topic detection by k-means clustering*

| Unsupervised topics (k=10) | | | | | Unsupervised topics (k=25) | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Predefined topic** | **Cluster keywords** | **Size (%)** | **WCV** | **Acc.** | **Topic** | **Cluster keywords** | **Size (%)** | **WCV** | **Acc.** |
| Common facilities | area, pool, spa | 8.8 | 1.05 | 73.4% | spa | area, pool, sauna, spa, treatment | 6.1 | 1.07 | 84.4% |
| Food and beverages | breakfast | 24.9 | 0.99 | 49.3% | breakfast | breakfast, place | 8.3 | 0.90 | 61.8% |
| | | | | | restaurant | restaurant | 3.7 | 1.13 | |
| | | | | | dinner | dinner, food | 3.5 | 1.10 | |
| | | | | | beverages | coffee, lobby, tea | 2.2 | 1.07 | |
| | | | | | food | breakfast, buffet, dinner, event, lunch | 3.2 | 0.87 | |
| Location | city, halmstad, location, station, town, train, walk | 9.2 | 1.04 | 71.5% | beach | beach, hotel, location | 3.1 | 1.18 | 68.5% |
| | | | | | centrality | center, city, halmstad, minutes, station, train, walk | 3.7 | 1.04 | |
| | | | | | town | town | 2.1 | 1.14 | |
| | | | | | parking | hotel, lot, parking, space | 3.0 | 1.10 | |
| Room | bathroom, floor, room, sea view, shower | 7.7 | 1.01 | 73.7% | bathroom | bathroom, floor, shower | 4.7 | 0.97 | 73.7% |
| | | | | | view | room, sea view | 3.0 | 1.17 | |
| | bed, room | 14.1 | 1.07 | | bed | bed, room | 3.5 | 0.75 | |
| | | | | | room costs | bit, room | 8.7 | 1.13 | |
| Staff | staff | 5.3 | 1.05 | 49.8% | reception | reception, staff | 4.6 | 1.25 | 76.2% |
| | food, service | 4.0 | 1.16 | | service | food, service, staff | 3.3 | 1.20 | |
| N/A | hotel | 14.2 | 1.06 | 67.4% | hotel | hotel | 8.1 | 1.10 | 62.3% |
| | | | | | art | art, scandic, work | 2.2 | 1.04 | |
| | | | | | experience | experience | 1.2 | 1.26 | |
| | | | | | Sweden | Sweden, weekend, year | 3.9 | 1.04 | |
| | | | | | value for money | money, value, visit, worth | 2.7 | 1.08 | |
| | | | | | going out | atmosphere, choice, drink, fun | 2.2 | 0.91 | |
| | | | | | day | day, time | 5.2 | 0.93 | |
| | | | | | night | night | 4.7 | 1.09 | |

*Note.* Keywords = words with TF-IDF value > 0.05. WCV = Within cluster variation. Acc. = Accuracy of assignment to predefined topic.

# 5. Implications

## 5.1. Theoretical implications

The current study presented a novel approach to topic detection (i.e., identifying the topic or product feature a user review or review statement is about). Specifically, the study introduced an extended topic detection approach based on hierarchical keyword clustering, enabling the identification of topics at different levels of granularity in an unsupervised manner.

Previous research used either word-list-based or supervised approaches to identify predefined topics (e.g., Höpken et al., 2017; Martinez-Torres & Toral, 2019; Luo et al., 2021) or unsupervised approaches to dynamically identify more fine-grained topics (e.g., Menner et al., 2016; Serrano at al., 2021; Luo et al., 2021; Ahani et al., 2021). To the best of our knowledge, our study is the first attempt to systematically compare both approaches to evaluate whether unsupervised approaches can replace supervised ones. One main finding of the study, and its primary contribution to the body of tourism literature, is that the unsupervised keyword-clustering approach can identify predefined high-level topics with an accuracy similar to supervised topic-detection approaches and thus replace them (RQ1). This is a remarkable insight, as unsupervised topic-detection approaches do not require any training data, which releases the analyst from the burden of providing work-intensive pre-classified training data and employing a simplified research design. Additionally, topics do not need to be predefined, and the approach can identify unknown issues, thereby avoiding potential researcher bias by falsely presupposing tourists' interests and preferences.

The current study's second and equally important finding is that, based on an unsupervised keyword-clustering approach, topics can be identified on different abstraction levels and be meaningfully composed into a topic hierarchy (RQ2). Although the task of topic detection has been heavily researched in the tourism domain, unsupervised approaches like topic modelling (e.g., Serrano et al., 2021; Luo et al., 2021; Ahani et al., 2021) cannot be used to identify topics at different prespecified levels of granularity. In contrast, based on the characteristics of the k-means clustering algorithm for identifying a predefined number of evenly distributed clusters, the keyword-clustering approach presented in this study proved its ability and appropriateness to identify topics on different abstraction levels and populate a meaningful topic hierarchy.

## 5.2. Managerial implications

As highlighted, sentiment analysis of tourism online reviews is a powerful method of supporting customer value creation because of its ability to make sense of large volumes of structured and unstructured data (Mariani et al., 2018). Sentiment analysis can support tourism businesses and destinations in decision-making by providing insights into travelers' and visitors' experiences, feelings, interests, and opinions and measuring actual behavior. By doing so, such analysis can potentially make tourism management decisions "smarter" (Bagherzadeh et al., 2021).

In the current study, we presented and empirically validated a novel approach to analyzing customer feedback dynamically and on a more fine-grained level to provide improved insights. This approach requires less preparatory effort by tourism and hospitality management (i.e., analysts) than other approaches. More precisely, the presented approach for topic detection (i.e., keyword clustering) enables the identification of topics unsupervised without the need to either predefine issues or provide pre-classified training data. This task must be repeated regularly if topics change significantly over time (i.e., topic drift). In addition to effort reduction, such an unsupervised approach avoids topical bias and can effectively address dynamic shifts in topical trends. Therefore, our proposed approach empowers hotel or destination managers to analyze customer feedback flexibly and naturally. The analysis is not restricted by predefined topical categorization, which reflects the management's point of view more than the customers' but is driven by the most critical aspects and latest trends affecting the customer experience.

Additionally, fine-grained topics along a topic hierarchy can be identified, enabling the coverage of niche topics, long-tail effects, and microtrends (Anderson, 2004). Using this approach, topics covered by review statements can be analyzed on different abstraction levels and drilled down from a more abstract to a more concrete level without losing the broader context in which these topics are mentioned. This offers hotel and destination managers a powerful and flexible instrument to analyze customer reviews, especially for niche products and specific customer segments.

Finally, most previous studies have focused on functional product dimensions and related consumption values. However, the presented approach allows more abstract but widespread and highly relevant product components or measurements to be identified, such as art, science, work, experience, value, money, worth, or atmosphere, as shown in Table 2. Thus, an important practical implication of the proposed approach for topic detection is its capacity to replace relatively simplistic decision support focused on merely functional product elements with more realistic decision support focused on differentiation based on unique experience profiles and hospitality values reflected in communicative aspects of the guest experience and expressed in customer reviews.

## 6. Conclusions

This study proposed an extended approach to sentiment analysis as an automated method of analyzing UGC through online tourism reviews. We presented the unsupervised topic detection approach of *keyword clustering*, which proved its ability to identify non-predefined topics on different levels of granularity in a hierarchical manner and validated its performance over both word-list-based methods as well as supervised machine learning approaches, such as support vector machines (SVM), naïve Bayes, and *k*-nearest neighbor (k-NN; Liu, 2012).

The results indicate that unsupervised machine learning-based approaches to dynamic topic detection can reach similar accuracy levels without defining them a-priori when comparing dynamically identified topics with predefined ones. Our proposed and empirically validated approach has the potential to replace supervised topic detection approaches, releasing analysts from the burden of predefining topics and the need to provide reclassified training data (research question 1). Additionally, unsupervised topic detection methods offer the crucial opportunity to identify issues at different levels of granularity. Again, the study results demonstrate that unsupervised topic detection approaches, such as our proposed and validated k-means clustering approach, are capable of reliably detecting fine-grained topics, which can be appropriately assigned to predefined issues as an extension on a lower level of abstraction, thus building a meaningful topic hierarchy (research question 2).

To conclude, our recommendation for future research follows Rahmani et al. (2018), who employed a psycholinguistic approach to topic modelling to analyze blogs and travelers' well-being. We agree with the authors that the detection of those experience topics that not only produce memorable experiences but also relate to travelers' eudemonic well-being over time should be pursued in the future using UGC-based data. Furthermore, we recommend future efforts toward automated topic detection in UGC data, and their interpretation be more firmly based on theories of meaning and language. Following subjectivist approaches, the intention of the UGC creator (e.g., the visitor or guest), which reflects a particular mental state and emotional outcome, should be studied in various tourism contexts.

## References

Ahani, A., Nilashi, M., Zogaan, W.A., Samad, S., Aljehane, N.O., Alhargan, A., Mohd, S., Ahmadi, H., & Sanzogini, L. (2021). Evaluating medical travelers' satisfaction through online review analysis. *Journal of Hospitality and Tourism Management, 48,* 519-537. https://doi.org/10.1016/j.jhtm.2021.08.005

Anderson, C. (2004). The long tail. The future of entertainment is in the millions of niche markets at the shallow end of the bitstream. *Wired Magazine, 12*(10), 170-177.

Aurchana, P., Plyyappan, R., & Periyasamy, P. (2014). Sentiment analysis in tourism. *International Journal of Innovative Science, Engineering & Technology, 1*(9), 443-450.

Bagherzadeh, S., Shokouhyar, S., Jahani, H., & Sigala, M. (2021). A generalizable sentiment analysis method for creating a hotel dictionary. *Journal of Hospitality & Tourism Technology, 12*(2), 210-238. https://doi.org/10.1108/JHTT-02-2020-0034

Eber, Z.F., Baggio, R., & Fuchs, M. (2018). Hyperlink network analysis of a multi-destination region: The case of Halland, South Sweden. *Journal of Information Technology & Tourism, 20*, 181-189. https://doi.org/10.1007/s40558-018-0108-9

Fuchs, M., Höpken, W., & Lexhagen, M. (2017). Business intelligence for destinations: Creating knowledge from social media. In M. Sigala & U. Gretzel (Eds.), *Advances in social media for travel, tourism, and hospitality* (pp. 290-310). Routledge. https://doi.org/10.4324/9781315565736

Geetha, M., Singha, P., & Sinha, S. (2017). Relationship between customer sentiment and online customer ratings for hotels. *Tourism Management, 61*, 43-54. https://doi.org/10.1016/J.TOURMAN.2016.12.022

Höpken, W., Eberle, T., Fuchs, M., & Lexhagen, M. (2020). Improving tourist arrival prediction: A big data and artificial neural network approach. *Journal of Travel Research, 60*(5), 998-1017. https://doi.org/10.1177/0047287520921244

Höpken, W., Fuchs, M., Keil, D., & Lexhagen, M. (2015). Business intelligence for cross-process knowledge extraction at tourism destinations. *Information Technology & Tourism, 15*(2), 101-130. https://doi.org/10.1007/s40558-015-0023-2

Höpken, W., Fuchs, M., Menner, Th., & Lexhagen, M. (2017). Sensing the online social sphere: The sentiment analytical approach. In Z. Xiang & D.R. Fesenmaier (Eds.), *Analytics in smart tourism design* (pp. 129-146). Springer. https://doi.org/10.1007/978-3-319-44263-1_8

Li, J., Xu, L., Ling, T., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management, 68*, 301-323. https://doi.org/10.1016/J.TOURMAN.2018.03.009

Li, Q., Wu, Y., Wang, S., Lin, M., Feng, X., & Wang, H. (2016). Vistravel: Visualizing tourism network opinion from user-generated content. *Journal of Visualization, 19*(3), 489-502. https://doi.org/10.1007/s12650-015-0330-x

Liu, B. (2011). *Web data mining* (2nd ed.). Springer. https://doi.org/10.1007/978-3-642-19460-3

Liu, B. (2012). *Sentiment analysis and opinion mining*. New Morgan & Claypool. https://doi.org/10.1007/978-3-031-02145-9

Luo, Y., He, J., Mou, Y., Wang, J., & Liu, T. (2021). Exploring China's 5A global geoparks through online tourism reviews. *Tourism Management Perspectives, 37*, Article 100769. https://doi.org/10.1016/j.tmp.2020.100769

Mariani, M., & Borghi, M. (2021). Environmental discourse in hotel online reviews: A big data analysis. *Journal of Sustainable Tourism, 29*(5), 829-848. https://doi.org/10.1080/09669582.2020.1858303

Mariani, M., Baggio, R., Fuchs, M., & Höpken, W. (2018). Business intelligence and big data in hospitality and tourism: A systematic literature review. *International Journal of Contemporary Hospitality Management, 30*(12), 3514-3554. https://doi.org/10.1108/IJCHM-07-2017-0461

Martinez-Torres, M.R., & Toral, S.L. (2019). A machine learning approach for the identification of the deceptive reviews in the hospitality sector using unique attributes and sentiment orientation. *Tourism Management, 75*, 393–403. https://doi.org/10.1016/j.tourman.2018.03.020

Mehraliyev, F., Chan, I.C.C., & Kirilenko, A. (2022). Sentiment analysis in hospitality and tourism: A thematic and methodological review. *International Journal of Contemporary Hospitality Management, 34*(1), 46-77. https://doi.org/10.1108/IJCHM-02-2021-0132

Menner, Th., Höpken, W., Fuchs, M., & Lexhagen, M. (2016). Topic detection: Identifying relevant topics in tourism reviews. In A. Inversini & R. Schegg (Eds.), *Information and communication technologies in tourism* (pp. 411-423). Springer. https://doi.org/10.1007/978-3-319-28231-2_30

Mitchell, M., Taylor, A., & MacIntyre, R. (1994). *Penn Treebank project*. https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html

Neidhardt, J., Rümmele, N., & Werthner, H. (2017). Predicting happiness: User interactions and sentiment analysis in an online travel forum. *Information Technology and Tourism, 17*(1), 101-119. https://doi.org/10.1007/s40558-017-0079-2

Qi, S., Wong, C., Chen, N., Rong, J., & Du, J. (2018). Profiling Macau cultural tourists by using user-generated content from online social media. *Information Technology & Tourism, 20*(2), 217-236. https://doi.org/10.1007/s40558-018-0120-0

Rahmani, K., Gnoth, J., & Mather, D. (2018). Hedonic and eudaimonic well-being: A psycholinguistic view. *Tourism Management, 69*, 155-166. https://doi.org/10.1016/j.tourman.2018.06.008

Rossetti, M., Stella, F., & Zanker, M. (2016). Analysing user reviews in tourism with topic models. *Information Technology & Tourism, 16*(1), 47-58. https://doi.org/10.1007/s40558-015-0035-y

Serrano, L., Ariza-Montes, A., Nader, M., Sianes, A., & Law, R. (2021). Exploring preferences and sustainable attitudes of Airbnb green users in the review comments and ratings. *Journal of Sustainable Tourism, 29*(7), 1134-1152. https://doi.org/10.1080/09669582.2020.1838529

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms. *Tourism Management, 58*, 51-65. https://doi.org/10.1016/j.tourman.2016.10.001

## Acknowledgements

The authors would like to acknowledge contributions to this article from the project 'The Knowledge Destination II' funded by the European Regional Development Fund (project id 20200778) and run in collaboration with Region Halland, Sweden, and the destinations of Halmstad, Varberg, Falkenberg, Hylte, Kungsbacka and Laholm.

Submitted: July 20, 2022
Revised: June 25, 2023
Accepted: September 06, 2023