Using textual volunteered geographic information to model nature-based activities: A case study from Aotearoa New Zealand

Ekaterina Egorova

Faculty of Geo-Information Science and Earth Observation, University of Twente, the Netherlands

Abstract: A boom in volunteered geographic information has led to extensive data-driven exploration and modeling of places. While many studies have used such data to explore human-environment interaction in urban settings, few have investigated natural, non-urban settings. To address this gap, this study systematically explores the content of online reviews of nature-based recreation activities, and develops a fine-grained hierarchical model that includes 28 aspects grouped into three main domains: *activity*, *settings*, and *emotions/cognition*. It further demonstrates how the model can be used to explore the variation in recreation experiences across activities, setting the stage for the analysis of the spatio-temporal variations in recreation experiences in the future. Importantly, the study provides an annotated corpus that can be used as a training dataset for developing methods to automatically capture aspects of recreation experiences in texts.

Keywords: nature-based recreation, cultural ecosystem services, volunteered geographic information, social media, digital geography

1 Introduction

Human activities today are triggering drastic environmental changes around the globe, reflecting an urgent need for a better understanding of human-environment interactions. Spatial information science plays a critical role in meeting this need by providing insights into the spatial and temporal dynamics of social and environmental processes. Indeed, recent advances in geospatial analytics and new sources of spatial data, such as volunteered

geographic information (VGI), put spatial information science in the position of providing active support for sectors ranging from mobility to agriculture [56, 64]. However, this support is heterogeneous, and some sectors, data sources and methods remain beyond the focus of spatial information science.

One example of research that uses only few mainstream data sources and methods is the body of work that leverages VGI for studying human-environment interaction. Derived from sources ranging from photo-sharing platforms to travel blogs, VGI offers data for investigating how people perceive their environments, interact with them through activities, develop sentiments toward particular places, and create and reproduce discourses about them [24]. Much of this research, however, investigates urban settings [13, 32, 40, 71] rather than non-urban, natural environments, which are especially contested where the interests of humans and ecosystems are closely interwoven, as in case with nature-based recreation. On the one hand, nature-based recreation supports the well-being of people and enhances their awareness of the environment, thus motivating public support for biodiversity and conservation efforts; on the other hand, it can bring lasting changes to the natural environments [75]. Understanding recreation behavior is thus crucial for increasing the resilience of human populations, ecosystems, and landscapes. Given that often-used methods for understanding recreation behavior (such as participatory mapping and in situ or household surveys) are generally characterized by a trade-off between the level of spatial and semantic detail and the amount of time invested, VGI could provide a valuable additional source of data.

Currently, VGI research on nature-based recreation usually leverages collections of georeferenced photographs from platforms such as Flickr, with the underlying assumption that the spatial distribution of photographs is as a digital proxy of recreation visits [3, 9, 23, 25, 28, 33, 38]. Some studies also involve the analysis of the content of photographs, which allows the exploration of recreation experiences in a more detailed way. However, information derived from images mostly represents aspects that are simplistic because they must be straightforward to operationalize and identify, such as the type of the recreation activity, elements encountered in the environment, or number of people who engaged in the activity. More nuanced aspects of human-environment interaction—for example, emotional responses—are elusive in image analysis [39]. Using *textual* VGI could address this gap and provide deeper insights into recreation experiences by semantically enriching datasets used in this sector.

This paper systematically explores the content of textual VGI in the form of reviews of nature-based recreation activities from an online platform in Aotearoa New Zealand. To make the results more generalizable, the study compiles and uses a dataset that contains reviews of ten nature-based recreation activities ranging from trekking to fishing. The ten activities vary in their attributes and are undertaken in diverse biophysical environments, both on land and on water. This study is distinct from an ontological inquiry into the concept of activity; rather, its aim is to identify and categorize aspects of recreation activities that are verbalized in online reviews, thus cataloging the richness with which recreation experiences are described in this type of VGI. To achieve this aim, the study develops a model of aspects of nature-based recreation experiences, and provides a coding scheme and a dataset of annotated reviews. It also reports on the differences in the distribution of aspects across the activities, demonstrating that this type of textual VGI goes beyond describing general aspects such as positive and negative sentiments and provides nuanced insights into recreation experiences.

The study contributes to spatial information science first by using a local online platform as a source of data to highlight the opportunities for exploring and diversifying sources of VGI (beyond the mainstream platforms such as Flickr or Twitter). Second, unlike previous work on VGI that performs text analysis mostly using photograph tags, this study explores semantically richer textual descriptions in the form of reviews. Third, text analysis in this study represents a hybrid approach that relies on both close and distant reading techniques and results in more nuanced insights than those achieved through commonly used techniques such as topic modeling. Importantly, the study also presents a side experiment comparing its approach to using topic modeling for a similar purpose to provide a number of methodological insights. Fourth, a significant contribution of the study is the documentation of the linguistic encoding of aspects of recreation experiences in the form of a coding scheme and an annotated dataset that can be used for multiple purposes. Fifth, the study enhances the understanding of how activities influence human conceptualizations of the environment beyond urban settings, and provides systematic insights into the aspects of human-environment interaction in the nature, including aspects related to emotion and cognition. The latter insights make an especially significant contribution to spatial information science, which has mostly used VGI to describe more tangible aspects of human-environment interaction, such as place elements or human activities [52,71].

Importantly, by systematically investigating the content of the data, the study helps infer research purposes that this type of VGI can serve to thus be of interest to research on nature-based recreation and, more generally, cultural ecosystem services. In the long run, spatial information science can equip these research sectors with a toolbox for extracting and integrating geographic information from multimodal VGI, providing empirical data to complement and advance existing research, resulting in richer, more qualitative, and more pluralistic perspectives on nature-based recreation experiences.

The immediate areas of applications for the study are as follows. First, the annotated dataset can be used as training data for developing methods for the automated annotation of aspects in text, whereby both supervised machine learning and rule-based approaches are possible. Developing such methods will allow researchers to leverage large textual datasets for studying human-environment interaction. Second, the proposed model can serve data integration purposes. In particular, the annotated dataset can be used to derive and expand lexicons for each of the aspects to help classify tags generated by image labeling tools such as Google Cloud Vision, which are being increasingly used in research on nature-based recreation and ecosystem services more generally [22]. Thus, this study makes an important step toward the extraction and integration of spatial information originating from diverse VGI data sources. As such, it contributes to the development of a toolbox capable of leveraging large multimodal datasets to model and explore the spatio-temporal variation of aspects of human-environment interaction in various contexts.

The remainder of this article introduces related work on text in spatial information science, on the role of VGI, and on nature-based recreation in Section 2, before describing the data and methods in Section 3. The aspects and their linguistic encoding are described in Section 4, followed by the analysis of the variation of aspects across activities in Section 6. The results of the study are discussed in Section 7 before a short conclusion.

2   Related work

Three lines of work are particularly relevant to this study. Research that leverages textual descriptions to explore geographic knowledge contextualizes methodological aspects of this study and its applications for the retrieval of geographic information. Research into nature-based recreation sets the thematic context for this study, and frames its potential applications in the area of recreation behavior modeling. Research that leverages VGI to examine human-environment interaction highlights the opportunities offered by this data source.

2.1   Textual data in spatial information science

Modeling spatial processes and human-environment interaction is one of the key tasks of spatial information science. Diverse data sources are leveraged for this purpose, including textual descriptions: formal texts such as traffic codes and glossaries reflect publicly shared knowledge [37, 60, 73], elicited spatial descriptions unveil the conceptualization of space in particular contexts such as wayfinding [15], and VGI provides insights into naturally occurring discourse across large spatio-temporal scales.

Methodologically, research into spatial language in text is multifaceted, and falls on a broad spectrum between close and distant reading techniques. The importance of geo-referencing textual documents for further analysis is reflected in the long-standing line of work that develops coding schemes for capturing spatial information—for example, nominal references to place—in texts [42, 53]. Development of the conceptual inventory of such schemes is based on examining real texts and making adjustments through multiple rounds of annotation. In a similar manner, techniques such as cognitive discourse analysis represent highly iterative processes [65], but offer powerful insights into the way humans conceptualize surrounding geographic space across a variety of contexts, with important practical applications. Analysis of topographic feature glossaries, for example, has unveiled how classes such as *processes* and *human intentions* reflect human environmental experiences and can inform queries for topographical data analysis and mapping [73]. Analysis of VGI route directions has showed which types of *landmarks* and *decision points* support navigation in mountain areas [18], while further research has suggested ways of integrating this information for rescue operations [70]. Content analysis can also be facilitated through corpus linguistics techniques such as frequency-based analysis, identification of synonyms, or analysis of co-occurrences, performed on text pre-processed with part-of-speech tags and syntactic relations [37,60]. Applied to VGI images tags, this type of analysis has revealed how three place-related facets—*elements*, *qualities*, and *activities*—capture the way in which people describe places and can be used to explore the differences between places [52]. Finally, distant reading techniques such as topic modeling and measures such as cosine similarity offer ways to analyze vast datasets, allowing researchers to link textual descriptions to their spatial footprints and thus explore the data in a spatially explicit manner [2, 32, 35].

Close reading techniques require a meticulous, iterative approach and are often perceived as time-consuming in the era of big data and machine learning. However, they produce structured, annotated data that can generally be used for training in geographic information retrieval, leading to the systematic extraction and integration of spatial concepts [17]. Distant reading techniques such as topic modeling, on the other hand, obscure

structural properties of discourse, do not allow to disambiguate between word senses, and their outputs have less potential for being reused for other purposes. They do, however, allow researchers to discover topics and gain insights into vast datasets relatively quickly, and they can be easily applied to new datasets. In general, there is a rising consensus that close and distant reading techniques have trade-offs between the level of semantic detail and the scalability of analysis, and neither is advantageous over the other in terms of accountability and replicability [7].

Provided the chosen methodology is adequate for the research purpose at hand, textual descriptions allow researchers to explore a wealth of geospatial knowledge. The variety of sources ranging from topographic dictionaries to VGI offer possibilities for unveiling how space is experienced and perceived by different actors in a vast variety of contexts. Models of human-environment interaction derived from such data "reflect major concepts of environmental experience, language pragmatics, cognitive imagination and embodiment" [73, p. 192] while also serving as a basis for the retrieval and communication of geographic data and information.

2.2 Nature-based recreation as a type of human-environment interaction

Modeling the structure of nature-based recreation experience has been addressed through multiple approaches ranging from phenomenology to netnography and through a variety of data, including pre- and post-trip interviews, in situ surveys, and telephone interviews [48, 59]. While some studies focus on certain recreation activities such as camping, others explore the recreation potential of areas and ecosystems (for example, rivers) [20]. A distinct line of work explores recreation behavior of groups (e.g., indigenous youth, elderly people), and investigates how the perception of experiences vary among tourists and local residents, as well as between genders [19, 48].

Several major themes emerge from these studies. First, since nature-based recreation activities differ in their attributes, one widely used typology differentiates between *appreciative-symbolic* (activities directed toward appreciation of the natural environment, e.g., horseback riding), *extractive-symbolic* (activities characterized by the quest for trophies extracted from the natural environment, e.g., fishing), *passive free-play* (activities requiring little effort, e.g., quiet boating), *sociable learning* (clearly social activities where the primary motivation is the social interaction, e.g., nature study), and *active-expressive* (activities that do not require a particular natural settings, e.g., frisbee) [30]. The choice of an activity is rooted in intrinsic motives such as enjoyment of nature, exhilarating physical exercise, appreciation of cultural resources, or socialization [46]. Thus, recreationists aiming at enjoying nature are attracted by natural features and activities that involve nature observation and study [45]. Those driven by novelty, adventure, a simpler lifestyle and escape from the ordinary are attracted by undisturbed environments and a chance to see wildlife, while those motivated by physical exercise prefer nature-based sports [50].

Apart from the general attributes of an activity, a recreation experience is shaped by multiple further aspects of human-environment interaction, including the perceived quality of the natural and social settings. Regarding the natural settings, many recreationists prefer wild, clean, isolated, and natural environments, whereby landscape diversity and the presence of scenic vistas play a key role [4]. Regarding the social settings, many recreationists appreciate the opportunity to have a good time with companions, and to communicate with other nature lovers on the route [5, 8]. In adventurous commercial group activ-

ities, participants also highlight the role of competent, safe, knowledgeable guides, as well as friendly, sharing, considerate, and sociable team members [4, 59]. In activities such as ecotourism trips, important aspects of experiences, according to participants, are the quality of facilities and food, as well as the choice of secondary activities [59]. The overall recreation experience is formed and remembered through the prism of emotions that it evoked. Previous studies report on the importance of senses of isolation, serenity, introspection, spirituality, personal growth, and renewal of self [4]. Place attachment and the sense of place emerge as salient aspects of nature-based recreation activities that take place repeatedly in particular locations [20, 34]. Meanings associated with such locations can change over time, unveiling a dynamic nature of the sense of place that reflects physical changes to locations but also to "participants' recreational interests; and broader life changes which reflect the interwoven nature of participants' recreation-life experience" [20].

Understanding recreation behavior patterns and aspects that shape recreation experiences is crucial for minimizing the impact on the environment while maximizing the benefits for the humans, thus increasing the resilience of both human populations and ecosystems. Given the resource-intensive nature of traditional methods such as surveys, VGI offers valuable additional sources of data to enhance our understanding of human-environment interaction in the context of recreation.

2.3 Volunteered geographic information and natural settings

VGI has recently established itself as one of the major sources of data in spatial information science. The geospatial component that VGI often contains in an explicit or implicit form allows researchers to explore a wealth of thematic and temporal information linked to various locations. This multi-dimensional information of unprecedented granularity, not available in more traditional spatial data, provides insights into human behavioral patterns such as urban mobility, as well as into the spatio-temporal unfolding of events such as natural disasters [55, 61, 77]. Much of this work is framed around the concept of place, whereby various forms of textual VGI are used to extract "soft" information in the form of local place knowledge and feelings about neighborhoods, model place properties, or identify potential for leisure activities, with topic modeling being a wide-spread methodology [32, 40, 52, 71].

VGI-based research in nature-based recreation mostly represents the analysis of images from photo-sharing platforms such as Flickr, Panoramio, Instagram and Geograph [3, 9, 12, 23, 25, 28, 33, 38, 41]. Since taking a photograph can be seen as some measure of the photographer's individual preference for the depicted subject matter, the spatial distribution of images can be interpreted as a digital proxy of the general "worthiness" of locations [3, 57]. There is, however, a remarkable variety in the interpretation of this worthiness, and image frequency distributions have been used as a proxy of the distribution of cultural ecosystem services [11, 25], landscape aesthetics [9, 38], cultural attachment [23], recreation potential [12, 21], and visitation rates [26].

Including image content analysis allows researchers to explore more nuanced aspects of human-environment interaction in natural settings. Image classification schemes either rely on existing frameworks or are defined in a bottom-up way, depending on the research question at hand. Thus, classifying images according to the Cultural Ecosystem Services framework (with categories including *recreation*, *aesthetics*, *scientific and educational*, *cultural heritage and identity*, *spiritual and religious*, *inspirational*) allows one to explore the spatial distribution of the use of particular cultural services [11]. Applying the Final Ecosystem

Goods and Services Classification System offers a way to further differentiate between images that depict a biophysical service, and images that depict humans benefiting from a service [3]. Image content analysis offers an opportunity to explore more nuances types of cultural uses of natural settings (*nature appreciation*, *landscape*, *social recreation*, *fishing recreation*, *history*, *research* and *infrastructure appreciation*) [57], types of interactions (*physical and experiential interactions* and *intellectual and representative interactions*) [66], types of natural environments (*spontaneous nature*, *productive nature*, *designed nature*, *domesticated nature*, and *biocultural nature*) [25]. Image classification can be also performed in an explorative way, as in [28], where the authors derive aspects of recreation activities in a national park based on the *presence or absence of people*, *activities*, *landscape*, *animals*, and *infrastructure*, followed by a more nuanced classification into revealing subcategories.

The fact that images often include explicit or implicit geolocation and time marks allows for the analysis at various spatial [9, 23, 38], and temporal [33, 41, 69] scales. Moreover, the spatio-temporal granularity of data opens a possibility for exploring factors that have spatial covariance with image distributions, and for linking various aspects of recreation activities to biophysical (distance to the ocean, clarity and depth of water, cumulative viewing area) [33, 38, 69], infrastructural (distance to roads, trails, urban areas, cultural sites, ski resorts) [11, 66, 72], and socioeconomic (users' home locations) variables [12, 21]. Following the identification of best explanatory variables for the spatial distribution of recreation patterns, this line of research suggests areas that have the same characteristics but are underused, thus showing a hidden potential and capacity for recreation activities.

Comparison of recreation patterns derived from image-based VGI to official survey data confirms strong correlations between the two data sources in terms of time series, spatial distribution, and types of recreation activities although the strength of correlation varies across VGI types [3, 28, 39, 41]. However, VGI should not be leveraged as a reliable proxy for the full range of aspects of recreation experiences—for example, the perceived wilderness of natural settings was found to be negatively correlated with image-based VGI [39].

2.4   Research gap and questions

The research gap to be addressed in this study is threefold. First, much of the research in spatial information science that leverages VGI focuses on urban settings, while natural settings have traditionally gained less attention. Second, the VGI-based studies that do investigate natural environments mostly analyze the distribution of georeferenced photographs from platforms such as Flickr, and opportunities for extracting semantically rich information from textual sources remain largely unexplored. Third, text analysis of VGI in spatial information science is confined to unsupervised methods for topic discovery, and while these methods are easily applicable to new collections of texts, they are less suitable for capturing semantically nuanced concepts such as the sense of place. Related to this, with the general prevalence of unsupervised methods, spatial information science is currently lacking coding schemes and annotated datasets that could be used for training supervised models to capture a rich set of semantically nuanced aspects of human-environment interaction.

Addressing these gaps will enrich the current methodological toolbox applied by spatial information science to extract the wealth of "soft" information from VGI, building and curating geographic text analysis resources. Exploring descriptions of human activities in non-urban settings will enhance our understanding of how human-environment interac-

tion varies across different contexts. Importantly, it will put spatial information science in the position of enriching datasets used by the cultural ecosystem services and nature-based recreation research communities. This study aims to make the first step in this direction by focusing on the following research questions:

1. *Which aspects of nature-based recreation experiences are documented in online reviews?*
   To systematically explore the scope and content of the data, this study develops a fine-grained hierarchical model of aspects of recreation experiences. In doing so, it relies on methodological triangulation and combines close and distant reading techniques to capture both frequently mentioned as well as more nuanced, specific aspects of recreation activities. The important tangible outcome of the study is the coding scheme and the annotated dataset that can be further specified and expanded by the involved research community.

2. *Do aspects of nature-based recreation experiences differ across type of activities, as reflected in online reviews?*
   This question explores the semantic richness of the reviews. The working hypothesis is that online reviews as a type of textual VGI are detailed enough to reflect differences in types of activities. A confirmed hypothesis will indicate that online reviews go beyond reflecting general aspects of recreation experiences such as positive and negative sentiments, and that this type of textual description can provide nuanced insights into human-environment interaction in the context of nature-based recreation experiences.

3   Data and methods

3.1   Data

The source of data for this study is https://www.rankers.co.nz. Set up in 2007 to collect trusted travel reviews, this locally owned resource is a major travel application about and for Aotearoa New Zealand, used by both locals and tourists. Reviews are in English, and the scope of reviews covers five major categories: activities, attractions, accommodation, transport, and regions. Within each activity, one can further search for a particular experience—for example, the search engine for trekking experiences includes parameters such as the trek's length, duration and difficulty.

A web page related to a particular experience (e.g., a diving spot or a biking route) contains three major blocks of information: general description, photographs, and reviews. The structure of the general description of the experience varies across activities, but some common elements include its title, expected duration, and location. The latter is presented on four levels of granularity: island, region, area, and coordinates. For example, the location section of the Abel Tasman Coast Track includes "South Island," "Nelson region," "Abel Tasman," "-40.9958362042124, 173.004884719849." A review typically includes the name of the reviewer, their country of origin, a rating of the experience on a scale of 1 and 10, the date when the review was added, the date of the experience, and the text of the review.

This study focuses on land-based (caving, biking, horse trekking, and walking/trekking) and water-based activities (diving, kayaking/canoeing, fishing, swimming with mammals, rafting, and sailing), and the corresponding sections of the web page were
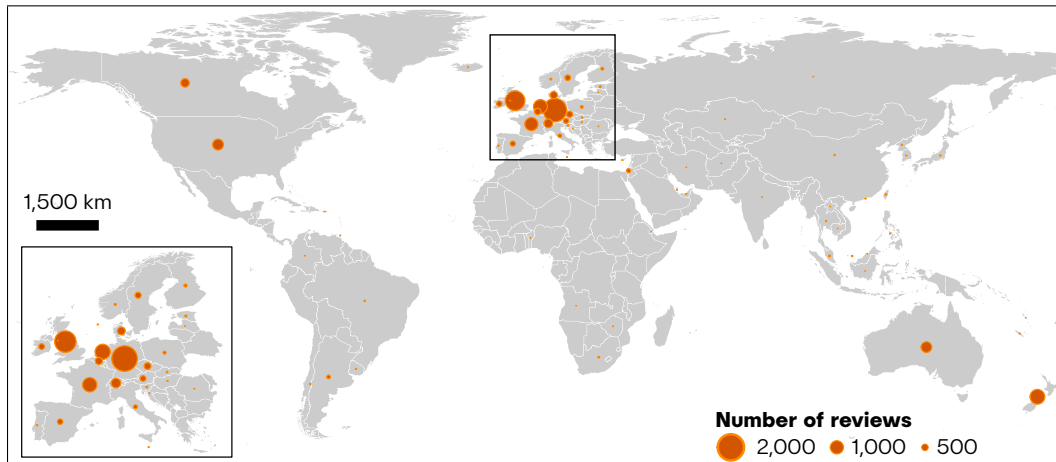
Figure 1: Reviewers' countries of origin.

crawled using Scrapy[1]. The data was stored in a database consisting of two tables: one table storing experience-related information (type of activity, experience ID, title and subtitle, the number of reviews and the coordinates), another table storing the actual reviews of each particular experience (experience ID, review ID, reviewer's name and country of origin, rating, review date, experience date, and review text). Characteristics of the sub-corpora[2] are represented in Table 1.

Activity	N reviews	N words	Avr N words/review
Land			
caving	179	5,292	29.56
biking	582	23,764	40.83
horse trekking	186	7,437	39.98
walking/trekking	10,666	238,959	22.4
Water			
diving	176	6,447	36.63
kayaking/canoeing	1,203	41,252	34.29
fishing	193	6,736	34.9
swimming with mammals	234	6,660	28.46
rafting	400	12,084	30.21
sailing	291	11,642	40

Table 1: Activity-based sub-corpora and their characteristics.

Out of 14,110 reviews in total, only 330 (3.8%) do not indicate the reviewer's country of origin. The total number of unique countries is 63, but the distribution is heavily biased toward a few of them (see Figure 1): 90.2% of reviews are written by tourists from 16

---

[1]https://scrapy.org
[2]"Sub-corpus" in this paper refers to a collection of reviews of one type of activity, for example, caving; there are thus ten sub-corpora in the corpus.
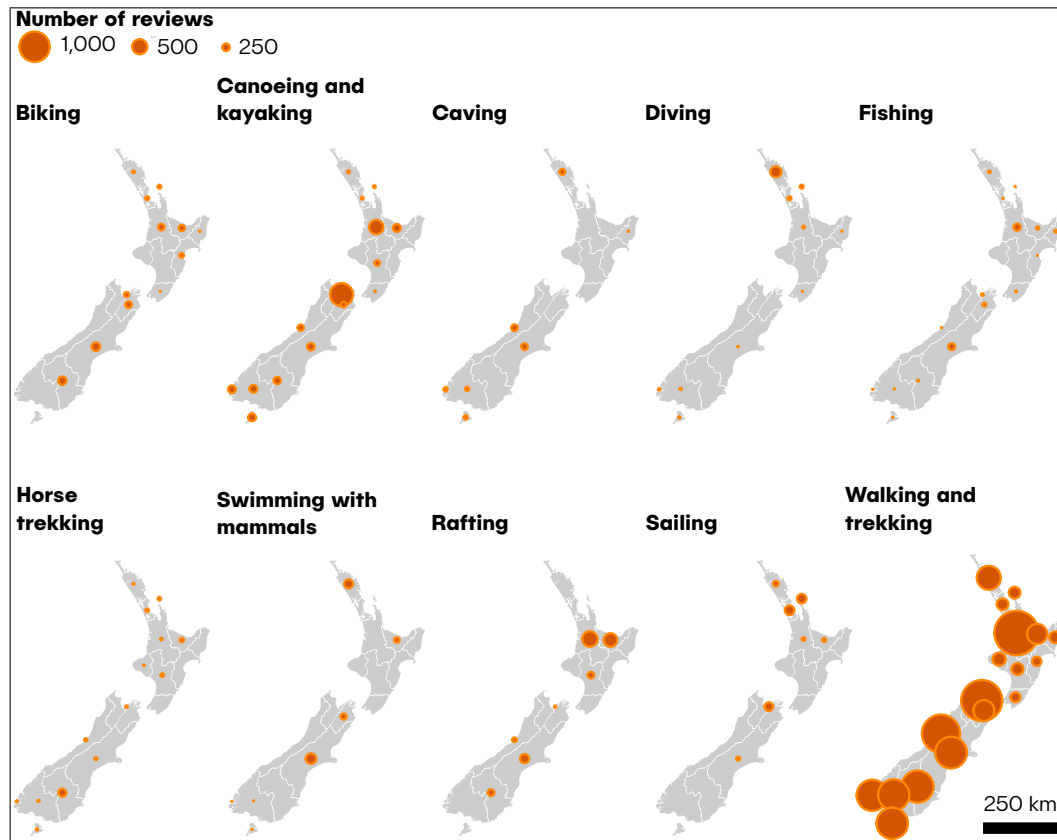
Figure 2: Number and distribution of reviews per activity and geographic region.

countries, and reviews from tourists from Germany alone constitute 26.6% of all reviews, followed by those from the UK at 11.4%. However, this distribution appears to largely reflect the actual trends in the special interest sectors of New Zealand's international tourism market where longer walks are particularly popular among German and UK visitors [67]. Given that the whole corpus is clearly dominated by reviews of walking/trekking (see Figure 2), reviewers on the platform can be seen as a representative sample.

In terms of contribution inequality, there is no bias toward a small number of prolific contributors. The majority of reviews—87.1%—are from reviewers who have written between 1 and 4 reviews. Reviews from more prolific reviewers (who have left 11 to 26 reviews) only represent 2.16% of the data. There is also no evident bias in the spatial coverage of the reviews: locations of reviewed experiences are spread across New Zealand regions relatively evenly, as shown in Figure 2.

3.2   Methods

The major methodological steps constituting the analysis in this study are the development of the model of aspects of nature-based recreation experiences (i.e., identification and grouping of aspects) and the examination of the variation of these aspects across activities.

The model development involved methodological triangulation by combining both close (content analysis of a sub-corpus) and distant (frequency-based analysis of the whole corpus) reading techniques.

The aspect identification involved a manual content analysis and a frequency-based analysis of the reviews. For the manual content analysis, 200 reviews (20 for each of the sub-corpora) were randomly selected and annotated with POS-tags using the Stanford coreNLP parser [43]. Nouns, verbs, adjectives, and adverbs were marked as candidates for aspects for the ease of manual content analysis. The goal of the manual content analysis was to identify whether these candidates represented aspects, and to capture aspects encoded through more complex structures (e.g., *also for kids*). For the frequency-based analysis, lists of most frequent nouns, verbs, adverbs, adjectives, bigrams, and trigrams were compiled for each sub-corpus, using a cut off frequency of 50. The lists were further merged to remove duplicates and reduce the effort for the manual inspection. Final lists were examined manually, whereby relatively straightforward terms (e.g., *splendid*, *path*) were identified as aspects without further exploration of the context, while lexemes offering multiple contextual interpretation (e.g., *day*, *water*) were further inspected in context by randomly selecting ten sentences where they were used. The goal behind this step was to identify aspects that were not captured in the sample of 200 reviews used for manual content analysis.

The labeling and grouping of aspects into a model followed several principles. First, the process aimed at a hierarchically structured model to ensure coherence and systematic coverage of aspects of human-environment interaction. Second, it aimed at a fine-grained model that is consistent with reliable coding but leaves the decision of the level of analysis to the data analysis stage and thus allows for a flexible reanalysis for a variety of research purposes. Third, the study aimed at a model that allows for adding new branches, provided further aspects emerge in textual descriptions of other types of nature-based recreation activities or localities. The development of the model represented a highly iterative process, characterized by multiple modifications.

The evaluation of the model and the method was performed with the help of the following experiments. First, an interannotator agreement analysis was conducted to determine consistency of the coding scheme using Krippendorff's alpha [27]. Second, the model was validated on "unseen" data from 70 random reviews of long-distance trekking (the sub-type of the walking/trekking activity that was not represented in the corpus of 200 reviews used for manual content analysis). Third, to compare the adopted resource-intensive method for identifying and categorizing aspects to a more computational process, a Latent Dirichlet Allocation (LDA) was performed using MALLET, an off-the-shelf package for machine learning applications to text, where the optimization of parameters such as alpha and beta is built-in into the package [47]. The optimal number of topics was chosen based on the coherence measure, and several LDA runs were performed on different subsets of the corpus to achieve the best interpretability and semantic granularity of topics (more details in Section 5). Discovered topics were mapped onto the corresponding aspects in the model, producing a basis for comparing the methods and their outcomes.

The variation of aspects across activities was explored through a series of chi-square tests of independence which compared the distribution of frequencies of aspects across activities.

Figure 3: Example of an annotated review.

First, a pairwise comparison was performed to identify activities that are most (dis)similar to each other; second, each activity was compared against all others.

4   Aspects of nature-based recreation activities

The model includes 28 aspects grouped into three main domains: activity, settings, and emotions/cognition (see Figure 4). This section describes each of the domains and its aspects in detail, while more examples of their linguistic encoding and use in the context are found in Table 3.

4.1   Activity

This domain encompasses three major sub-domains related to any activity: activities and actions, activity-related artifacts, and activity characteristics (duration and difficulty).

4.1.1   Activity

Descriptions of ACTIVITIES AND ACTIONS are naturally an important part of the reviews, and they expose a remarkable variation in granularity. First, apart from the ten primary activities in the corpus, reviews frequently mention secondary activities such as bird watching and wine tasting that one can engage into alongside with the main activities. Activities are encoded by various forms of verbs and nouns, as well as verb and noun phrases—e.g., *bike tour*, *biking*, *ride*, *horse trek*. Second, reviews include descriptions of actions undertaken as part of recreation activities. These actions can be both activity-specific (occurring in the reviews of one activity only) as well as more general (found in the reviews of several activities).

Activity-specific actions often represent physical actions involving locomotion—for example, paddling in the context of kayaking/canoeing, scrambling in the context of caving, swimming in the context of diving. These descriptions can be very detailed, with reviewers meticulously documenting their experience along segments of the route, as in "I couldn't

```
/
└─Activity
  └─Activities and actions
  └─Activity-related artifacts
  └─Activity characteristics
    └─Difficulty
    └─Duration
└─Settings
  └─Biophysical settings
    └─Physical settings
      └─Spatial properties and relations
      └─Natural and geographic features
      └─Settings in general
    └─Flora and fauna
    └─Environmental processes
      └─Time of the day and corresponding processes
      └─Seasons, weather and conditions
  └─Infrastructure
    └─Free access
    └─Facilities and their quality
    └─Services and their quality
  └─Social settings
    └─Social relations
      └─Team
      └─Personnel
      └─Other people
    └─Social interaction
  └─Cultural-historical environment
└─Emotions and cognition
  └─Sense of satisfaction
  └─Emotions
    └─Sense of aesthetic appreciation
    └─Sense of experiencing something special
    └─Sense of having an adventure
    └─Sense of having fun
    └─Sense of relaxation
    └─Sense of safety
    └─Sense of wilderness
    └─Other emotions
  └─Cognition
```
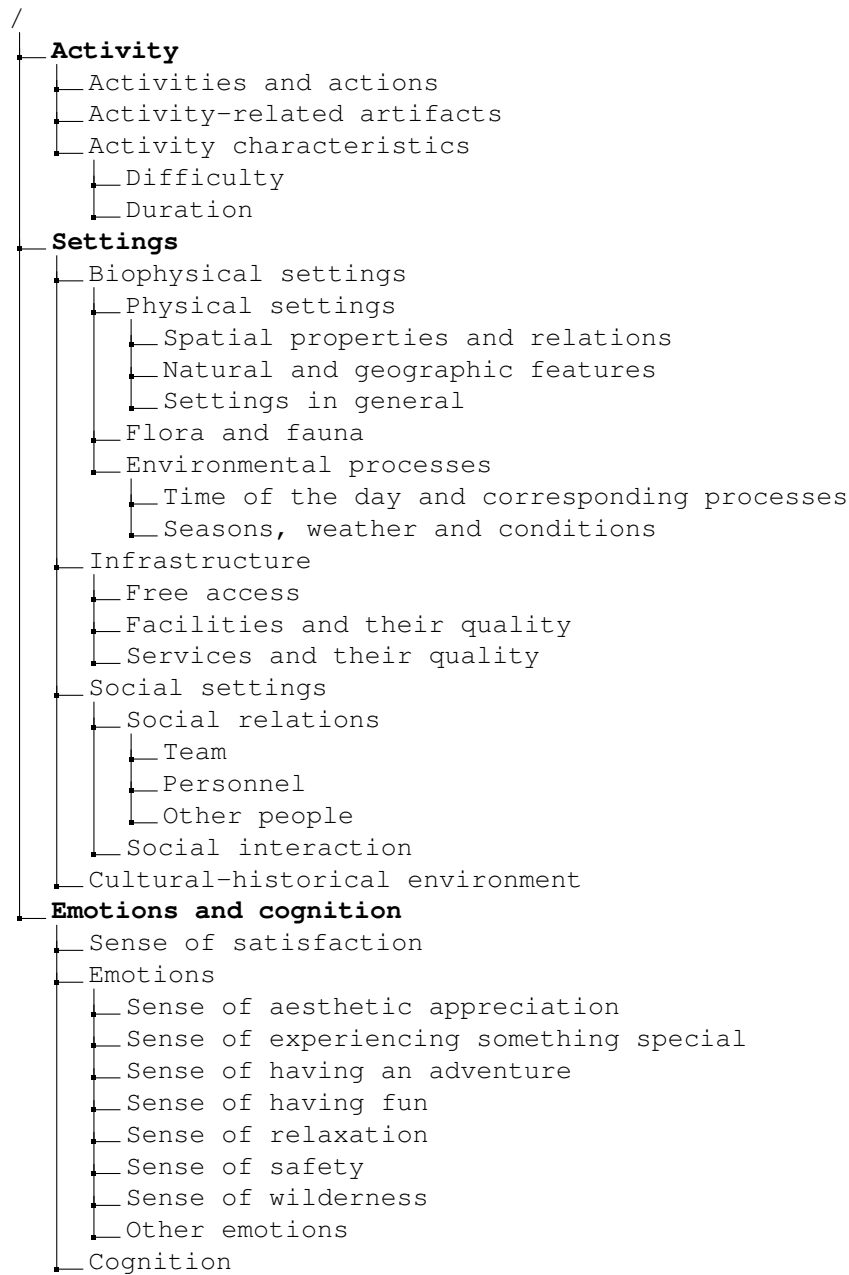
Figure 4: Model of aspects of nature-based recreation activities.

get over a waterfall in the middle of the run, tried to climb around it but I fell off like an id-iot." Activity-specific actions can also include actions of a higher order, such as chartering a yacht in the context of sailing, or doing a certification course in the context of diving.

Actions that are not specific to a recreation activity type relate to trip preparation: planning activities, booking accommodation, and bringing or hiring equipment. Further examples of activities that occur in multiple sub-corpora relate to navigation in the environment, such as getting to, discovering, exploring, and leaving locations, as in "It took me about 45 minutes to get there." Another aspect that shapes the recreation experience is the possibility to get close to and/or watch natural phenomena, ranging from stars to landmarks and wildlife. This aspect is represented by verbs and phrases such as *look for*, *come across*, *encounter*, *get close to*, *watch*, *see*, *take photos*, as in "watched a few dolphins from a distance." Finally, reviews reflect the importance of stopping, resting, or eating.

ACTIVITY-RELATED ARTIFACTS include references to gear and equipment such as trekking sticks, flashlights, torches, warm clothes, masks, or life jackets. Reviewers often describe means of assisted locomotion—bikes, rafts, boats, canoes, kayaks, or sailing yachts—as well as their qualities, as in "stable kayaks," "well-maintained bikes," or "immaculate equipment."

DURATION can refer to the duration of the experience, as in "It was a full day sailing," as well as to its particular segments or sub-activities, as in "The introduction was a bit too long." Duration expressions expose a peculiar variation in temporal granularity, ranging from minutes when the reviewers describe some exciting short-term experiences such as a short walk in a cave to see the glowworms, as in "Fantastic for the five minutes it lasted," to weeks for experiences such as sailing trips, as in "Two week sailing trip." Activity duration is also frequently mentioned in the context of optimal time that one can spend on it without getting bored or exhausted, as in "Two to three days maximum," "4 days is a perfect time." Further, this aspect is an important part of the experience of commercial activities, as in "Though I was told that we'd have hours of time on the water, I ended up rafting actively for only an hour or so." In terms of the linguistic encoding, duration can be represented both qualitatively, as in "a short walk," "a long introduction," as well as quantitatively, as in "2.5-hour trip," "30-minute walk." In reviews of activities that involve substantial movement in space, such as trekking and biking, duration can be encoded through references to distance, as in "not so far on a gravel road."

DIFFICULTY of the activity is another important aspect in the recreation experience. It is often encoded through adjectives such as *easy*, *ridable*, *tough*, *doable*, *strenuous*, as well as through descriptions of the type "beginners welcome," "also for amateurs," "for all ages." Apart from qualitative descriptions, reviews of some activities—mostly, rafting—include quantitative scales of difficulty assessment, as in "5 Grade waterfall."

4.2 Settings

There are four categories of settings mentioned in the reviews: biophysical settings, infrastructure, social settings, and cultural-historical settings.

4.2.1 Biophysical settings

Biophysical settings include aspects related to the physical environment, flora, and fauna. The physical environment includes three groups of aspects that can be distinguished based

on the level of granularity: spatial properties and relations, natural and geographic features, and settings in general.

SPATIAL PROPERTIES AND RELATIONS, sensed multimodally—through locomotion, visual, and olfactory perception—are an important part of recreation experiences. In land-based activities, reviews often describe the slope angle and the quality of the terrain (e.g., "slippery when wet") or the size and the extension of natural features (e.g., "a huge cave"). Where river crossings are involved, the level and depth of the water is important, as in "The river was not at a good water level." Experiences of water-based activities are also influenced by spatial properties: the extension of waterfall drops is important in rafting, as in "a 7-meter waterfall drop;" water visibility becomes crucial in diving, as in "Slightly underwhelmed, visibility was not the best on the day." This aspect also includes descriptions of colors, soundscapes and olfactory landscapes, as in "incredibly blue," "pretty quiet." Finally, apart from spatial properties, also spatial relations are frequently mentioned in the reviews; proximity to environmental elements—as encoded by lexemes such as *close* or *distant*—is particularly important in many activities, be it proximity to a landmark, a viewing platform, or wildlife.

NATURAL AND GEOGRAPHIC FEATURES are the most frequently mentioned elements of the biophysical environment, varying from the perspective of scale from oceans and mountains to stalactites and rocks. A closer inspection of linguistic contexts unveils a range of roles that natural and geographic features play in recreation activities. They can represent segments of the route that had to be followed or crossed, as in "Some crossings were difficult, especially the large pool of water midway through;" they can represent the environment where a main activity takes place, as for example with caving "Walk through a limestone cave;" they can also be mentioned as elements of vistas, as in "Watched the sun set over the sea," or landmarks, as in "You are rewarded at the end with beautiful Cathedral Cove." Geographic features of a larger scale such as islands are mentioned in the descriptions of the general whereabouts of an activity, most commonly in the context of diving, as in "I went diving at the Poor Knights Islands." Importantly, the same lexeme can have different meanings in the context of different activities, and thus fulfill different roles—for example, water is an unwanted obstacle in caving where reviewers encourage to check water levels before going. Water is a setting for an activity in swimming with mammals and in diving, as in "A great day out on the water," or "Made us feel comfortable in the water." Water can also be mentioned as a landmark in the vista, as in "A nice track with views to the water."

SETTINGS IN GENERAL is an umbrella term used in the model for more general descriptions of the surroundings. It includes references to the landscape, both in land-based and in water-based activities (e.g., "spectacular landscapes," "beautiful underwater world"), to the scenery and views (e.g., "peaceful scenery," "wicked views"), to places (e.g., "lovely place," "nice spots"), as well as to the area of the activity location (e.g., "a must do if you are in the area," "perfect location"). Examples of rarer descriptions are *back yard*, *down south*, or *land*, as in "Nice people that love showing off their back yard."

FLORA AND FAUNA is another prominent aspect of nature-based recreation activities—reviewers document encountered species and comment on the presence, quantity and diversity of wildlife and vegetation (e.g., "lots of fish, different species," "not so many dolphins," "a heap of marine life"). Also, the behavior of the encountered species is often described in the reviews, as in "The dolphins were feeding." TIME OF DAY when the activity was undertaken (e.g., dawn) is often reported together with corresponding natural

processes such as sunrise or tide, as in "Boat trip through the French Pass at full ebb tide." SEASON, WEATHER, CONDITIONS is another frequently mentioned aspect that shapes the nature-based recreation experience, as in "Bring clothing for warm and cold weather."

4.2.2 Infrastructure

Infrastructure includes mentions of the free access, facilities, and services.

FREE ACCESS—that is, the absence of an entrance fee, or a possibility to undertake an activity on one's own rather than as part of an organized commercial tour—appears to be an important aspect of a recreation experience. While some reviews highlight the financial side of the matter, as in "Free is a very good price!," others appear to appreciate the freedom of exploration and a more private contact with the nature, as in "Exciting to be able to go by yourself and see the caves." This aspect is mostly mentioned in the context of caving, since many famous caves in New Zealand can only be attended with an organized tour. However, also in horse trekking and kayaking/canoeing, being able to leave the organized group of riders or hire a kayak without having to join a group is mentioned by reviewers as an aspect that can shape the experience, as in "Independent canoe hire—brilliant."

FACILITIES AND THEIR QUALITY is another important aspect of recreation experiences, whereby reviewers comment on the presence or absence of certain facilities, as well as their quality and level of maintenance. In caving, biking, and walking/trekking, most frequent comments refer to signage, water fountains, viewing platforms, and their quality, as in "The track itself was excellent, well-groomed and almost all rideable." Showers gain importance in water-based activities such as diving and rafting, as in "They said there was a hot shower on the boat but there wasn't." Examples of other facilities mentioned include parking lots, camping spots, restaurants, shops, and kiosks.

SERVICES AND THEIR QUALITY become an important element in the experience of commercial activities, whereby reviewers frequently comment on the level of organization and professionalism of the company in general (e.g., "very professional company," "poor communication," "careful preparation"), and the trip in particular (e.g., "well-organized," "no hassles," or "went smoothly"). Interestingly, apart from the company's attitude toward customers, reviewers mention the company's attitude toward the natural environment and wildlife—as in case with dolphins in "They really cared about providing their passengers with a memorable, enjoyable and safe(!) day, while ensuring dolphin welfare was a key priority." The guides' professionalism, responsibility, and a personable attitude also have a key impact on the experience—reviews contain multiple descriptions of the type "The guides were really professional," or "The guides were skilled and strong." Examples of other services mentioned in the reviews are transport, food, accommodation, hire services, choice of tours, or insurance (e.g., "good transport," "great lunch," "great accommodation," "bike hire service," "good choice of tours," or "no insurances"). Finally, reviewers frequently assess the value for money, as in "overpriced," "reasonable value for money," or "too commercial, too expensive."

4.2.3 Social settings

Social settings and the quality of social interactions emerge as another important aspect shaping recreation experiences.

TEAM includes references to the close circle of familiar people who share the experience: partners, family, or friends. The possibility to engage into an activity together with children

appears to be particularly important in some of the recreation activities, as in "great with kids" or "also for kids."

OTHER PEOPLE are references to people met during the recreation experience. These can be people that engage in a different recreation activity on the same route—for example, walkers and cyclists, as in "Be careful at each corner to avoid collision with walkers or other cyclists." In commercial activities, these are people who are put together on the same team, and this aspect appears especially relevant in activities where the performance of each member influences the experience of the rest, as becomes evident in example "Got put in a ride with a family who had never ridden before," but also in the context of potential overcrowding, as in "too many people," "a crowded boat," "too busy." Finally, this aspect also includes descriptions of encounters with the local people, as in "Our guide Hillary knew many interesting people."

PERSONNEL—the crew, dive masters, guides and instructors—is a salient aspect of the social settings in commercial activities, as in "very professional guides" or "enthusiastic fun staff."

SOCIAL INTERACTION, its nature, and its quality are further elements of social settings that can shape the overall experience. Recreationists appreciate an easygoing atmosphere and the friendliness, hospitality, helpfulness, and enthusiasm of other people and personnel. Activities and actions related to social interaction include following trail etiquette (e.g., "taking care to defer to cyclists"), making friends (e.g., "got buddied with another lone diver"), laughing and talking (e.g., "great guides that make you laugh"), helping and caring (e.g., "local guy was fantastic in helping us"), and sharing knowledge, places, stories (e.g., "a great mix of scenery, storytelling and adrenaline").

4.2.4 Cultural-historical settings

Cultural-historical settings represent both tangible (i.e., physical elements of the landscape) and intangible (i.e., history and folklore, traditions, and practices) elements encountered in the described region. One frequently mentioned aspect of cultural-historical settings is the cultural heritage of indigenous groups—for example, cave sculptures, Māori carvings, and traditional ceremonies. Reviewers also comment on learning about the history of the region in general, sometimes specifying past events, such as a tragedy at a coal mine. Another important element of recreation activities is the encounter with local cultural practices such as mussel farming, as in "Learned a lot about the impact of mussel farming in the region." Physical elements that belong to cultural-historical settings range from small towns and lighthouses to wineries and shipwrecks, and can be encountered during the main activity (as in "This hike takes you by the prawn farm, power facility and hydro dam") or represent destinations in side trips (as in "She took us on several trips to view off the track towns").

4.3 Emotions and cognition

This domain groups aspects of recreation activities that represent descriptions of emotions and various forms and ways of knowing.

4.3.1 Sense of satisfaction

Given the genre of texts—reviews—the corpus includes multiple references to positive and negative sentiments of various intensity. They are mostly encoded by synonyms of adjec-

tives *good* and *bad*, and verbs *like* and *dislike*. Examples of more idiosyncratic expressions include "once in a life time," "made our trip," "robbed the experience for me," or "hooked for life." These general sentiments can relate to various aspects of the activity, including the experience in general (e.g., "had a wonderful time," "one of the 10 top walks") and its particular elements (e.g., "great to see seal pups at play this time of year"), activity-related artifacts (e.g., "boat OK," "got some nice fish"), settings (e.g., "great views," "cool places to relax"), weather (e.g., "nice weather"), facilities and infrastructure (e.g., "good information boards," "excellent paths," "facilities are second to none"), guides and services (e.g., "very good service," "excellent value"). Triggers for sentiments can also be activity-specific: in diving, the experience can be influenced negatively by the assigned partner or water clarity; in biking, it can be ruined by the necessity to share the trail with the hikers.

4.3.2 Emotions

In contrast to the SENSE OF SATISFACTION, emotions in this domain are of a more nuanced nature.

SENSE OF AESTHETIC APPRECIATION is reflected in the descriptions of something as being beautiful, and is encoded by a wide scope of adjectives with this core meaning. It is mostly mentioned in relation to natural and geographic features (e.g., "beautiful mountains"), routes taken (e.g., "a scenic ride," "a beautiful walk"), settings in general (e.g., "gorgeous views," "a stunning location," "scenic surroundings"), or in relation to natural phenomena (e.g., "a beautiful sunrise"). SENSE OF EXPERIENCING SOMETHING SPECIAL is encoded by adjectives that have a stronger positive denotation than those in the SENSE OF SATISFACTION aspect. Some of the examples include *exciting*, *fabulous*, *fantastic*, *special*, *unique*, *unforgettable*, *memorable*, as in "a unique wildlife experience." This aspect also includes lexemes that convey the surreal nature of the experience, such as *magic* and *surreal*, as well as metaphors that highlight the unearthly essence of the experience, as in "It's like going on a new planet." Reviewers also frequently report on the sense of being in awe and on finding something *impressive* and *epic*, usually in relation to views and natural and geographic features, but also in relation to experiences such as seeing glowworms in the complete darkness of a cave, as in "Keep your lights off and be in awe." SENSE OF HAVING AN ADVENTURE is mostly encoded by the lexemes *adrenaline*, *adventure*, and *adventurous*, as in "Perfect for everyone with a headlamp who loves a little adventure." SENSE OF HAVING FUN includes both the descriptions of having fun or finding something funny, as well as the descriptions of being bored or finding something boring. It is encoded by lexemes and phrases such as *fun*, *great fun*, *good fun*, or *entertaining*, as in "The weather didn't matter in the end, I had great fun in amusing company." SENSE OF RELAXATION is mostly described through lexemes such as *peaceful*, *relaxing*, *chill*, *relaxed*, as well as phrases such as *calm my nerves*, and rather idiosyncratic *stress buster*, as in "The ultimate stress buster for corporates or a very special occasion charter". SENSE OF SAFETY—or, fear—is another frequently mentioned aspect, usually represented by *scary* and *safe*, as in "Dove straight off the beach which was great—not as scary!" SENSE OF WILDERNESS is encoded by lexemes such as *nature*, *natural*, *wild*, or *wilderness*. Reviewers often report on the SENSE OF WILDERNESS in relation to the experience (e.g., "a very natural experience") or in relation to biophysical setting (e.g., "the wilds of the East Cape"), and describe the remoteness and pristine nature of certain locations (e.g., "a pristine lake"). Finally, OTHER EMOTIONS include less frequently mentioned emotions encoded by lexemes such as *shock*, *sorry*, *relief*, as in "When

you are there you get your surprise." Some examples of this aspect are descriptive, idiosyncratic, and pertain to the sense of place, as in "Awesome place to stay, like a little family" or "Sometimes I felt like true musters with the sheep going on up the hills ahead of us."

4.3.3 Cognition

Various types of knowledge and ways of learning and knowing manifest themselves at multiple stages and levels of recreation activities. Reviewers often refer to their expectations of and prior knowledge about destinations or activities, by verbs such as *know*, *expect*, or *warn*. These can be based on different sources of information, including prior experience with similar recreation activities in other places, reading, or being informed by others, as in "I had been warned that I would become wet." Not being informed about recreation opportunities such as non-commercial caves with glowworms can be one of the sources of regret in the reviews, as in "I wish I knew earlier about the caves." Another frequently mentioned type of knowledge is spatial knowledge and ability to locate oneself, which is crucial not only for successful navigation, but also for good timing to stay away from other groups, as in "I got an extra river map—very useful as you know exactly where you are (and timing to stay away from other companies)." As a result, the quality of spatial information representations such as maps and signs is often mentioned in the reviews, as in "Maps were vague and difficult to follow." In activities that take place in changeable environments—for example, in caving where the water level can change fast—the ability to perform a variety of judgments and make decisions about direction and safety gains importance, as in "Don't go alone and think well about going in." Commercial activities usually involve an introduction to the activity and related equipment, and reviews often include the assessment of the quality of this introduction, as in "good commentary," "good information before leaving," or "informative view of trout fishing." The educational aspect goes beyond learning about the activity in its narrow sense and encompasses learning about the wildlife and geology of the area, local history and traditions, or cultural practices such as farming or landscape engineering. This educational aspect is expressed in diverse ways: by describing the whole of the experience as interesting, by specifying what they have learned, by highlighting the knowledgeability of guides, or by assessing the quality of information boards with facts about the area. Recreationists especially appreciate when local people and guides share their knowledge, and they express disappointment when they feel like they lacked some tips. Finally, reviews also reflect how the cultural and knowledge background of the reviewer influences the educational aspect of the recreation experience, as in "Being of farming background it was very interesting to see how things are done in New Zealand."

5 Method and model evaluation

The reliability of the coding scheme was evaluated by the standard procedure of measuring the inter-annotator agreement. The data compiled for this experiment included two randomly chosen sentences for each of the 28 aspects, i.e. 56 sentences in total. Linguistic expressions to be annotated were marked in bold, and two independent annotators were given definitions of aspects and examples along the lines described in Section 4. The reliability measure Krippendorff's alpha was found to be 0.76, and the analysis of instances of disagreement suggested three main sources of the disagreement. First, one third of the

disagreement instances is found in the domain of emotions and results from varying subjective interpretations of the meaning of a lexeme: *phenomenal* in "phenomenal experience" was annotated as the SENSE OF EXPERIENCING SOMETHING SPECIAL by one annotator and as the SENSE OF SATISFACTION by the other, *lovely* in "a lovely place" was annotated as the SENSE OF AESTHETIC APPRECIATION by one annotator and as the SENSE OF SATISFACTION by the other. Second, the role of context becomes visible in cases such as "Got put on a more suitable horse," where *horse* was annotated as an ACTIVITY-RELATED ARTIFACT by one annotator, and as FLORA AND FAUNA by the other. Third, the experiment revealed that the annotation process can involve unnecessary inferences, as in case with *strong* in "Water was strong and deep." One annotator interpreted it as a SPATIAL PROPERTY, while the other annotator inferred that this description communicates the DIFFICULTY of the activity.

The process of the model validation, performed on 70 random reviews of long-distance treks, revealed a relatively high number of linguistic expressions that were not encountered in the dataset used for the manual content analysis; however, those could be unambiguously allocated to the aspects in the model, and mostly referred to WILDLIFE AND VEGETATION (e.g., *sandflies*, *tussock*), NATURAL AND GEOGRAPHIC FEATURES (e.g., *volcanic*, *coastline*), or FACILITIES and PERSONNEL (e.g., *huts*, *rangers*). Four linguistic expressions—*untouched*, *isolated*, *ancient*, *native*—added new semantics to the SENSE OF WILDERNESS aspect, and could also be singled out into separate aspects if their frequency were to suggests doing so for a given corpus.

Finally, to compare the approach chosen in this study to a more computational method, a topic modeling experiment was performed, with the ultimate goal of assessing efforts required for the discovery of topics corresponding to the aspects in the model. A detailed overview of the flow of this experiment is provided in Table 4. The experiment started by training LDA on a balanced corpus of reviews of all activities (176 for each activity, based on the size of the smallest sub-corpus), and the discovered topics corresponded to 12 aspects in the model. As one can see in Figure 5, many topics unsurprisingly reflect activities (biking for topic 27, caving for topic 24, kayaking for topic 18, horse trekking for topic 8, sailing for topic 7, swimming with mammals for topic 21). Other topics include guides (topics 17, 18, 25), views (topic 29), sense of having fun (topics 1, 15), weather (topic 22), general satisfaction (topics 1, 15), and time-related aspects of the experience (topics 5, 11, 16, 20). To rule out activity-related topics and discover more nuanced topics, another LDA run was performed on reviews of three activities—diving (two new aspects, TEAM and OTHER PEOPLE), fishing (two new aspects, FLORA AND FAUNA and KNOWLEDGE), and horse trekking (one new aspect, CULTURAL-HISTORICAL SETTINGS). [3] This run also produced some place-related topics, such as beach. Finally, to rule out both activity-related and place-related topics, three more runs were performed on the reviews of three most frequently reviewed trails in the "walking/trekking" sub-corpus, which unveiled one more aspect, the SENSE OF WILDERNESS. In total, experimenting with topic modeling allowed to uncover 19 out of 28 aspects in the model and to make the following observations. First, the process of topic interpretation was facilitated by the knowledge of texts. [4] As a result, keywords composing topics appeared semantically transparent, and the process of examining

---

[3] Given the main focus of the study, running LDA on reviews of each activity separately was beyond the scope of the experiment.

[4] For example, knowing that *free* is mostly used in the reviews the context of free access, and not in the context of a feeling of freedom, which might seem as a more plausible interpretation for someone not familiar with the corpus.

Figure 5: Topics produced by LDA run on the corpus of reviews of all activities.

topics was relatively straightforward and not as time-consuming as would be the case if the analyst was not familiar with the corpus. Second, achieving the level of semantic granularity comparable to the model resulting from the manual content analysis requires multiple iterations on sub-corpora, and the task becomes time-consuming relatively quickly. Third, the process of topic interpretation and discovery is not entirely objective, and, moreover, does not produce a model as an output—the latter still lies within the realm of the analyst.

6 Variation of aspects across activities

In total, 3,543 instances of aspects were annotated in the 200 reviews of the ten activities. Due to the size of the annotated corpus and the low frequencies of some of the aspects in certain activities, statistical testing could be only performed for the distribution of the main domains and sub-domains. Results of a set of chi-square tests support the hypothesis that the distribution of aspects varies across different types of activities; however, this does not hold true for all sub-domains and for all activities. Thus, a comparison of each activity against all others in Table 2 discloses that the settings domain is most responsive to the type of activity, followed by the activity domain, while no statistically significant differences are observed in the domain of emotions/cognition. Pairwise comparisons of sub-domains' distributions in Tables 5 and 6 provide further nuanced insights. For example, in the activity domain, walking/trekking is the only activity significantly different from most other activities. In the settings domain, most activities are significantly different from at least four other activities, and caving and horse trekking are significantly different from as many as eight activities.

Figure 6 allows to better understand the origin of the differences. Among the main domains, *settings* is the most prominent for most activities. One notable exception is fishing, where the *activity* domain predominates at 38.08%, which can be explained by the extractive-symbolic type of this activity [30]. The *emotions/cognition* domain fluctuates around 30%, being at its highest (38.08%) in rafting, which might be explained by the

	All domains	Activity	Settings	Emotions/cognition
caving	.0363	.2701	**< .005**	.0852
biking	.3059	**< .005**	.0067	.1630
diving	.6886	.4163	**< .005**	.0208
fishing	**< .005**	.0325	.0768	.4046
horse trekking	.0052	**< .005**	**< .005**	.2882
kayaking/canoeing	**< .005**	.9817	.0320	.4752
swimming with mammals	.0078	.8203	.0067	.4288
rafting	.0051	.2077	**< .005**	.0210
sailing	.1164	.3255	**< .005**	.3740
walking/trekking	.0080	**< .005**	**< .005**	.6021

Table 2: *P* values for the comparison of the distribution of domains and sub-domains in each activity against all others using the chi-square test, statistically significant values in bold ($P < 0.005$, after the Bonferroni correction).

fact that this is the most adventurous activity in the corpus, triggering a strong emotional arousal.

In the *activity* domain, ACTIVITIES AND ACTIONS are by far the most prominent aspect in all activities. Activity characteristics—DIFFICULTY and DURATION—are most prominent in the reviews of walking/trekking (41.82%), which involves active locomotion and long-distance routes (in comparison to caving). Reviews of walking/trekking abound with details on route segments and their difficulty. Activity-related ARTIFACTS are especially prominent in the reviews of horse trekking (38.30%), where many descriptions focus on horses, or in the reviews of other activities that involve means of assisted locomotion such as rafting and sailing. In caving—due to the nature of the environment—reviewers frequently mention torches and clothes for getting wet, which results in a relatively high proportion of this aspect (12.4%) in comparison to only 3.64% in walking/trekking.

In the *settings* domain, aspects related to the *biophysical settings* are most prominent in the caving (74.83%) and in the walking/trekking (62.40%) reviews. This can be explained, on the one hand, by the role of locomotion in these land-based activities, which makes spatial properties such as steepness relevant for the experience. On the other hand, both activities are appreciative-symbolic, which results in the detailed descriptions of natural and geographic features and landscapes. The smallest proportion of the *biophysical settings* is found in diving (23.23%), which might reflect the relative homogeneity of the environment in comparison to the settings of the land-based activities. In a number of activities, the *social settings* are more prominent than the biophysical environment. The most social activity in the corpus is rafting (46.9%), which can be explained by the importance of both teamwork and guides. *Social settings* are also salient in horse trekking, where the experience is largely shaped by the horse riding skills of other people on the team, and in diving, where a diving partner is required. Rather expectedly, the least social activities are caving (5.3%) and walking/trekking (4.8%), both of which can be undertaken on one's own and are mostly unguided. *Infrastructure* is most salient in the reviews of diving (40.65%), whereby reviewers comment on both the services and facilities. *Cultural-historical settings* are most frequently mentioned in biking reviews (10.4%). Riders cover larger distances and thus encounter more cultural elements of the landscape, such as wineries, villages, and small towns.
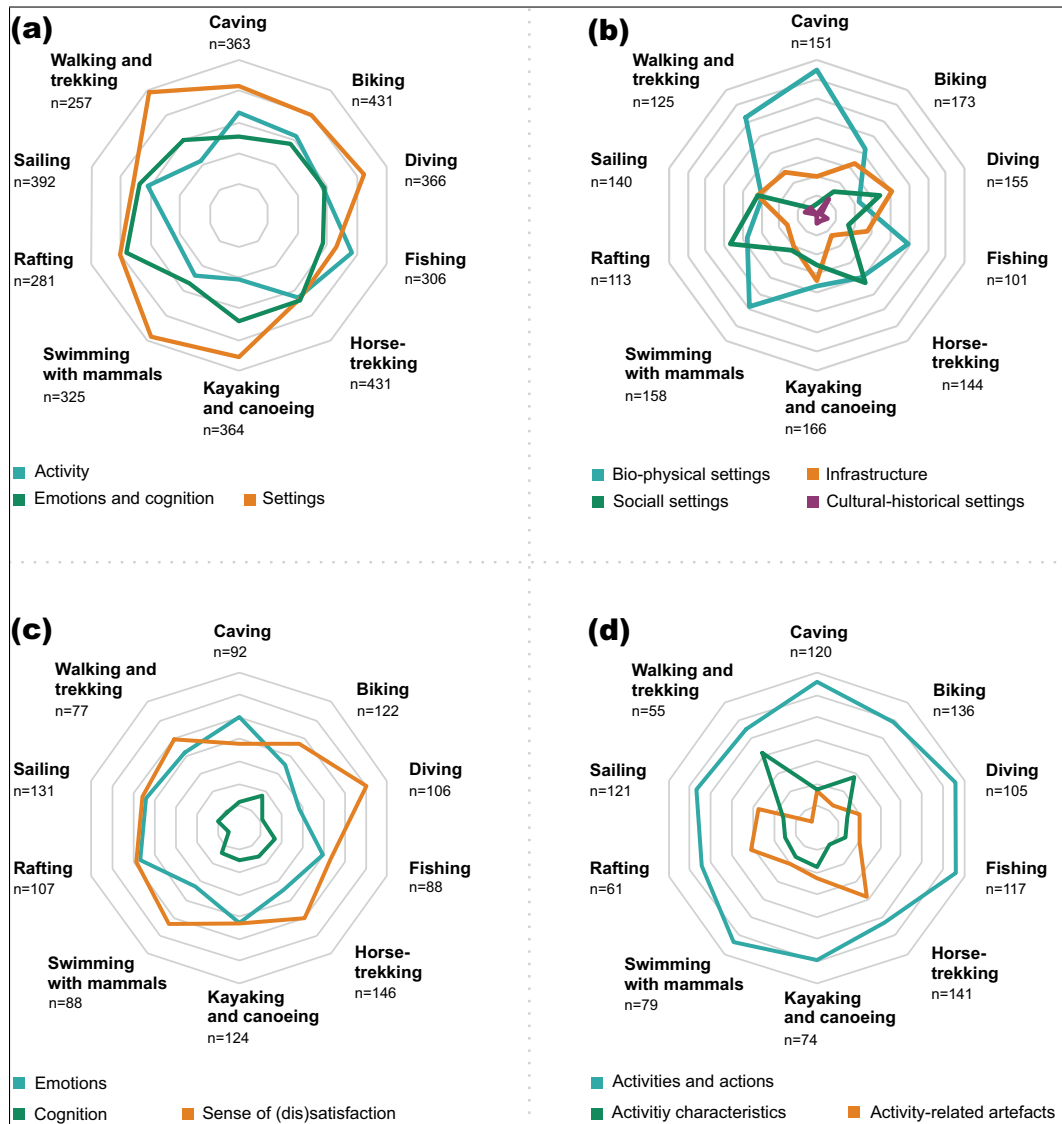
Figure 6: Distribution of main domains (a) and sub-domains(b-d) across the ten activities, in percent. Numbers indicate the total number of aspects in the (sub-)domain.
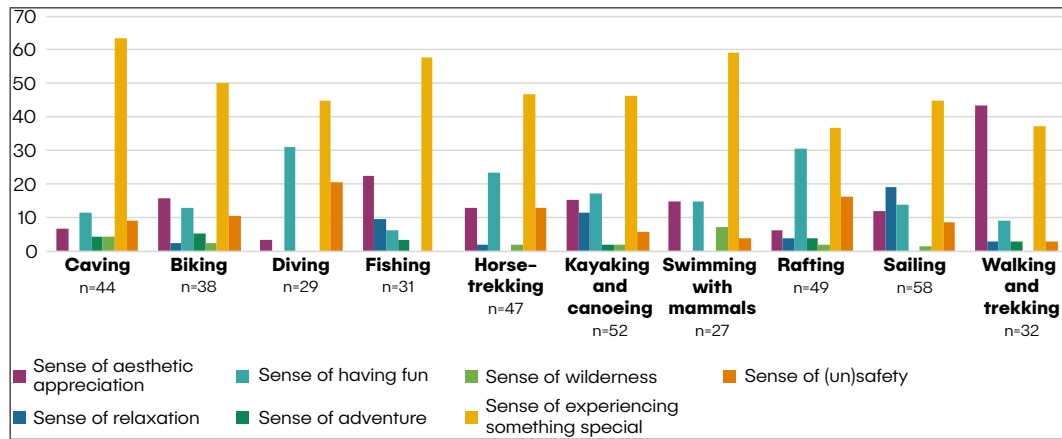
Figure 7: Distribution of aspects in the *Emotions* sub-domain across the ten activities, in percent. Numbers indicate the total number of aspect instances in the sub-domain.

In the *emotions/cognition* domain, the SENSE OF SATISFACTION prevails in the majority of the sub-corpora, essentially reflecting the nature of the data—that is, reviews. Cognitive aspects of recreation experiences are the least mentioned aspect, fluctuating between 4.67% and 18.03%. Some interesting patterns can be observed in the sub-domain of emotions, as shown in Figure 7 . First, while some activities—for example, caving—are dominated by one emotion, others—for example, diving and rafting—show a more balanced distribution of various emotions. Second, while reviews of activities such as caving, biking, and walking/trekking include mentions of (almost) the whole spectrum of emotions, reviews of activities such as diving and swimming with mammals are less rich and include references to four or five emotions only. Looking at the top emotions in each activity, the SENSE OF EXPERIENCING SOMETHING SPECIAL is the most frequent emotion in nine out of ten activities. Only walking/trekking is dominated by the SENSE OF AESTHETIC APPRECIATION (43.75%). The next most frequent emotion is the SENSE OF HAVING FUN, which is prominent in the reviews of diving, rafting, and horse trekking. As for the less frequent emotions, the SENSE OF SAFETY is found in all activities except fishing. It is the most frequent emotion in diving (20.69%) and rafting (16.33%), while the SENSE OF RELAXATION is found in almost activities but is most prominent in sailing (18.97%).

7   Discussion

7.1   Using textual VGI to model recreation activities

This study explored the content of reviews of nature-based recreation experiences in Aotearoa New Zealand by deriving a model of aspects of recreation activities. It involved a hybrid methodology that combined semi-automated frequency-based techniques with content analysis and was grounded in the analyst's knowledge of research areas related to nature-based recreation, cultural ecosystem services, and landscape perception. In the

era of machine learning, the chosen approach might be criticized as a resource-intensive task that is also prone to subjectivity. To address this potential critique by comparing the procedures and the results, the study included a topic modeling (LDA) experiment.

The LDA experiment clearly demonstrated that using topic modeling for the task undertaken in this study is neither more objective nor less time-consuming. First, it is important to understand that topic interpretation in this experiment was "blended" [7], since it was preceded by content analysis that contributed insights into the properties of discourse and helped disambiguate between word senses. A "naive" approach (in which topics are interpreted by an analyst who may be familiar with a topic area, but has no possibility for close reading) might have taken even more time, yielding less comprehensive results. Second, arriving at a fine level of semantic granularity of topics requires multiple LDA runs on different subsets of data, which in turn requires intuition and knowledge of data (e.g., ruling out place-based topics). Using topic modeling for a nuanced aspect detection thus appears to be as time-consuming as the manual screening of the most frequent terms and the content analysis, which formed the methodology of this study; also, while topic discovery appears objective and data-driven given its computational component, topic interpretation is still a subjective procedure. However, further methodological studies are needed to understand how these approaches compare and if and how they can be used in a complementary manner.

The dataset annotated in this study can be used to train supervised topic models, adding to the relatively new line of research that assesses the quality of supervised methods for leisure activity modeling [71]. Supervised topic modeling offers a solution to some of the pitfalls associated with both manual content analysis and unsupervised topic modeling but requires labeled datasets, coding schemes, resources and evaluation techniques. Indeed, initiatives such as SpatialML and ISO-space [42, 54] have facilitated the development of advanced methods for capturing spatial concepts related to location in text. Having similar resources for capturing aspects related to experiences would help extract and analyze corresponding concepts, enriching existing datasets with this "soft" information. In the era of crowdsourcing and citizen science, such efforts of data annotation and curation could be shared with the broader community of interested volunteers [6, 68].

7.2 Aspects of nature-based recreation activities

By developing a model of aspects of nature-based activities, this study captures and documents the richness with which recreation experiences are described in online reviews. In doing so, it provides initial insights into the nature and intensity of human-environment interaction in the context of recreation activities as reflected in this type of VGI. The model includes 28 aspects grouped into three domains: activity, settings, and emotions/cognition. These domains largely echo the components of models of human-environment interaction that have been developed previously from several disciplinary perspectives: the domain of the material world and physicality, the domain of human actions, the domain of meanings, and, more generally, cognition [63].

Further, the model developed in this study provides means for comparing aspects that can be derived from photographs and aspects that can be derived from text. Thus, some of the identified aspects largely correspond to those captured through image analysis: for example, the presence or absence of people, activities, landscape, animals, infrastructure, or history [14, 28]. Examples of aspects that have not been reported in the research that

leverages images include the difficulty and the duration of the activity, and aspects related to the domain of emotions/cognition. These findings provide first insights into the way in which different VGI modalities can be used in a complementary way, and they call for a more systematic analysis of the differences and overlaps in the sets of aspects that they reveal.

The set of aspects uncovered through textual VGI led to the following observations. In the activity domain, reviews report a variety of activities and actions in detail. As an example, reviews of caving describe locomotion, navigation, and exploration, as well as experiences like getting close to and seeing glowworms. Apart from the ten primary activities, reviewers frequently describe diverse secondary activities that they engaged into while in the area. This data source can thus be used in the line of work that extracts leisure activities potential from Web data [71]. Finally, an activity's difficulty and duration are another two prominent aspects in the reviews, suggesting that providing adequate information about these aspects is crucial for the quality of visitors' experiences.

Describing the settings where the activity takes place, reviews mention infrastructure and services and biophysical, social, and cultural-historical settings. Previous studies report that the biophysical settings and the perceived attractiveness of landscapes and natural features play a key role in the overall recreation experience [4, 14, 16, 59]. In line with these findings, descriptions of natural and geographic features abound in the reviews, ranging from mountains to stalactites. Reviews also document spatial properties such as steepness or comment on landscape in general. These varying levels of granularity in the spatial descriptions echo earlier findings that the activity influences the relevance and level of granularity of spatial information [18, 31]. Further, reviews reflect the importance of the social environment and of having an opportunity to engage into an activity with relatives to experience the "family togetherness" or to have a good time with other recreationists and local people [5, 8, 76]. Reviewers often describe the social interaction ranging from storytelling and cooking for each other to caring about other people on the trail. The role of guides and staff in adventurous outdoor activities is also clearly visible in the reviews, in line with previous studies [4, 59]. Apart from the positive aspects associated with the social settings, reviews also include negative emotions associated with overcrowding, known to have an impact on the experience from previous research [36, 58].

In the domain of emotions/cognition, reviews reflect the importance of learning something new, supporting previous findings [59, 74]. In line with earlier studies reporting on the role of social interaction in the learning activity in recreation, reviews often mention the knowledgeability of guides and local people [30]. Many of the aspects grouped in the emotions sub-domain have been previously reported in both recreation studies and landscape perception research. The sense of relaxation, safety, and the sense of having fun were identified as some of the major psychological benefits of recreation activities, resulting in place attachment and destination loyalty, while the sense of adventure and novelty is an important aspect in nature-based activities such as rafting [50, 76]. Interestingly, the sense of experiencing something special (the most frequent emotion in the reviews), encoded by adjectives such as *magic*, *surreal*, or *exceptional*, or metaphors such as *new planet* or *another world*, has not been discussed much in previous literature. At the same time, the corpus does not contain descriptions of the sense of tranquility and solitude, spirituality, nostalgia, identity, or belonging, found in previous studies on nature-based recreation [29, 34, 74]. Potential reasons could lie in the destination image of Aotearoa New Zealand, the fact that

reviews are mostly written by foreign tourists, and the fact that Aotearoa New Zealand is a remote destination for tourists to develop destination loyalty.

Finally, the comparison of the distribution of aspects across the ten recreation activities led to several observations that call for further investigation. The working hypothesis could be mostly confirmed for the settings domain—indeed, the proportions of descriptions of the infrastructure, biophysical, social, and cultural-historical settings differ significantly across activities, reflecting the unique attributes of activity types [30]. There is a prevalence of aspects related to the biophysical settings in reviews of appreciative-symbolic activities (e.g., caving, walking/trekking, or swimming with mammals), and the prevalence of aspects related to the social settings in rafting, an active-expressive activity that involves teamwork. In the activity domain, activity-related artifacts are most prominent in fishing, an extractive-symbolic activity, while activity characteristics such as difficulty and duration are most prominent in appreciative-symbolic activities that involve locomotion over large distances. Further, one can see differences in the emotions evoked by activities. While the size of the manually annotated corpus and the number of instances does not allow for statistical testing, comparing the distribution of emotions across the activities provides first insights into the way in which the type of a recreation activity influences the emotions that it evokes. Thus, the sense of relaxation has the highest proportion in sailing (a passive free-play activity), while the sense of aesthetic appreciation is most salient in walking/trekking (an appreciative-symbolic activity). In general, this comparative study suggests that online reviews are semantically rich, and invites further controlled studies of the variation in recreation experiences using larger datasets.

7.3   Limitations and future work

The proposed model is to a degree arbitrary and does not claim exhaustiveness or a finite level of granularity; it could also be expanded through additional layers of annotation—thus, the current coding scheme does not include the category of location since its focus is on general aspects of recreation experiences, and this layer could be added through the inclusion of toponyms into the scheme. As with any model, it is also an inevitable simplification of the modeled phenomenon. One example of such a simplification is the differentiation between the biophysical and the cultural-historical settings which obscures the fact that physicality is deeply rooted in the culture, and fails to recognize landscapes in a holistic sense [63]. Another example is the cognition aspect in the domain of emotions/cognition. Derived from the explicit knowledge-related descriptions in the reviews, this aspect simplifies the variety and complexity of cognitive processes that are involved in nature-based recreation activities but are not verbalized in the reviews—from risk-related problem solving such as wayfinding, to the perception of geographic features in landscapes [44]. Further, despite the meticulous iterative process of the model development, the resulting coding scheme and the annotated dataset can still contain annotation errors due to the human factor. The comparative study presented in this paper—while providing first insights into the variation of aspects across activities—has two major limitations. First, the size of the manually annotation corpus is relatively small for making final conclusions regarding specific aspects of recreation activities. Second, there is little control over the variation of factors *within* activities: thus, the perception of the same activity (e.g., walking/trekking) might differ if it is undertaken alone or through a commercial tour. Many

initial observations made in this study should be further tested through more controlled study design involving larger datasets.

More general limitations pertain to the "thin" type of data used in this study [40]. The absence of the background information about the reviewers—such as education, environmental attitudes, cultural background, or level of experience with the activity—prevents us from exploring how these factors may influence the perception of the recreation activity. The only background information available is the self-reported country of origin of a reviewer, and the analysis of distribution of those countries unveils a clear bias toward Europe. Indeed, the main language of the platform, English, serves as a gatekeeper and excludes other cultural and linguistic communities. It is important to keep this limitation in mind given that perception and emotional responses to the environment depend upon the understanding and awareness of the historical and cultural background of the landscape as well as the observer's own knowledge and cultural background [10]. Local populations, and Māori in particular, might have profoundly different relations with their ecologies in the context of nature-based recreation—for example, a more rooted sense of place [51, 62]. Thus, the results are primarily valid for the English-speaking international tourists from Europe and their experience of nature-based recreation activities in Aotearoa New Zealand, and transferability studies are needed to understand how well the model captures aspects of nature-based recreation activities as experienced by other groups in other localities. Further, reviews as a genre can be characterized by the confirmation bias that is found in post-trip survey questionnaires—namely, the tendency to describe aspects associated with positive (rather than negative) emotions as a result of the investment into the travel [1]. This can be particularly true in case of New Zealand, which represents an expensive and far-away destination for many.

Several research directions invite future work. As an immediate further step, one can use the dataset to develop methods for an automated annotation of aspects in text and leverage larger datasets in a series of case studies that will involve longitudinal and spatial analysis. Further, I aim to focus more closely on aspects related to emotions and involve methods from aspect-based sentiment analysis to link sentiments to aspects and gain a better understanding of recreation experiences. Finally, I also aim at exploring how data collected through traditional in situ surveys differ from image- and text-based VGI to set up methodological workflows that integrate and analyze multiple sources of multimodal data. I also aim to build upon the LDA experiment conducted in this study, and systematically compare insights produced by close and distant techniques, to understand how the latter can be used in a complementary manner.

7.4   Relevance and potential applications

VGI in the form of text and images has proved to be a rich source of data for studying collective perceptions of urban places [40]. However, leveraging VGI to study perceptions of natural environments is still in its infancy, and most research has only used the distribution of geolocated photographs as a proxy for landscape attractiveness and recreation visitation. This study highlights opportunities offered by the textual VGI data sources in the form of reviews. Its major contributions to spatial information science include the description of a VGI data source that has not yet been explored, the examination of semantically rich textual descriptions (as opposed to photograph tags used in previous research), and the in-

ventorization of the linguistic encoding of aspects of nature-based recreation experiences—including those related to emotion and cognition—through a model and a coding scheme.

The study has two immediate areas of application through the generated resources. First, the model provides basis for the extraction of geographic information in the nature-based recreation domain, and the resources made available through this study can be used to develop methods for the automated annotation of aspects in text. The annotated corpus can be used as a training dataset for machine learning; alternatively, it can be used to derive and expand lists of n-grams by using resources such as WordNet, providing a basis for a lexicon-based aspect detection [49]. Second, the model and the generated data can be used for the integration of multimodal geographic information in the nature-based recreation domain. An increasing number of studies of nature-based recreation explores the content of images by using tools such as Google Cloud Vision [22]. The latter generate tags for image descriptions, but the outputs are usually represented as frequency-based word clouds or clusters of tags, reminiscent of the topic modeling outputs discussed in Section 5, which leaves the semantic analysis rather coarse. The model and the annotated dataset developed in the current study provide a basis for exploring generated tags in a more semantically nuanced way and for integrating information derived from images and text. Importantly, they also offer means to compare and integrate information derived from multimodal VGI and data collected in more traditional ways—for example, in situ surveys.

The proposed model can thus facilitate the extraction and integration of geographic information from large multimodal datasets, and as such can be applied by a wide community interested in human-nature interaction for a variety of purposes. While aspects are grouped into broad domains, there is a possibility to disaggregate and subsequently recombine them into new domains and subdomains depending on the research question at hand. For example, park and recreation management agencies such as the Department of Conservation in Aotearoa New Zealand can use such data sources and methods to analyze and monitor walking/trekking reviews during the peak summer season, focusing on aspects related to facilities and infrastructure. Research into cultural ecosystem services and, in particular, the sense of place can use the outputs from this study as a starting point for a systematic expansion and further specification of aspects in the emotions sub-domain.

Taking a broader perspective, this study addresses the need to diversify data sources used in the study of human-environment interaction. Using the example of a VGI data source from Aotearoa New Zealand, it encourages researchers to explore more locally focused and activity-specialized sources, thus outbalancing data shadows cast by the few widely researched platforms such as Twitter and Flickr. Furthermore, text analysis should be effectively combined with image processing, since the exploration of multimodal VGI alongside more traditional methods can provide a more holistic data basis for studying complex and multifaceted phenomena such as human-environment interaction. Such research is important because of its potential to inform recreation management and related policy-making on multiple scales, with the final goal of minimizing the impact on the environment while maximizing the benefits for the humans.

8 Conclusion

This study set out to explore aspects of nature-based recreation activities through the prism of volunteered geographic information (VGI). Using reviews of ten land-based and water-

based activities, it developed a model that captures three major domains of aspects of recreation experiences (activity, settings, and emotions/cognition) and explored how these aspects differ across activities reflecting their unique attributes. This study contributes to several research directions. It advances VGI-based research that predominantly models places in urban contexts by highlighting the possibilities of modeling activities in natural environments. It contributes to the lines of research on nature-based recreation and cultural ecosystem services by suggesting a model that can be flexibly adjusted to answer a variety of research questions. It provides the spatial information science research community with a coding scheme and an annotated dataset that can be used as a basis for developing methods to annotate a rich set of aspects of recreation experiences in text automatically. The study thus makes an important contribution to the development of a methodological toolbox that is capable of leveraging geographic information of various modalities to provide a more holistic representation of human-environment interaction across multiple contexts.

Acknowledgments

References

[1] ALAEI, A. R., BECKEN, S., AND STANTIC, B. Sentiment analysis in tourism: Capitalizing on big data. *Journal of Travel Research 58*, 2 (2019), 175–191.

[2] ALAZZAWI, A. N., ABDELMOTY, A. I., AND JONES, C. B. What can I do there? Towards the automatic discovery of place-related services and activities. *International Journal of Geographical Information Science 26*, 2 (2012), 345–364.

[3] ANGRADI, T. R., LAUNSPACH, J. J., AND DEBBOUT, R. Determining preferences for ecosystem benefits in Great Lakes Areas of Concern from photographs posted to social media. *Journal of Great Lakes Research 44*, 2 (2018), 340–351.

[4] ARNOULD, E. J., AND PRICE, L. L. River magic: Extraordinary experience and the extended service encounter. *Journal of consumer Research 20*, 1 (1993), 24–45.

[5] BAUMGARTNER, R., AND HEBERLEIN, T. A. Process, goal, and social interaction differences in recreation: What makes an activity substitutable. *Leisure Sciences 4*, 4 (1981), 443–458.

[6] BONTCHEVA, K., DERCZYNSKI, L., AND ROBERTS, I. Crowdsourcing named entity recognition and entity linking corpora. In *Handbook of Linguistic Annotation*. Springer, 2017, pp. 875–892.

[7] BROOKES, G., AND MCENERY, T. The utility of topic modelling for discourse studies: A critical evaluation. *Discourse Studies 21*, 1 (2019), 3–21.

[8] BURCH JR, W. R. The social circles of leisure: Competing explanations. *Journal of Leisure Research 1*, 2 (1969), 125–147.

[9] CASALEGNO, S., INGER, R., DESILVEY, C., AND GASTON, K. J. Spatial covariance between aesthetic value & other ecosystem services. *PloS one 8*, 6 (2013), e68437.

[10] CHHETRI, P., ARROWSMITH, C., AND JACKSON, M. Determining hiking experiences in nature-based tourist destinations. *Tourism management 25*, 1 (2004), 31–43.

[11] CLEMENTE, P., CALVACHE, M., ANTUNES, P., SANTOS, R., CERDEIRA, J. O., AND MARTINS, M. J. Combining social media photographs and species distribution models to map cultural ecosystem services: The case of a Natural Park in Portugal. *Ecological indicators 96* (2019), 59–68.

[12] CUNHA, J., ELLIOTT, M., AND RAMOS, S. Linking modelling and empirical data to assess recreation services provided by coastal habitats: The case of NW Portugal. *Ocean & Coastal Management 162* (2018), 60–70.

[13] DAVIES, C. Reading geography between the lines: Extracting local place knowledge from text. In *International Conference on Spatial Information Theory* (2013), Springer, pp. 320–337.

[14] DELUCIO, J., AND MÚGICA, M. Landscape preferences and behaviour of visitors to Spanish national parks. *Landscape and Urban Planning 29*, 2-3 (1994), 145–160.

[15] DENIS, M. The description of routes: A cognitive approach to the production of spatial discourse. *Cahiers de Psychologie Cognitive 16*, 4 (1997), 409–458.

[16] DORWART, C. E., MOORE, R. L., AND LEUNG, Y.-F. Visitors' perceptions of a trail environment and effects on experiences: A model for nature-based recreation experiences. *Leisure Sciences 32*, 1 (2009), 33–54.

[17] EGOROVA, E., MONCLA, L., GAIO, M., CLARAMUNT, C., AND PURVES, R. S. Fictive motion extraction and classification. *International Journal of Geographical Information Science 32*, 11 (2018), 2247–2271.

[18] EGOROVA, E., TENBRINK, T., AND PURVES, R. S. Where snow is a landmark: Route direction elements in alpine contexts. In *International Conference on Spatial Information Theory* (2015), Springer, pp. 175–195.

[19] FAIRWEATHER, J. R., AND SWAFFIELD, S. R. Visitors' and locals' experiences of Rotorua, New Zealand: An interpretative study using photographs of landscapes and Q method. *International Journal of Tourism Research 4*, 4 (2002), 283–297.

[20] FARMINER, A. S. *People, Rivers and Recreation: Fluid Relationships of Place and Experience on the Clutha River, Otago, New Zealand*. PhD thesis, University of Otago, 2013.

[21] GHERMANDI, A. Integrating social media analysis and revealed preference methods to value the recreation services of ecologically engineered wetlands. *Ecosystem services 31* (2018), 351–357.

[22] GHERMANDI, A., SINCLAIR, M., FICHTMAN, E., AND GISH, M. Novel insights on intensity and typology of direct human-nature interactions in protected areas through passive crowdsourcing. *Global Environmental Change 65* (2020), 102189.

[23] GLIOZZO, G., PETTORELLI, N., AND HAKLAY, M. Using crowdsourced imagery to detect cultural ecosystem services: A case study in South Wales, UK. *Ecology and Society 21*, 3 (2016).

[24] GOODCHILD, M. F. Citizens as sensors: The world of volunteered geography. *GeoJournal 69*, 4 (2007), 211–221.

[25] GUERRERO, P., MØLLER, M. S., OLAFSSON, A. S., AND SNIZEK, B. Revealing cultural ecosystem services through Instagram images: The potential of social media volunteered geographic information for urban green infrastructure planning and governance. *Urban Planning 1*, 2 (2016), 1–17.

[26] HAMSTEAD, Z. A., FISHER, D., ILIEVA, R. T., WOOD, S. A., MCPHEARSON, T., AND KREMER, P. Geolocated social media as a rapid indicator of park visitation and equitable park access. *Computers, Environment and Urban Systems 72* (2018), 38–50.

[27] HAYES, A. F., AND KRIPPENDORFF, K. Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures 1*, 1 (2007), 77–89.

[28] HEIKINHEIMO, V., MININ, E. D., TENKANEN, H., HAUSMANN, A., ERKKONEN, J., AND TOIVONEN, T. User-generated geographic information for visitor monitoring in a national park: A comparison of social media data and visitor survey. *ISPRS International Journal of Geo-Information 6*, 3 (2017), 85.

[29] HEINTZMAN, P. Nature-based recreation and spirituality: A complex relationship. *Leisure Sciences 32*, 1 (2009), 72–89.

[30] HENDEE, J. C., GALE, R. P., AND CATTON, W. R. A typology of outdoor recreation activity preferences. *The Journal of Environmental Education 3*, 1 (1971), 28–34.

[31] HIRTLE, S. C., TIMPF, S., AND TENBRINK, T. The effect of activity on relevance and granularity for navigation. In *International Conference on Spatial Information Theory* (2011), Springer, pp. 73–89.

[32] HU, Y., DENG, C., AND ZHOU, Z. A semantic and sentiment analysis on online neighborhood reviews for understanding the perceptions of people toward their living environments. *Annals of the American Association of Geographers 109*, 4 (2019), 1052–1073.

[33] KEELER, B. L., WOOD, S. A., POLASKY, S., KLING, C., FILSTRUP, C. T., AND DOWNING, J. A. Recreational demand for clean water: Evidence from geotagged photographs by visitors to lakes. *Frontiers in Ecology and the Environment 13*, 2 (2015), 76–81.

[34] KIL, N., HOLLAND, S. M., STEIN, T. V., AND KO, Y. J. Place attachment as a mediator of the relationship between nature-based recreation benefits and future visit intentions. *Journal of Sustainable Tourism 20*, 4 (2012), 603–626.

[35] KLING, F., AND POZDNOUKHOV, A. When a city tells a story: Urban topic analysis. In *Proceedings of the 20th international conference on advances in geographic information systems* (2012), pp. 482–485.

[36] KOHLHARDT, R., HONEY-ROSÉS, J., FERNANDEZ LOZADA, S., HAIDER, W., AND STEVENS, M. Is this trail too crowded? A choice experiment to evaluate tradeoffs and preferences of park visitors in Garibaldi Park, British Columbia. *Journal of Environmental Planning and Management 61*, 1 (2018), 1–24.

[37] KUHN, W. Ontologies in support of activities in geographical space. *International Journal of Geographical Information Science 15*, 7 (2001), 613–631.

[38] LANGEMEYER, J., CALCAGNI, F., AND BARÓ, F. Mapping the intangible: Using geolocated social media data to examine landscape aesthetics. *Land use policy 77* (2018), 542–552.

[39] LEVIN, N., LECHNER, A. M., AND BROWN, G. An evaluation of crowdsourced information for assessing the visitation and perceived importance of protected areas. *Applied geography 79* (2017), 115–126.

[40] LIU, Y., YUAN, Y., AND ZHANG, F. Mining urban perceptions from social media data. *Journal of Spatial Information Science 2020*, 20 (2020), 51–55.

[41] MANCINI, F., COGHILL, G. M., AND LUSSEAU, D. Using social media to quantify spatial and temporal dynamics of nature-based recreational activities. *PloS one 13*, 7 (2018), e0200565.

[42] MANI, I., DORAN, C., HARRIS, D., HITZEMAN, J., QUIMBY, R., RICHER, J., WELLNER, B., MARDIS, S., AND CLANCY, S. SpatialML: Annotation scheme, resources, and evaluation. *Language Resources and Evaluation 44*, 3 (2010), 263–280.

[43] MANNING, C. D., SURDEANU, M., BAUER, J., FINKEL, J. R., BETHARD, S., AND MC-CLOSKY, D. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations* (2014), pp. 55–60.

[44] MARK, D. M., TURK, A. G., BURENHULT, N., AND STEA, D. *Landscape in language: Transdisciplinary perspectives*, vol. 4. Amsterdam, John Benjamins Publishing Company, 2011.

[45] MARQUES, C., REIS, E., AND MENEZES, J. Profiling the segments of visitors to Portuguese protected areas. *Journal of Sustainable Tourism 18*, 8 (2010), 971–996.

[46] MARQUES, C., REIS, E., MENEZES, J., AND SALGUEIRO, M. D. F. Modelling preferences for nature-based recreation activities. *Leisure Studies 36*, 1 (2017), 89–107.

[47] MCCALLUM, A. K. MALLET: A machine learning for language toolkit. *http://mallet. cs. umass. edu* (2002).

[48] MENG, F., AND UYSAL, M. Effects of gender differences on perceptions of destination attributes, motivations, and travel values: An examination of a nature-based resort destination. *Journal of sustainable tourism 16*, 4 (2008), 445–466.

[49] MILLER, G. A. Wordnet: A lexical database for English. *Communications of the ACM 38*, 11 (1995), 39–41.

[50] OH, H. C., UYSAL, M., AND WEAVER, P. A. Product bundles and market segments based on travel motivations: A canonical correlation approach. *International Journal of Hospitality Management 14*, 2 (1995), 123–137.

[51] PANELLI, R., AND TIPA, G. Placing well-being: A Māori case study of cultural and environmental specificity. *EcoHealth 4*, 4 (2007), 445–460.

[52] PURVES, R., EDWARDES, A., AND WOOD, J. Describing place through user generated content. *First Monday 16*, 9 (2011).

[53] PUSTEJOVSKY, J. ISO-Space: Annotating static and dynamic spatial information. In *Handbook of Linguistic Annotation*. Springer, 2017, pp. 989–1024.

[54] PUSTEJOVSKY, J., MOSZKOWICZ, J., AND VERHAGEN, M. A linguistically grounded annotation language for spatial information. *TAL. 53*, 2 (2012), 87–113.

[55] RAHIMI, M. M., NAGHIZADE, E., STEVENSON, M., AND WINTER, S. Service quality monitoring in confined spaces through mining Twitter data. *Journal of Spatial Information Science 2020*, 21 (2020), 229–261.

[56] RAUBAL, M. Spatial data science for sustainable mobility. *Journal of Spatial Information Science 20* (2020), 109–114.

[57] RICHARDS, D. R., AND FRIESS, D. A. A rapid indicator of cultural ecosystem service usage at a fine spatial scale: Content analysis of social media photographs. *Ecological Indicators 53* (2015), 187–195.

[58] ROSSI, S. D., BYRNE, J. A., PICKERING, C. M., AND RESER, J. "Seeing red" in national parks: How visitors' values affect perceptions and park experiences. *Geoforum 66* (2015), 41–52.

[59] RUHANEN, L. The prominence of eco in ecotourism experiences: An analysis of post-purchase online reviews. *Journal of Hospitality and Tourism Management 39* (2019), 110–116.

[60] SEN, S. Use of affordances in geospatial ontologies. In *Towards Affordance-Based Robot Control*. Springer, 2008, pp. 122–139.

[61] SERNA, A., GERRIKAGOITIA, J. K., BERNABÉ, U., AND RUIZ, T. Sustainability analysis on urban mobility based on social media content. *Transportation Research Procedia 24* (2017), 1–8.

[62] STEPHENSON, J. Many perceptions, one landscape. *Landscape Review 11*, 2 (2007), 9–30.

[63] STEPHENSON, J. The cultural values model: An integrated approach to values in landscapes. *Landscape and Urban Planning 84*, 2 (2008), 127–139.

[64] TAYLOR, K., AND AMIDY, M. Data-driven agriculture for rural smallholdings. *Journal of Spatial Information Science*, 20 (2020), 125–135.

[65] TENBRINK, T. *Cognitive discourse analysis: An introduction*. Cambridge University Press, 2020.

[66] TENERELLI, P., DEMŠAR, U., AND LUQUE, S. Crowdsourcing indicators for cultural ecosystem services: A geographically weighted approach for mountain landscapes. *Ecological Indicators 64* (2016), 237–248.

[67] TOURSIM NEW ZEALAND. Special interest sectors. https://www.tourismnewzealand. com/media/3163/special-interest-infographic.pdf, 2018. Last Accessed October 2, 2021.

[68] TSUENG, G., NANIS, S. M., FOUQUIER, J., GOOD, B. M., AND SU, A. I. Citizen science for mining the biomedical literature. *Citizen science: theory and practice 1*, 2 (2016).

[69] VAN BERKEL, D. B., TABRIZIAN, P., DORNING, M. A., SMART, L., NEWCOMB, D., MEHAFFEY, M., NEALE, A., AND MEENTEMEYER, R. K. Quantifying the visual-sensory landscape qualities that contribute to cultural ecosystem services using social media and LiDAR. *Ecosystem services 31* (2018), 326–335.

[70] VAN DAMME, M.-D., OLTEANU-RAIMOND, A.-M., AND MÉNEROUX, Y. Potential of crowdsourced data for integrating landmarks and routes for rescue in mountain areas. *International Journal of Cartography 5*, 2-3 (2019), 195–213.

[71] VAN WEERDENBURG, D., SCHEIDER, S., ADAMS, B., SPIERINGS, B., AND VAN DER ZEE, E. Where to go and what to do: Extracting leisure activity potentials from web data on urban space. *Computers, Environment and Urban Systems 73* (2019), 143–156.

[72] VAN ZANTEN, B. T., VAN BERKEL, D. B., MEENTEMEYER, R. K., SMITH, J. W., TIESKENS, K. F., AND VERBURG, P. H. Continental-scale quantification of landscape values using social media data. *Proceedings of the National Academy of Sciences 113*, 46 (2016), 12974–12979.

[73] VARANKA, D. E., AND CARO, H. K. Spatial relation predicates in topographic feature semantics. In *Cognitive and Linguistic Aspects of Geographic Space*. Springer, 2013, pp. 175–193.

[74] WEBER, D., AND ANDERSON, D. Contact with nature: Recreation experience preferences in Australian parks. *Annals of Leisure Research 13*, 1-2 (2010), 46–69.

[75] WINTER, P. L., SELIN, S., CERVENY, L., AND BRICKER, K. Outdoor recreation, nature-based tourism, and sustainability. *Sustainability 12*, 1 (2020), 81.

[76] YOON, Y., AND UYSAL, M. An examination of the effects of motivation and satisfaction on destination loyalty: A structural model. *Tourism management 26*, 1 (2005), 45–56.

[77] ZAHRA, K., IMRAN, M., AND OSTERMANN, F. O. Automatic identification of eyewitness messages on Twitter during disasters. *Information processing & management 57*, 1 (2020), 102107.

Domains and aspects	Examples of their linguistic encoding	Examples of the context
Activity		
Activities and actions	*caving, sailing, biking, bird watching, sheep shearing; galloping, paddling, scrambled, climb; activity, trip, boat trip, return trip; saw, looked for, got close; arrived, discovered, explored; booked, hired, rented, paid*	"I *saw* dolphins and *caught* some nice fish," "*Galloping* along the beach was wonderful," "*Paid* $140, 2 hours *motoring*, *drove* around *stopping* at the same beaches," "I *stopped* enroute at the Lake Pukaki Salmon farm and *picnicked* whilst *watching* people fish;"
Activity-related artifacts and their qualities	*clothes, torch, trekking stick, boat, yacht, swim bar, bikes, kayaks, photos, pictures, gear, newish, fish, ex racer, well-trained, well-equipped, suitable, comfortable*	"Need to have a good *headlamp*," "Great *photos* and a 7 metre waterfall drop," "Jane got us sorted with the *bikes*," "I hired *snorkelling equipment* for a day," "Carl knew exactly where to go to catch *Yellowtail Kingfish*," "*Stable kayaks* but heavy to paddle;"
Duration	*hours, minutes, full day, days, h, min, short, long*	"It takes approximately 45 *minutes*," "*Short* walk from the car park," "The training in the beginning was *not too long*," "Very good 2.5 *h* trip around the Punakaiki river and on the beach;"
Difficulty	*strenuous, doable, easy, demanding, intense, for beginners, for any age, no brainer, gnarly*	"Some few steep sections must be walked but 98% was *ridable*," "A bit *strenuous* but *good exercise* trekking up to Rawhiti cave," "Happy that there were *easier* and *tough* routes to choose from;"
Settings		
Biophysical settings		
Spatial properties and relations	*huge, long, steep, flat, hilly, slippery, muddy, dry, water level, close, distant, fresh air, noise, quiet*	"Very *small* and *wet* cave." "Water was *strong* and *deep*," "The river was not at a good *water level*," "*Closest* largest town is Whangarei," "Lloyd explained that the sea bed of Opotiki was very *flat*," "Seeing all the *colours of* the surrounding environment, the crater...;"
Natural and geographic features	*mountain, beach, cave, rainforest, river, stalactite, rock, coastline, limestone escarpments, sea bed, hilltop, bay, islands*	"Really nice walk through the *forest* with an awesome end at the *beach*," "And then a paddle up a *stream* to this amazing *waterfall*," "I rode right up into the *hills* above Cardrona, almost up to the *snow line*;"
Settings in general	*area, landscape, location, place, region, site, spot, views, surroundings, scenery, undersea, underwater landscape, setting*	"Amazing *location*," "Excellent *spot* and walkway," "One of the most spectacular *landscapes* you could imagine," "Nice *underwater landscape*, lots of fish, different species," "Beautiful *surroundings*," "As big LOTR fans, I got to see breathtaking *scenery* from the movies;"
Flora and fauna	*seals, glowworms, penguins, marine life, underwater wildlife, a Weka family, dolphins, vegetation, colony, sea birds*	"Plus I saw a heap of *marine life* and amazing coastal views," "A lot of *fish*, *seals*, *black coral*," "Great to see *seal pups* at play this time of year," "Once I saw *dolphins* cruising around us I forgot the cold water and enjoyed the *Hector dolphins*;"
Time of the day and processes	*dawn, evening, tide*	"My friend and I went for *sunset* and the lighting on the mountains was beautiful," "Boat trip through the French Pass at full ebb *tide* was just amazing;"
Seasons, weather and conditions	*March, summer, weather, wind, rain, conditions, hot, sunny*	"I visited it at the end of *winter*," "A nice thing to do on a *sunny* day," "Perfect *conditions*;"
Infrastructure		
Free access	*free, for free, unguided, by yourself, independent*	"Great, easy and *free*," "You can go *by your own* through a 600m long cave," "Fantastic *free* long distance trekking track," "The best trip *for free*;"
Facilities and their qualities	*parking lot, camping spot, dive spot, cruising ground, information boards, showers, WiFi, viewing platform, well maintained, well marked, well formed*	"Great *facility* with *showers* and *WiFi*," "Excellent, *well-maintained* paths," "Beautiful but too long and no breaks—no *water fountains/caves/loos*— nothing!," "The track is *well-marked* with *orange triangles*;"
Services and their qualities	*service, transport, accommodation, taxi service, canoe hire, company, personalised, organised, cheapest*	"Very good *personalised service*," "Really good small *company* with great guides," "Good *choice of tours*, friendly *service*, *transport* for Queen Charlotte Sound," "Three boats full of people, *organised with military precision*;"

Domains and aspects	Examples of the linguistic encoding	Examples of the context
Social settings		
Team	*family, buddy, wife, for the family, boys, daughter, my other half*	"My *husband* and I had a great time," "A fantastic weekend trip *for the family*," "*Our son* took us through this with his *wife*," "Great with *kids*;"
Personnel	*crew, dive master, instructor, guide, host, staff, team*	"Very professional *guides*," "Enthusiastic fun *staff*," "*Operator* was very friendly and informative," "Excellent *crew*;"
Other people	*overcrowded, crowded, alone, people, hikers, guests, other passengers, other cyclists, other companies*	"Not too many *people* on the track," "Good trip, but *crowded* boat," "At peak trekking season it could become disappointing to pass the *hikers*," "Be careful at each corner to avoid collision with *walkers* or *other cyclists*;"
Social interaction	*sense of humour, friendly, patient, laugh, talk, help, share, respect, storytelling, welcoming, treated*	"*Unfriendly* staff," "Good equipment, *easy going* atmosphere," "Thank you so much for *sharing* this beautiful place!," "Butler is a *real character* and very entertaining."
Cultural-historical settings	*Māori, sculpture, carvings, history, village life, historical, traditional ceremony, coal mining, farming, Lord of the Rings, wineries, vineyards, lighthouse, town, old road*	"The *vineyards* I went to were absolutely fantastic," "Easy day walk towards—if not to—*Pencarrow Lighthouse*," "Opotiki is a relaxed little *town* with lots of beaches close by," "I expected to see *Lord of the Rings* locations," "Very knowledgable about sea birds and *local folklore*;"
Emotions and cognition		
Sense of aesthetic appreciation	*beautiful, lovely, breathtaking, spectacular, scenic, gorgeous*	"*Gorgeous* bays and views," "*Beautiful* beech forest next to river," "*Scenic* views;"
Sense of wilderness	*hidden, wild, nature*	"But I awoke to a beautiful sunrise looking north towards the *wilds* of the East Cape;"
Sense of relaxation	*relaxing, relaxed, chill, chilling out, peaceful, leisurely*	"I had a *relaxing* time," "Good guide, *peaceful* scenery—Abel Tasman," "It was brilliant being able to have a *leisurely* canoe down the beautiful Whanganui;"
Sense of having fun	*fun, funny, good fun, great fun, epic fun, amusing*	"*Great fun*," "*Epic fun* and awesome guides," "The weather didn't matter in the end, I had *great fun* in *amusing* company!;"
Sense of adventure	*adrenaline, adventure, adventurous*	"Beautiful and *adventurous*," "The experience is fantastic, a great mix of scenery, storytelling and *adrenaline*," "A little *adventure*, walked along a dry river bed then up a track to find this huge cave;"
Sense of experiencing something special	*exciting, fabulous, fantastic, magic, memorable, special, stunning, surreal, unbelievable*	"A very *unique* experience," "Very *special* cave if you are up for a steep walk," "The trip through French Pass is definitely an *unforgettable* thing to do;"
Sense of safety	*treacherous, comfortable, safe, safety*	"Could prove *treacherous* in heavy rain," "Once there, awesome beach and setting, *safe* swimming, great for families;"
Other emotions	*shock, surprise, regret, nervous, relief, chaotic*	"Amazing and *chaotic* and that just made it better," "He loved every minute of it and never once looked like he was *nervous*;"
Sense of satisfaction	*enjoy, like, good, excellent, best, wonderful, disappointed*	"I really *enjoyed* my ride," "A *wonderful* experience," "*Great* to be able to visit the caves without paying for a tour," "*Perfect* tracks for beginners," "The diving a bit *disappointing*;"
Cognition	*knowledge, facts, tips, information, commentary, mine of knowledge, learn, interesting, informative, well explained*	"Steve was a great guide and *knew* loads of fascinating *facts* about the glow worms and the surrounding area," "*Informative* tour but the prawn fishing was the best bit," "John and Ginny were fantastic hosts and ...*introduced* us to village life;"

Table 3: Aspects, examples of their linguistic encoding and context.

Aspects	All reviews	Diving	Fishing	Horse trek.	Trail 1	Trail 2	Trail 3	In all runs
activities and actions	++++++++	++++++	+++++++	+++++		+++	+	+
activity-related artifacts	+++	+++	+++	++				+
difficulty								
duration	++++			+++	++	+	+	+
spatial properties and rel.								
natural and geogr. features	+	+			+	++	+++	+
settings in general	++	+	+	++	+++	++		+
flora and fauna			+					+
time of the day and processes							+	+
seasons, weather and cond.	+				+			+
free access								
facilities		++			+	++		+
services	++	++			+	+		+
team		+						+
personnel	+++	++	+	+++				+
other people		+		+	+			+
social interaction								
cultural-hist. sett.				+				+
sense of aesthetic appreciation	+	+			+		+	+
sense of wilderness					+			+
sense of relaxation								
sense of having fun	++		+					+
sense of adventure								
sense of exper. smth special	+	+					+	+
sense of safety								
sense of satisfaction	++++	+	++	+++++		+	++++	+
other emotions								
cognition			+					+
New aspects in the run	12	3	2	1	1	0	1	**Total: 20**

Table 4: Aspects captured through topics produced by LDA runs on different datasets. The number of "+" indicates the number of topics that fell into this aspect category. "+" indicates an aspect that has not been encountered in previous runs.

	caving	biking	diving	fishing	horse trek.	kayak/ can.	swim. w/m.	rafting	sailing	walk./ trek.
caving		.0969	.7036	.5984	**< .001**	.5358	.8120	.0813	.1381	**< .001**
biking	.0969		.0179	.0087	**< .001**	.0597	.0740	.0031	.0025	.0706
diving	.7036	.0179		.9890	.0075	.6878	.9131	.2414	.3652	**< .001**
fishing	.5984	.0087	.9890		.0076	.6476	.8632	.2465	.3588	**< .001**
horse trek.	**< .001**	**< .001**	.0075	.0076		.0355	.0145	.4024	.0876	**< .001**
kayak./can.	.0597	.0597	.6878	.6476	.0355		.8824	.5574	.7844	**< .001**
swim.w/m.	.8120	.0740	.9131	.8632	.0145	.8824		.3348	.5237	.**< .001**
rafting	.0813	.0031	.2414	.2465	.4024	.5574	.3348		.8612	**< .001**
sailing	.1381	.0025	.3652	.3588	.0876	.7844	.5237	.8612		**< .001**
walk./trek.	**< .001**	.0706	**< .001**	**< .001**	**< .001**	**< .001**	.**< .001**	**< .001**	**< .001**	

Table 5: P values for the pairwise comparison of the distribution of aspects in the *Activity* domain using the chi-square test, statistically significant values in bold (P < 0.001, after the Bonferroni correction).

	caving	biking	diving	fishing	horse tr.	kayak/ can.	swim. w/m.	rafting	sailing	walk./ trek.
caving		**< .001**	**< .001**	**< .001**	**< .001**	**< .001**	**< .001**	**< .001**	**< .001**	.1184
biking	**< .001**		**< .001**	.4899	**< .001**	.0729	.0033	**< .001**	**< .001**	.0060
diving	**< .001**	**< .001**		**< .001**	**< .001**	.0260	**< .001**	**< .001**	.2524	**< .001**
fishing	**< .001**	.4899	**< .001**		**< .001**	.0565	.1897	**< .001**	.0038	.0083
horse tr.	**< .001**	**< .001**	**< .001**	**< .001**		**< .001**	**< .001**	.7712	**< .001**	**< .001**
kayak./can.	**< .001**	.0729	.0260	.0565	**< .001**		**< .001**	**< .001**	.4278	**< .001**
swim.w/m.	**< .001**	.0033	**< .001**	.1897	**< .001**	**< .001**		**< .001**	**< .001**	**< .001**
rafting	**< .001**	**< .001**	**< .001**	**< .001**	.7712	**< .001**	**< .001**		.0038	**< .001**
sailing	**< .001**	**< .001**	.2524	.0038	**< .001**	.4278	**< .001**	.0038		**< .001**
walk./trek.	.1184	.0060	**< .001**	.0083	**< .001**	**< .001**	**< .001**	**< .001**	**< .001**	

Table 6: P values for the pairwise comparison of the distribution of aspects in the *Settings* domain using the chi-square test, statistically significant values in bold (P < 0.001, after the Bonferroni correction).