*January 2022*

# An Exploratory Survey of Recreational Activities Using Twitter Data with Logic-Based Location Categorization

Ferenc BÉRES [a], Gábor LUKÁCS [a], András J. MOLNÁR [a] and Pál SZEILER [b]
[a] *SZTAKI – Institute for Computer Science and Control, Budapest, Hungary*
[b] *ELTE – Eötvös Loránd University, Budapest, Hungary*

**Abstract.** Social media analysis has become a major instrument for data-driven tourism. It allows surveying visitor behavior on multiple scales. Considering the geographical characteristics of users' posts from social media platforms, we were able to address more specific questions related to the place type selection patterns of the visitors. In this paper, we present OpenLostCat, our first-order-logic-based location categorizer applicable for modeling location types depending on OpenStreetMap data. We report our findings revealed by this tool on more than one year's collection of global-scale Twitter data, focused on potential trail-related hiking and trekking activities. We categorized visited locations in our experiments based on place accessibility – transport and trail infrastructure –, and analyzed these categories according to the travel distance taken by visitors in general to reach these areas. Our comparisons reveal seasonal characteristics, continental differences (between Europe and North America), as well as specifics related to selected recreational areas. Besides these preliminary findings available for further verification, we show both the perspectives and limitations of our approach for future improvements and experiments.

**Keywords.** outdoor recreation, visitor monitoring, social media, volunteer-based mapping, location types, logic-based modeling

## 1. Introduction and Related Work

### *1.1. Outdoor Recreation, Trails and Visitor Management*

Outdoor tourism and recreation, especially visiting natural sites and protected areas while exercising physical activities such as walking or biking, has been getting increased popularity in recent years. While it is desirable for a number of reasons, including mental and physical health, as well as education and experience about our environment, the impact of human presence, especially in fragile habitats, may cause deterioration if not managed well. The advantage of visitor monitoring and management is manifested not only in mitigating such impacts but in general, gaining better knowledge about visitation patterns related to outdoor activities ensures improved decision-making about potential infrastructural developments, policymaking, maintenance efforts, or business and marketing analyses and endeavors in local or regional settings, or even on a global scale.

A crucial infrastructure of such outdoor activities is the network of marked or designated touristic and recreational *trails*, whose visitation patterns are mostly studied on a local or regional level, such as specific parks and sites for specific management and marketing purposes. Studying trail-related activities and visitation patterns on a global scale with the most probable geographical accuracy seems to be an interesting and not yet well-studied topic. Such an approach would not only provide common references for regional or park-based surveys but would also extend the scope of such studies into areas not directly or differently managed, or fall outside of the usual scopes of research and monitoring projects for any reason. For example, hiking and pilgrimage trail routes, especially in Europe, form an interconnected and broad network across all different countries, types of landscape or terrain, administrative and land management units with a wide variety of stakeholders.

Traditionally, on-site surveys are primarily utilized for visitor monitoring (such as in e.g. [25], [8]), in order to gain knowledge about the actual number, behavior, preferences and satisfaction of visitors in an area or at specific sites. With the development of technology, this has been extended with different methods such as trail counters and secondary data sources [23]. More large-scale aspects are revealed by secondary surveys and studies like [9]. A relatively early work on social media and tourism in general is [19], which is more into business and marketing aspects. Tracking movement trajectories of users by mobile devices has emerged as a possible approach, but it needs appropriate tools, willingness and explicit consent by the visitors.

### *1.2. Social Media Analysis in Visitor Monitoring*

In recent years, social media analysis has become a hot topic even for visitor management and monitoring. A recent paper [28] shows that the majority of related research is focused on the area of the USA. It mentions some studies on Twitter, and reveals that geolocation-based analysis is dominant over terms and tags looked for in textual contents, but very few consider associating volunteer-generated public map data (such as OpenStreetMap). The study identifies the following types of analyses: spatial analysis, temporal, cultural ecosystem services, economic values and sentiment analysis. Amongst the studied platforms, Flickr, Panoramio, Instagram, OpenStreetMap and Twitter are shown to have been the most popular. None of the mentioned works consider combining micro-level location characterization based on OpenStreetMap data with geotagged social media posts.

Another remarkable work is [32], a systematic review on social media and visitor use management in parks and protected areas, which also discusses limitations. In their study, Twitter follows directly after Facebook as the second most widely used platform, and Asia is coming right after Europe and North America in continental comparison. The popularity of different social media platforms has been changing over the years, and in many cases, and in many cases, data needs manual processing so that proper automated tools and methods provide high value. They also report that only a few studies have utilized fine-grained social media resolution in space and time, and it should be investigated more on how spatial and temporal visitation patterns can be estimated based on geolocated user posts. At the same time, they give warnings about limitations, inaccuracies and biases of such approaches, mainly regarding social media posts not being representative for the population of all the visitors or the visited sites. They give best practices, some of which can be utilized to verify or improve our study results in the future as well.

Multi-source analyses and correlation studies with other data sources include [29], comparing different social media platforms and official statistics in South Africa and Finland and identifying potential sources of mismatch such as geography (cell signal coverage), sudden events, park profiles, visitor profiles. [22] works with thematically categorized photos and compares them with user profiles revealed in online surveys, without fine-tuned location categorization or geotags, but based solely on photo content. An Instagram-based study with photo classification for user interests and activities, as well as home location detection and temporal patterns is [16], comparing social media findings with visitor survey. A fine-tuned but somewhat similar research approach is followed by [11] using geotagged photos along particular trails, comparing the results with official forest service surveys for validation and generalizability. The latter uses Flickr images, trip reports and on-site monitoring tools, and suggests social media analysis as a complementary, 'gap-filling' means in addition to regular on-site surveys and monitoring actions. Flickr is also used for quantifying nature-based tourism and recreation (visitation rates, origins, changes over time) by [33], concluding that Europe and the USA are the most popular areas, followed by Japan and New Zealand, almost a decade ago. Geotagged photos are used by [13] to identify hotspots in a smaller area using statistical methods and an overlay of particular on-site facilities.

Further works include [27] about national parks in Germany (using Flickr and a grid-based geographical splitting method and residence resolution of users), [12] about the Jeju Island in Korea analysed using Twitter data overlayed with other data sources such as land cover and OpenStreetMap, but also used grid cells for dealing with locations, and [14] using photo data with fine-grained land-cover to tackle hotspots for cultural and heritage tourism in a coastal region of Mexico.

Amongst other approaches, it is worth mentioning that [30] uses Foursquare venue check-ins by twitter collection related to travel diaries and activity preferences, [15] applies machine learning and natural language processing for classifying users and visitation areas based also on photos, [26] uses Twitter for studying seasonality of activities in urban parks, and [24] applies sentiment analysis for georeferenced visitor posts created in a single theme park.

### *1.3. Location Modeling using Volunteer-Generated Geographical Data*

One of the main questions related to analyzing social media or other user-generated data is to determine and categorize the actual location of the posting user. As we saw above, some studies analyze photos taken by visitors, and the type of location may be determined based on image content analysis and classification without even knowing the exact photo location in terms of geographic coordinates. Other approaches use explicit place type information, such as Foursquare check-ins. In our paper, we explore a different approach: considering only geotagged content with (supposedly) exact locations, determine the type of place based on the geographical features found at the location.

Our approach is similar to the *Recreational Opportunity Spectrum (ROS)* [10], which categorizes locations based on the environment and the proximity of specific types of assets and facilities. ROS has a normative status in the USA and has been adapted in slightly different variations to some other areas as well. Its rule-based nature makes it straightforward to apply to different areas if the appropriate data is available. For instance, [18] applies a rule-based approach for defining ROS categories in New Zealand, applies a rule-based approach for defining ROS categories in New Zealand, but their classification is more complex than ours. Instead, we aim for a simple and possibly uniform definition of global-scale location categories to explore rough visitation patterns. Another, more recent publication [21] uses a complex approach based on ecosystem services and ROS, including accessibility categorization, and among other data sources, uses the volunteer-generated public mapping database OpenStreetMap.

OpenStreetMap [3] has become a widely used data source and platform for a multitude of purposes, including tourism and humanitarian applications. It is free to use and based on volunteer contributions. Its power not only lies in its flexibility and the many tools and applications already built atop, but also in the way it organizes geographical data and enforces contributors to organically extend the existing content with their additions and modifications instead of uploading independent content in parallel about the same area or place. This way, a single global unified mapping database is being built and improved step-by-step, although there can be slight differences in the exact way of representing similar items across regions or countries.

Details on the data model of OpenStreetMap can be found in its community-based documentation [4], which contains a specific part for representing walking- and hiking-related facilities [5]. OpenStreetMap uses a free-form *tagging* approach instead of predefined table schemata for storing data assigned to geometric shapes: a geographic element (node or way) may have an arbitrary number of *tags* associated with it, with arbitrary names and values, and the rules are defined as community conventions [4]. Tags describe the type as well as any properties or measures of the geographic object. An additional data element type is the *relation*, which acts as a container for nodes and ways. A relation can hold its own tags and defines a higher-level object composed of its parts, such as a (usually longer) route, containing its parts and pieces as ways already existing in the database as street segments. OpenStreetMap uses the relation approach to represent marked and designated trail routes for hiking, biking, and other activities. As the database contains semantic and not presentation information such as styling, it has a multitude of map visualizations, and potentially any rule-based map symbology can be defined, depending on the purpose. As an example, a relevant interface for our topic is the WaymarkedTrails website, having customized maps of different types of routes [17]. Figures 1 & 2 show the national and international trails as an example, mapped on OpenStreetMap and presented by WaymarkedTrails.

In our exploratory study of this paper, we will define a simple yet flexible and powerful way of location type modeling by filter-based categorization, formulated as a subset of first-order logic. Since we do not need joins or complex, correlated subqueries for location type definitions, but only to determine the existence or nonexistence of geographic objects with specific tag values or value combinations in the proximity of each queried location, we do not need the full expressive power of relational calculi. A similar approach based on filtering and key-value equality testing is common practice in map visualization as well, to determine which geographical objects should be displayed on a particular map visualization and how (styling). We consider sets of all geographic objects in proximity of a point instead of single features. It is similar to the ROS approach but simpler and more universal.

Our rule language for location categorization will be a subset of JSON [1], i.e. all well-formed formulae in our language will also be a well-formed JSON expression. The main advantage of JSON is the direct human readability while at the same time it is ma-

**Figure 1.** Marked national and international hiking trails in North America mapped on OpenStreetMap [17]. Regional and local trails are not included here.

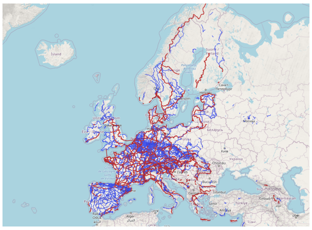


**Figure 2.** Marked national and international hiking trails in Europe mapped on OpenStreetMap [17]. Regional and local trails are not included here.

chine processible. A similar language has been developed as JsonLogic [31], but we do not need that complexity. One of the keys to keeping our expressions simple and intuitive is the natural correspondence between JSON constructs and logical operations, as we will see later in the paper. Another key is the implicit quantification we utilize through wrapping subformulae with default logical quantifiers based on their operands and context, similarly to how it was proposed in the system by [34], but with an opportunity for explicit variable quantification wherever convenient. Implicit quantification is a common practice in more complex languages and settings as well, such as, for example, in [7].

### *1.4. Goal and Outline of this Paper*

Our current study aims to explore whether any relevant potential visitation patterns can be discovered for outdoor recreation activities (especially for hiking, trekking, or other means related to trails) using geotagged Twitter data combined with OpenStreetMap location data on a global scale. For that purpose, we had selected a set of keywords characteristic to such activities and locations and collected matching Twitter data for more than a year. We applied our simple logic-based location categorization method on nearby geographic objects to infer the type of each geotagged location, mainly according to its accessibility by transportation and/or by marked trails. If they seem to be relevant, any preliminary findings should be verified and investigated in more detail in future studies, as we assume the data cannot be representative but can give hints for patterns and phenomena to look for more specifically. Our exploratory study aims to show the advantages, perspectives, and limitations of our approach and discuss possible future improvements, experiments, and applications.

The outline of the rest of our paper is as follows: Section 2 presents our tool used for location categorization with characteristics and examples of its rule language. Section 3 describes our Twitter data collection and the related data preprocessing steps, including user residence assignment. Here we also present geographical aspects of the specific locations in the highlight of our analysis (Section 3.3). Our results are exhibited and discussed in Section 4 in detail, while Section 5 concludes with some general remarks on the advantages and limitations of our approach, as well as possible future directions and applications.

## 2. Location Categorization with OPENLOSTCAT

### 2.1. *Introducing* OPENLOSTCAT*, the Logic-Based Location Categorizer*

OPENLOSTCAT (*Open Lo*gic-based *S*imple *T*ag-bundle *Cat*egorizer) is a free open-source utility designed and implemented by the authors in Python for data analysts, engineers and scientists who want to determine the characteristics of geolocated points in their datasets [20]. OPENLOSTCAT does the job by assigning category labels to each point based on logical rules defined in JSON for tags of nearby OpenStreetMap objects.

The tool can query OpenStreetMap objects at exact locations given by WGS 84 coordinates via the Overpass API (with a customizable proximity distance) [6]. User-defined location category labels are then generated for the given locations, based on logical formulae defined for available categories and evaluated for the set of queried OpenStreetMap objects in the proximity of each location. The set of database tags of each OpenStreetMap object (attribute instances, i.e. the actual data associated with the object is similar to a data tuple with no fixed schema) is called a *tag bundle*. Therefore, a logical rule is evaluated over a *tag bundle set*, giving a true or a false result for each defined location category.

A simple and comprehensive JSON format is used to describe a category rule collection, called a *category catalog*. Each category is defined by a rule expression, in which references to other, previously defined (sub)expressions can be reused. Reusable subexpressions are called *references* inside a location category catalog. For the assignment of location categories, the category catalog may prescribe its evaluation strategy as *single-category* (applying the first matching rule) or *multi-category* (applying all matching rules).

### 2.2. *Location Category Definition Language Characteristics and Examples*

The formal language of *OpenLostCat* consists of location category definition rules is a univariate first-order logic. Allowing only a single, implicit variable in predicates keeps the language and its evaluation overly simple. It can also be viewed as a simplified tuple calculus (without cross-product or join operations) for tuples without a fixed schema because OpenStreetMap allows arbitrary (finite number of) data tags assigned to a geographic object in its tag bundle. The order of tags is irrelevant as they are identified by their names. The language is implemented in JSON format, which allows utilizing a natural correspondence between arrays and disjunction, as well as object (record) notation and conjunction. For some cases, special keywords in the form of JSON field name prefixes are introduced, as it will be shown in the example formulae.

During the evaluation process, a condition formulated as a predicate formula is evaluated for each queried geographical object (actually represented as a tag bundle) in the proximity of a specific location being categorized, thus producing a true/false value for every single geographical object. Eventually, these values are aggregated into a single true/false value for a given location to be decided whether the location belongs to a particular category. Therefore, each rule defining a category must be quantified, either explicitly or implicitly. Variable quantification is added implicitly according to the actual operations of the formula, when not specified explicitly. The language can easily be extended in the future with more complex operations as needed.[1]

Because only a single variable is used in each (sub)formula, but already quantified (sub)formulas can also be connected via logical operators, this results in our language having two explicit levels of (sub)expressions:

- *Item-level (a.k.a. geographic-object-level or filter-level)* subexpressions are non-quantified expressions composed of atomic formulae and logical connectives actually correspond to set operators as being evaluated one-by-one on elements of the input set of queried map objects in the proximity of a given location, thus resulting a in subset of their input set containing the matching geographical objects (actually, the tag bundles of them) for further processing.
- *Category-level (a.k.a. bool-level)* (sub/)expressions are (explicitly or implicitly) quantified expressions, resulting in a single boolean value. Such an expression can be a directly quantified item-level subexpression with its result set aggregated into a single aggregated boolean value, or composed of such formulae using logical connectives whose operators work on the category level with boolean inputs and produce boolean results.

*References* as named subexpressions can also be defined and can be used and referenced from multiple category definitions, like building blocks, referring to distinct (sub)concepts. This way, repeated parts of rules do not have to be explicitly duplicated, and whenever a change is necessary, it can be done in one place. The language distinguishes item-level and category-level references, prefixed with a single and a double hashmark, respectively. A single-hashmark reference refers to a concept defined over a single geographic object under evaluation for filtering, while a double-hashmark reference refers to a location under question having already a set of geographic objects being evaluated in its proximity.

Some representative examples are shown in Table 1 with their JSON syntax and corresponding first-order formulae.

The smallest building block of this language – besides the boolean constants *true* and *false* – is the *atomic filter*, which checks whether a key is present in a tag bundle and the value of the tag equals the desired value, or values listed in an array (actually a compact disjunction). A *NULL* value is used to indicate non-existence of the named tag in the tag bundle. Atomic filters can be combined into a *conjunctive* formula by adding more than one tag filter to a single *JSON object* in curly brackets – or using the special name prefix __AND_ as needed. Alternatively, a *disjunctive* formula can be created by the *JSON array* notation (values or objects listed in square brackets) – or by the prefix __OR_, whichever is appropriate. The rest of first-order operations are expressed by the prefixes __NOT_ (negation), __IMPL_ (for implication), __ANY_ (existential quantifier) and __ALL_ (universal quantifier), and the prefix __REF_ is used to indicate a reference wherever needed, but the hashmark notation already identifies references where no additional JSON attribute name is necessary as in a key-value notation setting. The reason for using prefixes is that each attribute name must be unique in a JSON object and adding an

[1] Implicit quantification is achieved by default quantifier wrapping: if a (sub)formula occurs in a context where no free variable should be present anymore, the system 'wraps around' the subformula with a $\exists$ or $\forall$ quantifier, depending on the actual content of the (sub)formula. For simple cases, a positive statement gets a $\exists$ while a negative (negated) statement becomes quantified with $\forall$. More details follow.

| Reference Name | OPENLOSTCAT JSON Rule | First-Order Logical Formula |
|---|---|---|
| **a) Defining geographic *item-level* location concepts** as references in form of predicate formulae with a free variable: | | |
| *#residential_area* | {”landuse” : ”residential” } | $x[landuse] = "residential"$ |
| *#marked_trail* | {”route” : [”bicycle”, ”canoe”, ”foot”,<br>”hiking”, ”horse”, ”inline_skates”,<br>”mtb”, ”piste”, ”running”, ”ski”]} | $x[route] \in \{"bicycle","canoe","foot",$<br>$"hiking","horse","inline\_skates",$<br>$"mtb","piste","running","ski"\}$ |
| *#public_motor_accessible* | {”__NOT__” : [<br>{”access” :[false, ”private”]},<br>{”motor_vehicle” : false} ]<br>} | $\neg \bigvee \big( x[access] \in \{"no","private"\},\ x[motor\_vehicle] = "no" \big)$ |
| *#common_public_road* | {<br>”highway” :[”motorway”, ”trunk”,<br>”primary”, ”secondary”, ”tertiary”,<br>”unclassified”, ”residential”,<br>”motorway_link”, ”trunk_link”,<br>”primary_link”, ”secondary_link”,<br>”tertiary_link”, ”living_street”,<br>”service”, ”pedestrian”, ”bus_guideway”,<br>”escape”, ”raceway”],<br>”surface” :[null, ”paved”, ”asphalt”,<br>”concrete”, ”concrete:lanes”,<br>”concrete:plates”, ”paving_stones”,<br>”sett”, ”unhewn_cobblestone”,<br>”cobblestone”, ”metal”, ”wood”],<br>”__REF_not” : ”#public_motor_access”<br>} | $\bigwedge \big($<br>$x[highway] \in \{"motorway","trunk","primary","secondary","tertiary","unclassified","residential","motorway\_link","trunk\_link","primary\_link","secondary\_link","tertiary\_link","living\_street","service","pedestrian","bus\_guideway","escape","raceway"\},$<br>$x[surface] \in \{NULL,"paved","asphalt","concrete","concrete:lanes","concrete:plates","paving\_stones","sett","unhewn\_cobblestone","cobblestone","metal","wood"\},$<br>*#public_motor_accessible* $\big)$ |
| *#transport_accessiblility* | [<br>{”amenity” : [”ferry_terminal”, ”parking”]},<br>{”public_transport” : ”platform”},<br>{”aerialway” : ”station”},<br>{”railway” : ”station”},<br>{<br>”highway” : [”services”, ”trailhead”,<br>”rest_area”,”emergency_bay”,<br>”elevator”, ”bus_stop”],<br>”__REF__” : ”#public_motor_accessible”<br>},<br>{ ”__REF__” : ”#common_public_road” }<br>] | $\bigvee \big($<br>$x[amenity] \in \{"ferry\_terminal","parking"\},$<br>$x[public\_transport] = "platform",$<br>$x[aerialway] \in "station",$<br>$x[railway] \in "station",$<br>$\bigwedge \big( x[highway] \in \{"services","trailhead","rest\_area","emergency\_bay","elevator","bus\_stop"\},$ *#public_motor_accessible* $\big),$<br>*#common_public_road* $\big)$ |
| **b) Defining *category-level* location concepts** as references in form of quantified (closed) logical formulae: | | |
| *##Easy_access* | [ ”#residential_area”, ”#transport_accessibility ] | $\exists x \bigvee \big($ *#residential_area*, *#transport_accessibility* $\big)$ |
| *##Trail_close* | [ ”#marked_trail” ] | $\exists x$ *#marked_trail* |
| *##Trail_close_2* | { ”__ANY__”: ”#marked_trail” } | $\exists x$ *#marked_trail* |
| *##Calm_streets_only* | { ”__NOT__” : {<br>”__ANY__” :<br>{”highway”: [”primary”, ”secondary”]} }<br>} | $\neg \exists x\ x[highway] \in \{"primary","secondary"\}$ |
| *##No_public_transport _access* | {” __ALL__”: {<br>”__NOT_1”: { ”public_transport”:<br>[”stop_position”, ”platform”] },<br>”__NOT_2”: { ”amenity”: ”ferry_terminal”},<br>”__NOT_3”: { ”aerialway” : ”station” },<br>”__NOT_4”: { ”railway” : ”station” },<br>”__NOT_5”: { ”highway” : ”bus_stop” } }<br>} | $\forall x \big($<br>$\neg\, x[public\_transport] \in \{"stop\_position","platform"\},$<br>$\neg\, x[amenity] = "ferry\_terminal",$<br>$\neg\, x[aerialway] = "station",$<br>$\neg\, x[railway] = "station",$<br>$\neg\, x[highway] = "bus\_stop" \big)$ |
| *##Fully_wheelchair_ accessible_station* | { ”__IMPL__”: [<br>{”public_transport”: [”stop_position”, ”platform”] },<br>{”wheelchair”: [true, ”designated”]} ],<br>”__ANY__”:<br>{”public_transport”: [”stop_position”, ”platform”] }<br>} | $\bigwedge \big($<br>$\forall x\ x[public\_transport] \in \{"stop\_position","platform"\} \rightarrow x[wheelchair] \in \{"yes","designated"\},$<br>$\exists x\ x[public\_transport] \in \{"stop\_position","platform"\} \big)$ |

**Table 1.** Example logical rules defined as (sub)expression references used by OPENLOSTCAT for location categorization. The JSON syntax is explained by corresponding first-order logic formulae. Note the implicit logical variable $x$, the connectives generated from arrays and objects, and cases of implicit quantification.

arbitrary suffix to any of the operator prefixes allow adding further subexpressions with the same operator on the same level into a conjunctive formula expressed by a JSON object.

Implicit quantification for a formula is defined generally as *existential*, except when *each* of the disjunctive subformulas in its *conjunctive normal form* has *at least one negated* atomic operand (or an operand being the *false* boolean constant). In that case, the implicit quantifier applied will be *universal*. This principle is exactly that of [34] and has the advantage of the assigned implicit quantifiers being invariant of the actual way of expressing equivalent logical (sub)formulae. It naturally corresponds to the common-sense interpretation of a negation of a single atomic filter (or their conjunction) as being universal, and for the implication, which is implicitly wrapped into a universal quantifier, reflects the assumed intention by the implication being declared as a universal rule. Logical connectives inherit the default quantifier type from their subexpression(s) and resolve them according to the rule of each connective.[2] If a quantified subformula is directly connected to a free-variable subformula via a logical connective, then implicit quantification will be enforced on the latter, or for the whole formula if a free-variable formula is used for defining a category or a category-level reference.

A detailed description of language rules and definitions for all the syntax and semantics are given in the documentation files of [20].

### *2.3. Location Categorization for Our Study*

Location categories used for the exploratory analysis of this paper reflect the assumed accessibility for each location. Places located in residential areas, or in the proximity of conventional roads, public or motorized transport services are defined as *frontcountry* locations. Generally, visitors do not need much hiking or other physical activity to reach these areas. On the contrary, backcountry points are further away from such infrastructure in our terms and generally assumed to be reached by active physical movement only, at least based on the locally found OpenStreetMap objects (items) around them. Note that these concepts are used here somewhat differently from the standard ROS categories or their usual interpretation in recreation geography or ecology, but these terms are appropriate for our current purpose. Furthermore, each location may or may not have marked or designated *trails* in their proximity, thus forming four different categories together with the two above.

More precisely, the four location categories are based on some of the concepts defined as subexpressions in Table 1. We set the OpenStreetMap query radius as $100m$, so the location of each geotagged tweet is categorized by evaluating its logical formula on the geographical objects found in its proximity of 100 meters. The definitions are given below as a *category catalog*, with the logical subformula references defined in Table 1. Its evaluation strategy assigns the first matching category to each location being categorized:

[2] Negation switches the default quantifier type of its operand between $\exists$ and $\forall$, conjunction defaults to existential $\exists$ if any of its operands default to $\exists$ (otherwise $\forall$), and disjunction inherits the universal default quantifier $\forall$ if any of its operands have it as default (otherwise defaults to $\exists$). Therefore, note that e.g. in Table 1 at ##*No_public_transport_access* the quantifier *__ALL_* could have been omitted as it would produce the same, universally quantified formula due to being a conjunction composed only of negated atomic filters.

**Frontcountry_with_trail:** $\#\#Easy\_access \wedge \#\#Trail\_close$ (being at most 100m distant from a marked trail and a residential area or a place accessible by conventional motorized or public transportation), or else:

**Frontcountry:** $\#\#Easy\_access$ (being at most 100m distant from a residential area or a place accessible by conventional motorized or public transportation), or else:

**Backcountry_with_trail:** $\#\#Trail\_close$ (being at most 100m distant from a marked trail but further away from any residential area or transportation access point), or else:

**Backcountry:** $\top$ (everything else: being more than 100m distant from a marked trail and from any residential area or transportation access point).

Note that we do not utilize the full power of OPENLOSTCAT's logical language for our current location categorization setting.

## 3. Recreational Data Collection from Twitter

### *3.1. Twitter Data Collection*

In this work, we collected geotagged tweets from August 2020 to November 2021 to uncover recreational user activity patterns for different countries and geographic locations. Table 2 presents the rich keyword set that we specified for the public Twitter search API during the data collection. It contains general phrases related to nature, wilderness parks, and various recreational activities (mainly hiking, trekking or pilgrimage), but popular keywords and hashtags referring to distinct relevant and popular geographic locations and landmarks are also included.

Initially, we were also experimenting with the geolocation-based Twitter search API, which returns tweets within a pre-defined radius of a given reference point. Unfortunately, the volume of trekking-related data in this setup was susceptible to the choice of the radius parameter. For example, close to populated areas, most of the tweets returned by geolocation search was not related to recreational activities. Due to this behavior, we decided to use the keyword-based search approach without any *a priori* geolocation restrictions.

During data collection, we found that Twitter users have a very low tendency to geotag their tweets. Thus, we excluded the majority of collected tweets from our experiments, as we consider geotagged tweets only. Next, we excluded spam Twitter accounts related to bots, weather, earthquake, fire, and traffic reports based on their activity patterns in the collected data. In total, we managed to collect 437 098 geotagged tweets from 116 916 different users.

### *3.2. User Residence Locations*

In this work, we rigorously assess the travel distance Twitter users are willing to take to reach their desired recreational areas (e.g., marked trail, wilderness park, nearby lake, or park). For this reason, we needed to assign a residence location to users in our data set. Twitter users may opt to publish their residence (home location) on their profile in the form of textual information. We first extracted the underlying city or district names from these text snippets along with the corresponding latitude and longitude coordinates

| Group | English | Keywords |
|---|---|---|
| General | Yes | nature, walking, lake, view, views, landscape, trail, hike, hiking, trekking, climbing, mountains, mountain, mount, mountainlife, mountainlovers, rockclimbing, trailrunning, backpacking, naturebeauty, naturelovers, naturephotography, hikingadventures |
| | No | wandern, randonnee, randonnées, excursionismo, escursione, escursioni, montaña, montagna, montagne, senderismo, naturaleza, peja |
| Natural parks | Yes | stateparks, nationalpark, national/state/regional + park/forest/grassland/seashore/reserve, protection/wilderness/recreation/conservation area, mountain park |
| | No | parque/parco/parc/réserve + national/nacional/natural/regional/régional/faunique, parco nazionale, reserva da biosfera, naturpark, naturschutzgebiet |
| Pilgrimage | Mixed | pilgrimage, peregrinacion, peregrino, pilgrim, pelerin, camino, caminho, caminopeople, elcaminopeople, caminodesantiago, caminofrances, buencamino |
| Location hashtags | Mixed | prisojnik, julijskealpe, adirondacks, dolomiti, dolomites, austrianalps, pyrenees, zermatt, matternhorn, montblanc, snowdonia, mountmonadnock, mountwhitney, whitemountains, runyoncanyon, lagodicarezza, valldenuria, valdifassa, valledaosta, valfiorentina, altoadige, hautesavoie, peakdistrict, lakedistrict, vignemale, laketahoe, tahoe, bardonecchia, appalachia, trentino, südtirol, southtyrol, yosemite, mttamalpais, appalachiantrail, glaciernationalpark, valdisole, karersee, banffnationalpark, himalayas, everest, cantwellcliffs, chiefloganstatepark, hudsonvalley, hockinghillsstatepark |

**Table 2.** Set of key phrases used during the Twitter data collection.

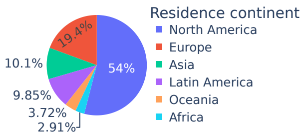



**Figure 3.** Distribution of user residence by continents.

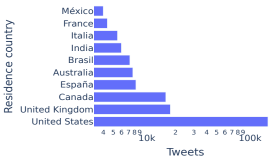



**Figure 4.** Number of tweets posted by users who live in the top 10 most popular countries by user residence.

using the *wikipedia* Python package. Then, by querying the *Nominatim* geocoding API [2], each user was additionally assigned a country based on its approximate residence coordinates. This way, we assigned a valid residence location to 83 779 users. In Figure 3, we show their distribution by continents, while Figure 4 presents the tweet volume posted by users with residence in the most popular countries occuring in the data set. Unfortunately, many users publish invalid or inconclusive locations in their Twitter profiles (e.g., 'Around the world', 'Mars', 'London&NYC'). We excluded tweets posted by these users from further experiments.

### *3.3. Selected Countries and Areas*

Our data contains both local and global recreational user patterns as we did not restrict data collection to specific countries. However, as Figures 3 & 4 showed, Twitter users are more active in the Western Hemisphere (e.g., USA, Canada, Western Europe). Fur-

| Area | Tweets | Users | Circle centers (lat,long) | Radius (km) |
|---|---|---|---|---|
| Western Alpes | 1543 | 589 | (44.141754,6.9726151) | 50 |
| | | | (45.0437458,6.3892906) | 50 |
| | | | (45.9666772,7.6373506) | 50 |
| | | | (45.8323849,6.8637096) | 50 |
| Dolomites | 865 | 323 | (46.4118581,11.8216688) | 50 |
| Pyrénées | 856 | 357 | (42.6997233,-0.4364245) | 50 |
| | | | (42.6551014,0.6589325) | 60 |
| | | | (42.3923069,1.9831357) | 50 |
| Snowdonia | 552 | 307 | (53.070095,-3.969984) | 20 |
| Peak District | 276 | 134 | (53.419918, -1.771878) | 10 |
| Lake District | 1058 | 443 | (54.464247, -3.035512) | 25 |
| Scotland | 697 | 377 | (57.4834038,-5.0739052) | 150 |
| Rockies | 2471 | 1209 | (39.6933868,-105.8914474) | 50 |
| | | | (39.1723719,-106.8358464) | 50 |
| | | | (40.5709711,-105.8804096) | 50 |
| White Mountains | 1051 | 410 | (44.0908472,-71.5055852) | 50 |
| Smokies | 1122 | 529 | (35.139817, -83.750694) | 60 |

**Table 3.** The number of tweets and users related to the selected recreational areas in our data set. We also present the covering circles with radius for each location.

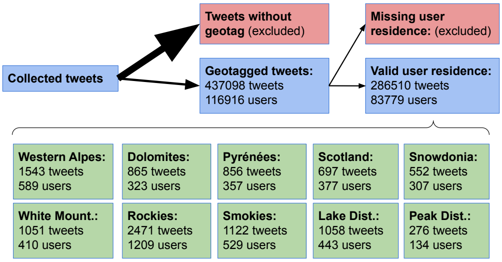



**Figure 5.** Data cleaning and filtering steps

thermore, we selected ten specific locations to compare and analyze recreational areas during Covid-19 within the USA and Europe. Each area is defined by one or more circles presented in Table 3, along with the number of tweets posted within these areas. For most of the circles, we set a radius of 50 kilometers, but some further adjustments were made to adapt to the dimensions of the given recreational area. For example, the radius is decremented in the case of nearby populated areas. Figure 5 includes a flow chart explaining the major data preprocessing steps we applied before filtering the data for the selected recreational areas.

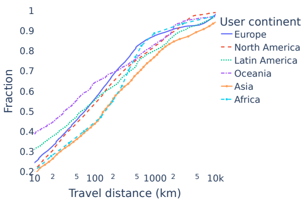



**Figure 6.** Cumulative distribution function for travel distance with respect to different residence continents.

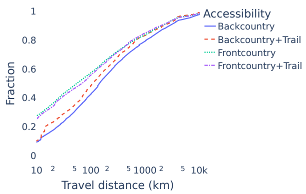



**Figure 7.** Cumulative distribution function for travel distance that visitors took to reach places that belong to different accessibility categories.

## 4. Results and Discussion

### *4.1. Users and Travel Distances*

In order to understand recreational user patterns shown by the Twitter posts we have collected, we calculate the Haversine distance between the location of each post and the residence of the corresponding user. This way, we get the approximate distance the user took to visit a given place. In the knowledge of these distances, we can extract traveling trends for users that belong to the same residence. For example, the travel distance distribution of users living on different continents is shown in Figure 6. Our results reveal that Twitter users located in Oceania and Latin America tend to go less further on average, which reflects the geographical properties of these regions. Namely, both continents consist of various islands, plus only a small fraction of Australian land is suitable for recreational activities, being usually close to populated areas.

Similarly, for a selected place, we can analyze the travel distance of the visitors. Figure 7 assesses the travel distance distribution for the accessibility categories described in Section 2.3. Our results indicate that users in general travel much further to reach backcountry locations compared to recreational areas in the frontcountry.

### *4.2. Accessibility Comparison*

This section further analyzes the travel distance with respect to different accessibility categories (e.g., frontcountry, backcountry). Almost two-thirds of the collected tweets were posted from frontcountry locations (see Figure 8). It means that users in our data rarely leave areas that are easily accessible through a motorized road or public transport network. Unfortunately, we do not have data from the pre-Covid period. Thus, we cannot properly quantify the effect of different Covid-19 waves or other travel restrictions on recreational patterns of Twitter users, but these events might have also encouraged people to visit frontcountry places more frequently.

In our next experiment, we compare travel distance patterns over time for users who live in Europe or North America. By choosing these regions, we cover almost 75 percent of all users (see Figure 3). By collecting data for more than a year, we had the chance to observe seasonal changes in travel distance for each accessibility category. Figure 9

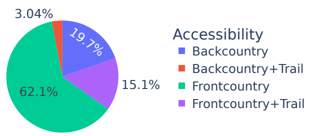



**Figure 8.** Distribution of accessibility categories in the collected data

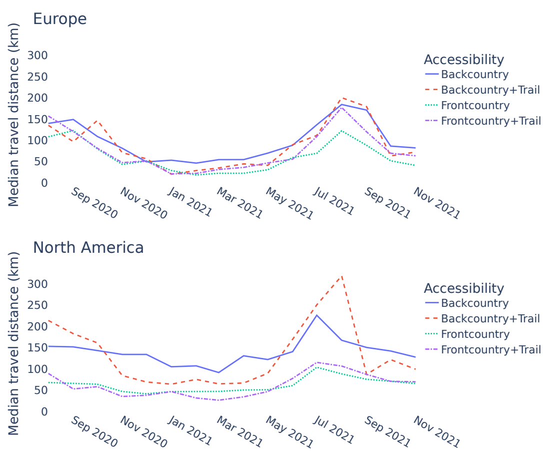



**Figure 9.** Median travel distance of visitors by location categories restricted to Europe and North America.

reveals that North American users are willing to travel further to reach backcountry locations even during the winter. On the other hand, half of all European tweets were posted no more than 50 kilometers from the corresponding user residence location during Q1 of 2021. This behavior might also be the effect of various Covid-related traveling restrictions that took place in several European countries during this period.

### 4.3. Comparison of Selected Recreational Areas

Our previous results were related to large geographic regions like continents or every backcountry location on Earth that our data covers. Here, we assess recreational user patterns on a more fine-grained level. We compare and analyze the travel distance taken by visitors of multiple recreational areas introduced in Section 3.3. It is important to note that users can post multiple tweets from the same area (e.g. photographers). In order to

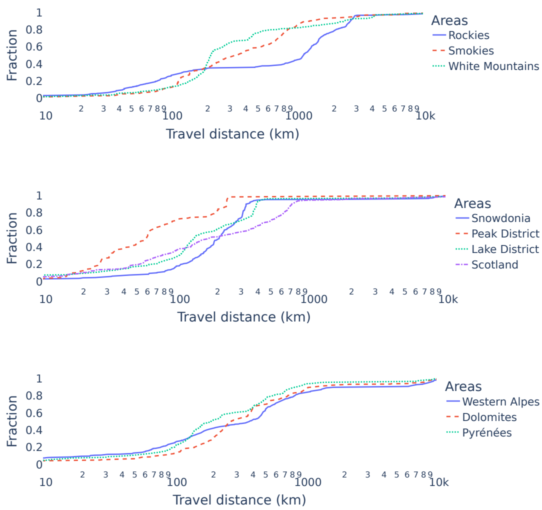
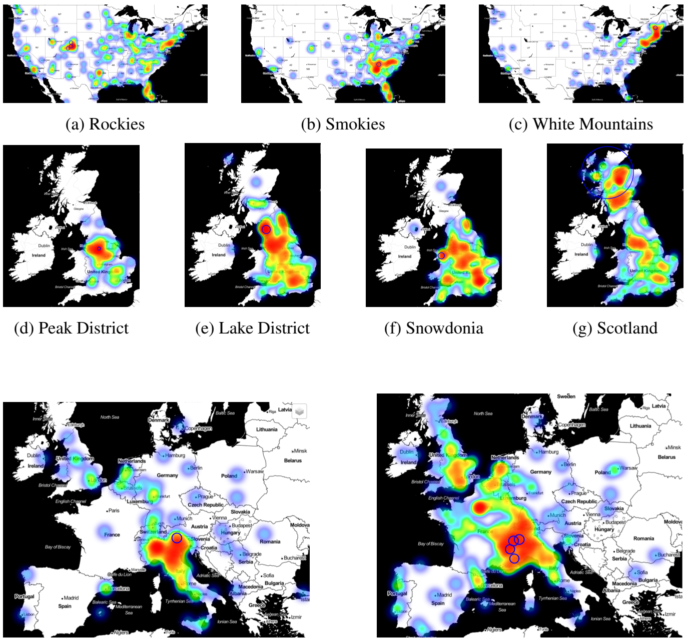



**Figure 10.** Cumulative distribution function for travel distance that visitors took to reach selected recreational areas in the USA **(top)**, UK **(middle)**, and continental Europe **(bottom)**.

avoid the possible user bias during our analysis, for each recreational area, we keep only a single post per user, the one with the largest travel distance.

In Figure 10, the cumulative distribution functions on travel distance clearly signal whether a given area is a popular tourist hotspot or is only visited by regional users. For example, almost half of the Rocky Mountains visitors come from its 200-kilometer radius, while most of its remaining tourists live further than 1000 kilometers. Thus, it is a real paradise for interstate hikers in contrast to the White Mountains, which addresses relatively more regional hikers. Figure 11 also expresses this behavior where the residence distribution of visitors is shown for each recreational area. Similar case studies are presented for the United Kingdom and continental Europe. The Peak District national park in the UK and the Dolomites in Italy are good examples of sites visited mostly by regional users. On the contrary, Lake District and Snowdonia national parks are more popular in the UK nationwide, while the area of the Western Alps indeed attracts international travelers from further distances. It is also interesting to observe that the selected recreational areas in the US and UK are more diverse in terms of their cumulative travel distance distribution function than the three continental European locations (Western Alps, Dolomites, Pyrénées), see Figure 10.

By deploying our logic-based location categorization tool, OPENLOSTCAT, we can further analyze travel distance patterns for different accessibility groups within these areas. Our results exhibit a significant separation of user interests for popular tourist



(a) Rockies (b) Smokies (c) White Mountains

(d) Peak District (e) Lake District (f) Snowdonia (g) Scotland

(h) Dolomites (i) Western Alpes

**Figure 11.** Visitor residence distribution for the selected recreational areas in the US **(first row)**, UK **(second row)**, and continental Europe **(third row)**. The covering circles for each selected area are marked by blue.

hotspots (e.g. Rockies, Western Alps); see Figure 12. For example, frontcountry locations are visited by users living much further than those interested in the backcountry. These are probably international or interstate travelers who only visit the most popular landmarks within these recreational areas or simply not familiar with backcountry trails or areas. On the other hand, for areas most visited by regional hikers (e.g. Dolomites, Pyrénées, Peak District), the median travel distance for the backcountry and frontcountry are more balanced. Interestingly, visitors of the UK national parks and similar areas are traveling the most for marked trails in the backcountry. This behavior was specific for these regions from the 10 selected recreational areas.


**Figure 12.** Median travel distance by accessibility category for the selected recreational areas in the USA **(top)**, UK **(middle)**, and continental Europe (**bottom**).

## 5. Conclusion and Future Work

In this paper, we presented an initial exploratory study on analyzing geotagged social media posts on a global scale with micro-level location categorization. Location types are modeled based on geographical features and their descriptive properties found in the proximity of each visited data point. The goal was to find out whether this approach is useful for identifying any potentially verifiable differences in large-scale visitation patterns of different location type settings, depending on geographical areas at multiple levels, seasonality, and traveling distance of visitors.

For such purposes, we designed and implemented OpenLostCat, a free, open-source, logic-based location categorization tool by which the modeler can define location categories over OpenStreetMap data found around a point on Earth. To achieve this, a JSON-based rule language can be used as a customized form of univariate first-order logic with convenience features such as implicit variable quantification.

We collected Twitter data of visitors by keyword-based filtering from August 2020 until November 2021, applied location categorization according to transportation and trail accessibility characteristics to each of the geotagged posts, determined the user home location of users where appropriate, then aggregated and visualized these data according to different aspects relevant for seeking potential large-scale visitation patterns for each location category.

Results of our experimental analysis are presented by the visited place types, their seasonal characteristics, and distribution of visitor travel distances. Geographical comparisons between North America and Europe, as well as for specific selected recreational areas show significantly different visitation patterns by users of geotagged tweets posted in those areas. Based on our collected data, heatmaps reveal the catchment areas of selected recreational areas, i.e., from where their identified Twitter visitors arrive. Observations can be refined according to the visited location categories, showing the median of distances traveled by each accessibility category. These findings may be subjects for further verification according to evidence from alternative sources or surveys in further studies since we have seen the low ratio of geotagged tweets, the inherent bias caused by keyword selection, as well as the variety in the popularity of Twitter amongst users in different areas.

Our main contribution on the methodological level is showing the feasibility and potential of enhancing social media content analysis with logic-based location modeling on a global scale by potentially characterizing any point on earth based on its local geographical features and looking at large scale visitation patterns by location categories such as home locations of users visited trailless natural areas compared to users at places accessible by motorized transportation. In contrast to most related studies, our major advantage is the global-scale approach, with no *a priori* particular focus area, that is achieved by the power of potentially global coverage of OpenStreetMap.

The main technical contribution of the study is the reusable open-source tool called OPENLOSTCAT, which does the actual location categorization with its simplified and convenient first-order-logic-based rule language for modeling location types, where the single-variable approach, the JSON formalism, and the implicit quantification mechanism provides simplicity in formulating many common rules or subexpression types and is capable of effective evaluation with large-scale datasets.

Although these initial results seem to be promising, interpretation of the actual survey outcomes needs caution due to data coverage and quality limitations, as potential biases are also reported in the related literature. Our study has shown the potential in our approach to discover relevant patterns and to be a viable source of information for complex surveys, in which such findings can be verified by other means. We highlight that our results are restricted to users posting geotagged tweets on Twitter with specific keywords related to hiking and nature-based recreation activities and whose location of residence can be resolved. Thus, our findings should not be interpreted as representative visitation patterns in general.

There are obvious differences in the popularity of Twitter usage in different countries, as shown by or data in Figures 3–4, which may highly affect the outcomes if we wanted our results to reflect actual visitor frequencies on a global scale. Our selected keywords are also far from complete and keyword-based filtering is somewhat biased by nature. It is a future issue whether a more fine-tuned filtering of social media posts data can be achieved. As location categorization is currently based solely on local features in close proximity, one must consider casual patterns while interpreting these categories. For instance, some city park locations will be categorized as backcountry in our terms (if not covered directly by a residential area polygon), which is completely acceptable by our category definition based on transportation and marked-trail-accessibility. However, it may be surprising from a common-sense viewpoint. OpenStreetMap data quality and coverage also affect the results in general, as some areas may have incomplete data, and the different trail designation and signage systems in different regions and countries, the phenomenon or concept of a marked trail will be very different. Unfortunately, we do not have data from the pre-covid era, so it is not possible at the moment to make comparisons with the time of the pandemic as we face limitations of the Twitter API with querying historical data. Finally, only a fraction of the collected data proved appropriate for our analysis as Twitter users generally do not tend to share geotagged tweets. Thus, it would be essential to characterize visitors who post geotagged tweets during their hiking, trekking, or other outdoor tours compared to others who do not, and so for which segment(s) of users are our results typical, compared to the whole population of visitors at specific areas. Nevertheless, our results revealed some relevant and interesting patterns, especially by the continental and area-based comparisons and the timeline charts, which can be validated by other types of surveys outside of the scope of this paper.

As a future prospect, this approach or its tools may be adapted to surveys of other fields with similar needs and characteristics. The rule definition language or even the location categorization method may also be extended to capture more sophisticated modeling of location settings. The flexibility and extensibility of the JSON format allows a potentially complex and heterogenous modeling language to be developed, where special keywords and prefixes may identify even the type or (sub)language of the rule being defined at some point. On the other hand, the fine-tuning of our results can be achieved by refining our keyword set in combination with more sophisticated methods for relevant post selection. For example, sentiment analysis, user, and location profiling can reveal additional details on the collected users and content.

Analyzing similar content posted to other social media platforms can make our analysis more robust and complementary to the coverage of Twitter data. Comparisons with field surveys such as questionnaire-based user studies may validate our results and orient their interpretation by discovering additional user-profile characteristics.

## References


[1] Introducing JSON. `https://www.json.org`, [Online; accessed 14-Jan-2022]

[2] Nominatim – open-source geocoding with OpenStreetMap data. `https://nominatim.org`, [Online; accessed 16-Jan-2022]

[3] OpenStreetMap. `http://www.openstreetmap.org`, [Online; accessed 13-Dec-2021]

[4] OpenStreetMap Wiki. `https://wiki.openstreetmap.org`, [Online; accessed 13-Dec-2021]

[5] OpenStreetMap Wiki. Walking routes. `https://wiki.openstreetmap.org/wiki/Walking_Routes`, [Online; accessed 13-Dec-2021]

[6] Overpass API. `https://wiki.openstreetmap.org/wiki/Overpass_API`, [Online; accessed 14-Jan-2022]

[7] Arndt, D., Schrijvers, T., De Roo, J., Verborgh, R.: Implicit quantification made explicit: How to interpret blank nodes and universal variables in Notation3 logic. Journal of Web Semantics **58**, 100501 (2019). https://doi.org/https://doi.org/10.1016/j.websem.2019.04.001

[8] Benkhard, B.: Determination of tourist flow patterns in a low mountain study area. Tourism & Management Studies **14**(3), 19–31 (2018). https://doi.org/10.18089/tms.2018.14302

[9] Bühler, G., Hermann, M., Lambertus, M.: Sehnsuchtsort natur in der digitalen leistungsgesellschaft. `https://sotomo.ch/site/sehnsuchtsort-natur` (05 2019), [Online; accessed 13-Dec-2021]

[10] Clark, R.N., Stankey, G.H.: The recreation opportunity spectrum: A framework for planning, management, and research. Tech. rep., Department of Agriculture, Forest Service, Pacific Northwest Forest and Range Experiment Station. (1979)

[11] Fisher, D., Wood, S., White, E., Blahna, D., Lange, S., Weinberg, A., Tomco, M., Lia, E.: Recreational use in dispersed public lands measured using social media data and on-site counts. Journal of environmental management **222**, 465–474 (06 2018). https://doi.org/10.1016/j.jenvman.2018.05.045

[12] Fisher, D.M., Wood, S.A., Roh, Y.H., Kim, C.K.: The geographic spread and preferences of tourists revealed by user-generated information on Jeju Island, South Korea. Land **8**(5) (2019). https://doi.org/10.3390/land8050073

[13] Ghermandi, A.: Analysis of intensity and spatial patterns of public use in natural treatment systems using geotagged photos from social media. Water research **105**, 297–304 (09 2016). https://doi.org/10.1016/j.watres.2016.09.009

[14] Ghermandi, A., Camacho, V., Trejo-Espinosa, H.: Social media-based analysis of cultural ecosystem services and heritage tourism in a coastal region of Mexico. Tourism Management **77** (10 2019). https://doi.org/10.1016/j.tourman.2019.104002

[15] Gosal, A., Geijzendorffer, I., Václavík, T., Poulin, B., Ziv, G.: Using social media, machine learning and natural language processing to map multiple recreational beneficiaries. Ecosystem Services **38** (06 2019). https://doi.org/10.1016/j.ecoser.2019.100958

[16] Heikinheimo, V., Di Minin, E., Tenkanen, H., Hausmann, A., Erkkonen, J., 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**, 85 (03 2017). https://doi.org/10.3390/ijgi6030085

[17] Hoffmann, S.: Waymarked trails: Hiking. `http://hiking.waymarkedtrails.org`, [Online; accessed 11-Jan-2022]

[18] Joyce, K., Sutton, S.: A method for automatic generation of the recreation opportunity spectrum in New Zealand. Applied Geography **29**(3), 409–418 (2009). https://doi.org/https://doi.org/10.1016/j.apgeog.2008.11.006

[19] Leung, D., Law, R., van Hoof, H., Buhalis, D.: Social media in tourism and hospitality: A literature review. Journal of Travel & Tourism Marketing **30**(1-2), 3–22 (2013). https://doi.org/10.1080/10548408.2013.750919

[20] Lukács, G., Molnár, A.J.: OpenLostCat – a location categorizer based on OpenStreetMap. `https://github.com/lukacsg/openlostcat`, Github code and documentation repository [Online; modified 17-Jan-2022]

[21] Martínez-López, J., Bagstad, K.J., Balbi, S., Magrach, A., Voigt, B., Athanasiadis, I., Pascual, M., Willcock, S., Villa, F.: Towards globally customizable ecosystem service models. Science of The Total Environment **650**, 2325–2336 (2019). https://doi.org/https://doi.org/10.1016/j.scitotenv.2018.09.371

[22] Moreno, R., Méndez, P., Ros-Candeira, A., Alcaraz-Segura, D., Santamaría, L., Ramos-Ridao, Á., Revilla, E., García, F., Vaz, A.: Evaluating tourist profiles and nature-based experiences in biosphere reserves using flickr: Matches and mismatches between online social surveys and photo content analysis. Science of The Total Environment **737**, 140067 (06 2020). https://doi.org/10.1016/j.scitotenv.2020.140067

[23] Muhar, A., Arnberger, A., Brandenburg, C.: Methods for visitor monitoring in recreational and protected areas: An overview. In: Arnberger, A., Brandenburg, C., Muhar, A. (eds.) Monitoring and Management of Visitor Flows in Recreational and Protected Areas. pp. 1–6 (01 2002)

[24] Park, S., Kim, J.W., Lee, Y.K., Ok, C.: Visualizing theme park visitors' emotions using social media analytics and geospatial analytics. Tourism Management **80** (04 2020). https://doi.org/10.1016/j.tourman.2020.104127

[25] Raadik, J., Cottrell, S.P., Fredman, P., Ritter, P., Newman, P.: Understanding recreational experience preferences: Application at Fulufjället National Park, Sweden. Scandinavian Journal of Hospitality and Tourism **10**(3), 231–247 (2010). https://doi.org/10.1080/15022250.2010.486264

[26] Roberts, H., Sadler, J., Chapman, L.: Using twitter to investigate seasonal variation in physical activity in urban green space. Geo: Geography and Environment **4**, e00041 (07 2017). https://doi.org/10.1002/geo2.41

[27] Sinclair, M., Mayer, M., Woltering, M., Ghermandi, A.: Using social media to estimate visitor provenance and patterns of recreation in Germany's national parks. Journal of Environmental Management **263**, 110418 (06 2020). https://doi.org/10.1016/j.jenvman.2020.110418

[28] Teles da Mota, V., Pickering, C.: Using social media to assess nature-based tourism: Current research and future trends. Journal of Outdoor Recreation and Tourism **30**, 100295 (2020). https://doi.org/https://doi.org/10.1016/j.jort.2020.100295

[29] Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., Toivonen, T.: Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. Scientific Reports **7** (12 2017). https://doi.org/10.1038/s41598-017-18007-4

[30] Vu, Q., Li, G., Law, R., Zhang, Y.: Tourist activity analysis by leveraging mobile social media data. Journal of Travel Research **57** (06 2017). https://doi.org/10.1177/0047287517722232

[31] Wadhams, J.: JsonLogic. `https://jsonlogic.com/`, [Online; accessed on 14-Jan-2022]

[32] Wilkins, E.J., Wood, S.A., Smith, J.W.: Uses and limitations of social media to inform visitor use management in parks and protected areas: A systematic review. Environmental management **67**, 120–132 (2021). https://doi.org/10.1007/s00267-020-01373-7

[33] Wood, S., Guerry, A., Silver, J., Lacayo-Emery, M.: Using social media to quantify nature-based tourism and recreation. Scientific reports **3**, 2976 (10 2013). https://doi.org/10.1038/srep02976

[34] Zeman, J.J.: A system of implicit quantification. The Journal of Symbolic Logic **32**(4), 480–504 (1967), `http://www.jstor.org/stable/2270176`