

Exploring problems through social media: the case of Beach Quality

Carlos Perrián-Pascual

Universitat Politècnica de València
Paranimf 1, 46730 Gandia (Valencia), Spain
jopepas3@upv.es

I. INTRODUCTION

Social networking has become a global communication phenomenon, where users post their text messages, images, video, audio, or a combination to express opinions, report facts, and show situations of interest. A current line of research related to social-media platforms is social sensing (Goswami and Kumar, 2016). One of the objectives of this field involves the analysis of digital communications to detect events or situations in the real world where social-media users play the role of social sensors (Sakaki et al., 2013; Musto et al., 2015; Arthur et al., 2018). In this context, this chapter is part of the research carried out in the framework of ALLEGRO (Adaptive muLti-domain social-media sEnsinG fRamewOrk), a multimodal system (i.e. text, audio and image processing) for reconstructing the state of society as interpreted by the collective intelligence of social-media users. In other words, ALLEGRO analyses user-generated content (UGC) (Barbosa dos Santos, 2022) to extract relevant knowledge for detecting problems in a given community, thus considering users as witnesses of society. ALLEGRO consists of two modules (i.e. Data Analysis and Data Fusion). The Data Analysis module comprises a dedicated component for each type of data to be analysed in the UGC unit (e.g. a tweet). However, each type of data is processed separately. Therefore, data from the different components are aggregated in the Data Fusion module, where the quality of aggregated data is enhanced by rejecting irrelevant information, minimising redundancy, resolving inconsistencies, and completing missing information.

One of the critical components in the Data Analysis module of ALLEGRO is DIAPASON (unified hybrid Approach to microtext Analysis in Social-media crowdsensing), which explores English and Spanish text messages in UGC units by integrating natural language processing, text mining, and knowledge engineering techniques. This chapter focuses on the DIAPASON ontology, illustrated by the case of Beach Quality problems. To this end, we describe the knowledge acquisition, specification and conceptualisation stages that allow modelling this part of the ontology. Moreover, we demonstrate that some problem types associated with this case pertain to various domains and dimensions in the social and physical realms of the DIAPASON ontology.

II. THE DIAPASON ONTOLOGY

1. Overview

The taxonomic hierarchy of the DIAPASON ontology consists of four levels of thematic granularity, each including a different kind of element: problem realm, problem dimension, problem domain, and specific problem type. The upper level is structured into two problem realms: the social realm, which subsumes the LIVING, ECONOMY and GOVERNANCE problem dimensions, and the physical realm, which

subsumes the MOBILITY, INFRASTRUCTURE and ENVIRONMENT problem dimensions. The modelling of the dimension level primarily results from Smart City frameworks (Giffinger et al., 2007; Govada et al., 2017; Appio et al., 2019). At the next level, each problem dimension is organised in problem domains (e.g. COMMUNITYFACILITY, CRIMEJUSTICE, ECOLOGICALHAZARD, HEALTH, MOTORIZEDVEHICULARTRAFFIC, and POPULATION, among others) on which citizens can show an attitude of disapproval towards some specific aspect of the community. Finally, the lower level describes particular types of community problems that can affect some, most or all citizens for each problem domain (e.g. BEACHCROWDING, BEACHPARKING, BEACHSECURITY, MARINESPILL, and SEAWATERTRANSPARENCY, among others).

2. Social problem dimensions and domains

A. Problem domains in the LIVING dimension

The problem domains in the LIVING dimension were primarily determined by considering Quality of Life (QoL) reports. The global economic crisis of 2008 triggered a lively debate on the QoL issue, which resulted in the report produced by the commission led by Joseph Stiglitz, Amartya Sen, and Jean-Paul Fitoussi. Since then, several initiatives of QoL indicators have been developed based on the recommendations of the Stiglitz-Sen-Fitoussi Report (Stiglitz et al., 2009), e.g. Eurostat (2017), Helliwell et al. (2020), and OECD (2020). Apart from such international initiatives, there have also been several national experiences with work on QoL indicators relevant to our research, e.g. Bundesregierung (2018), Office for National Statistics (2019), and Istituto Nazionale di Statistica (2020). Sociological encyclopedias and textbooks (e.g. Parrillo, 2008; Eitzen et al., 2014; Marginean, 2014; Seccombe and Kornblum, 2020) were also examined to determine the problem domains in this dimension. In the end, eleven domains were identified:

- a) CRIMEJUSTICE deals in general with the integrity of the individual, as the measure of right and the response to wrong, and in particular with security issues such as criminality, violence or vandalism. It is worth mentioning that the perceived lack of physical safety may affect subjective well-being more than the actual effects of a physical threat because the effects of crimes are socially magnified, thus affecting the QoL of many people that feel afraid.
- b) CULTURE encompasses aspects such as language, literature, music and song, non-verbal communication, religion or belief systems, rites and ceremonies, sport and games, food, clothing, customs and traditions, and the arts, among others, through which individuals, groups of individuals, and communities express their humanity and the meaning they give to their existence. UN Committee on Economic, Social and Cultural Rights (2009) profoundly influenced this notion of *culture*.
- c) EDUCATIONTRAINING focuses on the knowledge, competences, skills and activities related to formal education and lifelong learning.
- d) HEALTH refers to three main areas: (a) health status, i.e. physical and mental problems of the individual, (b) risk factors for the individual's health, resulting from lifestyle or personal choices (e.g. obesity, dietary habits, tobacco, and alcohol consumption), and (c) access to healthcare services, facing barriers such as cost, distance, and waiting times, among others.

- e) **INEQUALITY** includes, but is not limited to, three primary types of discrimination of a group in a society: i.e. racial and ethnic, gender, and economic inequality. Eurostat (2017) explained two ways to approach this topic: it can be a separate domain or can be explored as a distinct aspect within some domains. For example, whereas Bundesregierung (2018) chose the first approach, Eurostat (2017) and OECD (2020) opted for the latter. In the DIAPASON ontology, Inequality was regarded as a separate domain. Therefore, a given problem of discrimination can be subsumed by one or more domain classes, where **INEQUALITY** should be one of them.
- f) **LEISURE** refers to the free time outside work and the desirable activities that people undertake during that time because of their intrinsic value rather than necessity. **LEISURE** embraces sporting activities and recreation, i.e. activities that people do to refresh their bodies and minds so that their free time becomes more enjoyable (e.g. walking, reading, playing games, dancing, watching TV, and listening to music, to name a few).
- g) **POPULATION** is a broad category that primarily groups together not only those aspects reflecting cumulative processes on population (e.g. life expectancy and increase/decrease of the population) but also issues related to specific segments of the population (e.g. older people and people with functional diversity).
- h) **RESEARCHINNOVATION** focuses on any issue related to the creation, application or diffusion of knowledge (e.g. patent protection and brain drain).
- i) **SOCIALRELATIONSHIP** is based on two main areas: personal relationships (e.g. family and friends) and community relationships, i.e. civic engagement (e.g. volunteering).
- j) **SUBJECTIVEWELLBEING** represents a personal exploration of how the individual feels. In particular, the individual evaluates their life as a whole in terms of dissatisfaction, which requires some reflection on multiple aspects of past experiences regarding personal standards (i.e. overall assessment of life), or the individual indicates negative emotions about their current situation (i.e. affective state).
- k) **TERRORISMWAR** involves extreme violence motivated by political, ideological or strategic aims and inflicted by one group of individuals against another (Gomes, 2012).

Although there is a fair amount of consensus about the inventory of social problem domains in this dimension, it is crucial to explain the decisions taken when this part of the DIAPASON ontology was modelled:

- a) Economy, politics, and the natural environment are multidimensional concepts, so they were treated as separate problem dimensions rather than as simple problem domains in the **LIVING** dimension. In particular, **ECONOMY** and **GOVERNANCE** were regarded as the other two social dimensions and **ENVIRONMENT** as a problem dimension in the physical realm of the Smart City framework.
- b) Poverty was not represented as a separate problem domain because all the specific problem types about economic inequality pertain to the **LIVING** and **ECONOMY** dimensions.
- c) Housing and urbanisation refer to the places where we live, thus related to the **INFRASTRUCTURE** dimension in the physical realm of the Smart City framework.

- d) Any issue related to the quality of services was assigned to INFRASTRUCTURE, e.g. social services (i.e. health and social care) and public utilities (i.e. energy, water, and waste), or MOBILITY (i.e. transport), which are problem dimensions in the physical realm.

B. *Problem domains in the ECONOMY dimension*

In the DIAPASON ontology, the ECONOMY dimension corresponds to material well-being, which is about satisfaction with a variety of economic concerns: personal and family financial situation, perceived income adequacy, material possessions and the extent to which physical needs are met (i.e. standard of living), feelings of financial security, feelings about access to major goods and services, aspirations and attainment of material goods, and marketplace activities (Sirgy, 2018). Following Eurostat (2017), there are two types of domains in this dimension: those related to material living conditions (i.e. income, consumption, and wealth) and those related to productive activities (i.e. work). On the one hand, income (which determines the purchasing power) allows people to satisfy their needs and pursue their goals, and wealth (i.e. financial assets) makes it possible to sustain these choices over time. Moreover, the consumption of goods and services results in expenditure, which represents a critical determinant in personal finance. On the other hand, work primarily focuses on the quantity (e.g. unemployment and underemployment) and the quality of employment (e.g. benefits, stability, over-qualification, health and safety, and work-life balance). Finally, beyond personal material well-being (i.e. microeconomic perspective), other studies (Bundesregierung, 2018; Office for National Statistics, 2019) show that QoL can be affected by a variety of macroeconomic issues that strengthen or weaken the economy at a national level, such as inflation, investment rate, national debt, and public expenditure, among others. Therefore, the ECONOMY dimension comprises five main domains: CONSUMPTION, INCOME, MACROECONOMICS, WEALTH, and WORK. When considering the categorisation of this dimension, it was noticed that many problem types could be linked to both ECONOMY and LIVING. For example:

- economic disparities within society are also part of the INEQUALITY domain,
- burnout from an excessive workload or accidents at the workplace are also related to the HEALTH domain, and
- tight working schedules and lengthy commuting, which hinder the reconciliation of work and family responsibilities, also correspond to the SOCIALRELATIONSHIP domain.

C. *Problem domains in the GOVERNANCE Dimension*

The GOVERNANCE dimension, identified with politics, was grounded on Christensen's model (2016). From a sociological perspective, this author explored the interpretation of political dissatisfaction based on two issues: political support, which is an issue of satisfaction with democracy and trust in institutions, politicians and political parties, and subjective political empowerment, which concerns the extent to which citizens feel they can influence political decisions. This analysis framework constructs a typology of citizens concerning their political attitudes, thus resulting in satisfied, unsupportive, disempowered or disenchanted citizens. In our case, however, the entity that raises the political attitude was the focus of attention: politicians, political parties, institutions, and the political system in the case of political support, and political activity in the case

of subjective political empowerment. Therefore, the GOVERNANCE dimension comprises five domains: INSTITUTIONS, POLITICALACTIVITY, POLITICALPARTIES, POLITICALSYSTEM, and POLITICIANS. It should be noted that the distinctive transversality of inequality also affects this dimension, as illustrated in problem types such as the absence of women in a national parliament.

3. Physical problem dimensions and domains

A. Problem domains in the MOBILITY dimension

The MOBILITY dimension and its domains were characterised by the idea that three essential concepts related to transport systems can affect the quality of urban life: traffic, mobility, and accessibility (Litman, 2003). In this regard, traffic refers to vehicle movement, mobility refers to the movement of people or goods, and accessibility refers to the ability to reach opportunities (e.g. destinations, goods, and services). In the DIAPASON ontology, *mobility* is defined as the capacity of a given population to move from one place to another, where differences lie in the mode of travel (Böhler-Baedeker and Durant, 2015). From this context, a functional perspective of the definition of *traffic* was employed, thus referring to “activities designed to move objects and people around, with the help of means of transportation and the corresponding infrastructure” (Götz, 2014, p. 6705). Therefore, mobility is a broader concept than traffic in the DIAPASON ontology. Finally, as in Litman (2003), *accessibility* is understood as the ultimate goal of transport, so it does not refer to the movement itself. However, when accessibility is thought of in physical terms, i.e. “physical access to home, work, or community as exemplified by ramps into buildings, doors that are wide enough for a wheelchair to fit through” (Harniss, 2014, p. 9), this issue is treated in the INFRASTRUCTURE dimension. As a result, MOBILITY is modelled as a dimension that is structured into different types of traffic, each one corresponding to a distinct domain: MOTORIZEDVEHICULARTRAFFIC (e.g. buses, cars, e-scooters, motorcycles, taxis, and trucks), NONMOTORIZEDVEHICULARTRAFFIC (e.g. bicycles, skateboards, and roller-skates), and PEDESTRIANTRAFFIC.

B. Problem domains in the INFRASTRUCTURE dimension

The INFRASTRUCTURE dimension refers to the built facilities and services that give cities their form and function. A traditional view of the concept of *infrastructure* is found in Neuman (2005), who differentiated four categories: community facilities (e.g. schools, parks, hospitals, libraries, prisons, emergency services, and civic buildings), public works (e.g. roads and bridges, dams and canals, ports and airports, and railways), telecommunications (e.g. mobile, radio, television, and the Internet), and utilities (e.g. gas and electricity, water supply, sewerage, and waste collection and disposal). Howes and Robinson (2005) provided a more detailed account of physical infrastructure according to function, organised into three types: social, trade, and technical. On the one hand, social infrastructure, i.e. facilities related to human and social welfare, is classified into education and culture (e.g. schools, libraries, churches, and colleges), health and social services (e.g. hospitals, clinics, and advisory centres), public administration (e.g. fire station, city hall, police station, prison, and government buildings), and sport and recreation (e.g. stadium, swimming pool, zoo, park, museum, theatre, and cinema). Indeed, this type of infrastructure corresponds to traditional community facilities, i.e. “cultural, educational, recreational, health and civic structures

or places administered by local government authorities and available to the public” (McShane, 2008, p. 1). On the other hand, trade infrastructure, i.e. facilities directly used to produce goods and services, comprises factories, offices, shops, and warehouses. Finally, technical (or economic) infrastructure, i.e. long-lived networked, capital-intensive and engineered structures indirectly supporting economic production, is classified into public utilities (e.g. power, water supply, piped gas, and telecommunication), public works (e.g. roads, bridges, and dams), sanitary (e.g. sewerage and solid waste collection), and transport (e.g. railways, ports, and airports). In this context, the INFRASTRUCTURE dimension was structured into six domains, i.e. COMMUNITYFACILITY, HOUSING, PUBLICWORK, TELECOMMUNICATIONS, TRADE, and UTILITIES. In particular, it was based on Neuman (2005) to identify domains such as COMMUNITYFACILITY, PUBLICWORK, TELECOMMUNICATIONS and UTILITIES and on Howes and Robinson (2005) for TRADE. Finally, the urban landscape was completed with HOUSING, i.e. buildings or other types of shelter construction in which people live (Parrillo, 2008), which includes specific problem types about unwanted housing conditions to live a decent life, such as structural problems of the dwelling (e.g. damp walls and leaking roof) and overcrowding.

It should be noted that specific problem types in INFRASTRUCTURE are related not only to physical places but also to the accessibility and effectiveness of the services provided by these facilities (Istituto Nazionale di Statistica, 2012; Office for National Statistics, 2019). For example, irregularities in the water supply become a problem associated with UTILITIES and limited broadband coverage with TELECOMMUNICATIONS. Istituto Nazionale di Statistica (2012) stated three main areas of services that can be analysed, i.e. social services (i.e. health and social care), public utilities (i.e. energy, water, and waste) and mobility (i.e. transport). Accordingly, this dimension has many specific problem types linked to other dimensions. For example, the shortage of beds in healthcare facilities is related to both COMMUNITYFACILITY in the INFRASTRUCTURE dimension and HEALTH in the LIVING dimension, and people killed or injured in traffic accidents to any of the domains in the MOBILITY dimension, PUBLICWORK in the INFRASTRUCTURE dimension, and HEALTH in the LIVING dimension.

C. *Problem domains in the ENVIRONMENT dimension*

The ENVIRONMENT dimension is about how the environment can affect our QoL and how our activities can affect the environment. Therefore, this dimension deals with threats to the environment (e.g. various forms of pollution) and threats from the environment (e.g. drought), causing damage to human life, health, and property. As a result, the domains in the ENVIRONMENT dimension were organised according to the common types of hazards identified in books such as Lindell et al. (2006), Schwab et al. (2007), and Haddow et al. (2011), i.e. ECOLOGICALHAZARD, GEOLOGICALHAZARD, HYDROLOGICALHAZARD, and METEOROLOGICALHAZARD. ECOLOGICALHAZARD originates from the interactions of human beings with the environment, e.g. oil spills, pollution, and wildfires. GEOLOGICALHAZARD occurs with processes related to the ground—including internal geological processes, driven by the movement of tectonic plates (e.g. earthquakes, tsunamis, and volcanic eruptions), and external geomorphological processes, involving landform changes (e.g. landslides and avalanches). HYDROLOGICALHAZARD involves water processes, such as drought and floods. METEOROLOGICALHAZARD is driven by processes related to temperature and wind, including hail, heat/cold waves, and windstorms (e.g. cyclones, hurricanes,

tornadoes, and typhoons). Most studies on well-being and QoL (Istituto Nazionale di Statistica, 2012; Eurostat, 2017; Bundesregierung, 2018; Office for National Statistics, 2019; OECD, 2020) deal with threats to the natural environment (e.g. exposure to air pollution and noise, climate change, and water, air and soil quality) as a critical dimension. However, it should be noted that, although studies such as Eurostat (2017) and OECD (2020) also include access to green areas, this issue is treated within the COMMUNITYFACILITY domain in INFRASTRUCTURE.

Using social sensors to help emergency managers in decision-making about environmental issues, such as earthquakes (Sakaki et al., 2013) and hurricanes (Huang and Xiao, 2015), has become a relevant research topic over the last decade. Crowe (2012, pp. 102-103) explained that “initiating protocols and systems to monitor social media conversations—particularly during disasters—is critical for both emergency public information and situational awareness”. Collecting and reviewing social-media information in real time can help emergency managers respond to incidents efficiently and effectively by mobilising in-situ stakeholders such as firefighters, police officers or medical staff. In this regard, the hazard-oriented modelling of the ENVIRONMENT dimension facilitates using the DIAPASON ontology for such emergencies.

III. THE CASE OF BEACH QUALITY

1. Knowledge acquisition

To illustrate ontological modelling in DIAPASON, we take the case of Beach Quality problems. To this end, the first stage consists in identifying potential problems by adopting two primary strategies: a manual approach, where we explored scholarly studies (e.g. books, papers, reports, etc.), and a semi-automatic approach, where we explored social-media posts on the Internet (e.g. tweets).

On the one hand, as regards the scholarly literature, research papers such as Ariza et al. (2010) and Sardá et al. (2014), who created the Beach Quality Index (BQI), served as the guideline for developing this topic. BQI was devised to examine significant environmental features, together with their impacts and effects, for measuring and assessing Beach Quality. Indeed, the BQI is a useful tool for environmental management systems for beaches, as it evaluates the three environmental functions of the socio-ecological system of urban and urbanised beaches:¹ recreational, natural, and protective. For each component, a set of partial indices and several indicators for each partial index are also defined. The different importance of partial indices and indicators is regulated by weighted coefficients so that the corresponding function can be correctly assessed. Thus, detecting strengths and weaknesses in beach environments can help local managers in their decision-making processes. Table 1 shows the partial indices included in the BQI.

Table 1. Structure of the BQI.

Function	Partial Index	Description
Recreational	Water Quality	Detecting bacteriological indicators of faecal contamination in seawater

¹ Ariza et al. (2010) and Sardá et al. (2014) classified beaches into two different types: urban beaches, i.e. those located in the main town center (high density), and urbanised beaches, i.e. those located in residential areas on the outskirts of a town (low density). Natural beaches were not considered in these studies.

	Beach Crowding	Measuring the level of crowdedness
	Environmental Quality	Monitoring the aesthetic and hygienic environmental quality of beaches
	Services and Facilities	Monitoring the adequate provision of services and facilities
	Activities	Detecting the presence of annoying behaviour
	Access and Parking	Assessing the accessibility to beaches
	Comfort	Monitoring aspects that affect users' recreational experience
	Surrounding Area Quality	Monitoring the degradation of the coastal landscape
	Beach Safety	Measuring safety and rescue services
Natural	Natural Conditions	Monitoring the quality of the natural system
	Water-Sand Pollution	Monitoring frequent pollution events that cause beach closures
	Physical Quality	Monitoring changes in the physical properties of beaches by human activities
Protective	Protection	Monitoring the vulnerability of coastal facilities

On the other hand, we compiled a subcorpus of text messages from Twitter. Indeed, one of the goals of the ALLEGRO project is to construct a medium-sized corpus about social problems to be employed as a training dataset for deep learning-based text classification. Regarding the corpus design, we adopted a divide-and-conquer approach. In particular, our social-problem corpus is structured into many subcorpora, each oriented to specific problem types. Therefore, whereas the corpus is characterised by linguistic heterogeneity, each subcorpus displays more homogeneity. As a result, this division of labour considerably reduces the time of corpus compilation. As shown in Figure 1, we developed Twitter Searcher for the ALLEGRO research team, a user-friendly application that accesses the Twitter API.

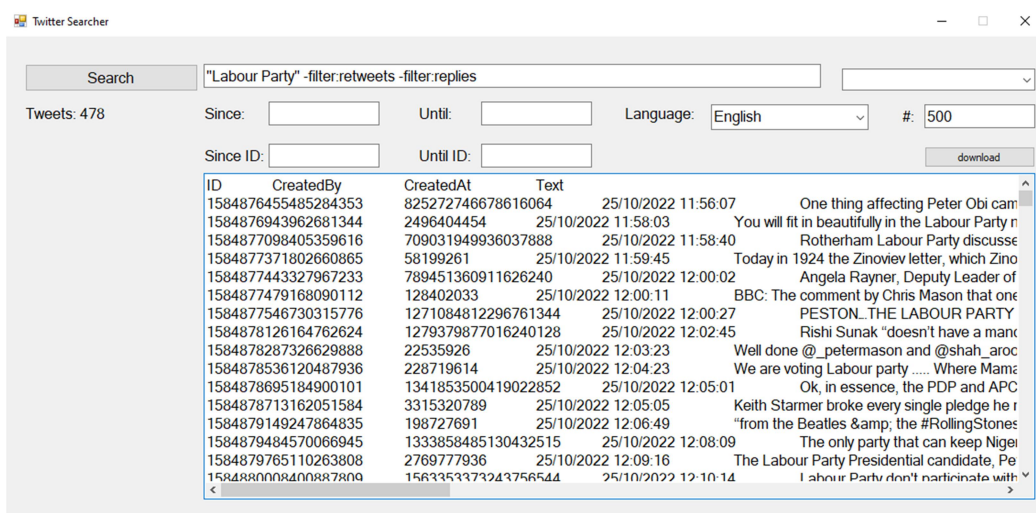


Figure 1. Twitter Searcher.

In this application, the search string has the same flexibility and limitations as the Twitter API, so researchers can employ the standard search filters in Twitter to build valid queries in Twitter Searcher.² For subcorpus compilation, we tend to make queries based on hashtags, usernames or named entities, thus minimising the bias introduced by keyword-based approaches in corpus development, i.e. searching for tweets containing selected keywords. Moreover, we filter out retweets and replies to other tweets (i.e. *-filter:retweets -filter:replies*). Researchers can also set some query parameters through the graphical user interface. For example, apart from determining the number of tweets to be retrieved, we can also search messages that Twitter has classified as being of a particular language, tweets created before or after a given date, and tweets including an ID greater than (i.e. more recent than), less than (i.e. older than) or equal to a specified ID. Only the language is a required parameter in Twitter Searcher. Once the search query is specified, the application compiles a collection of tweets automatically, recording the tweet ID, the user ID that created the tweet, the creation date of the tweet, and the text message. It should be noted that texts are pre-processed—particularly, URLs are removed, and tabs, new lines, and multiple consecutive spaces are replaced with a single space. Moreover, duplicate tweets are discarded. From time to time, researchers make further queries based on the same or different hashtags, usernames or named entities to expand a given subcorpus. When reusing the same query, the since-ID parameter is helpful to return tweets not included in the existing subcorpus.

In this regard, two controversial issues about Twitter could seemingly distort research results: restrictions on data acquisition and the veracity of information. However, such issues do not really undermine the adequacy of Twitter as an information-providing platform for social sensing. On the one hand, the Twitter Streaming API can only return up to 1% of all content published at a given time.³ However, Ayora et al. (2018) empirically demonstrated that neither the lack of completeness nor the latency of Twitter data results in flawed data that could lead to wrong conclusions. On the other hand, Doran et al. (2016, p. 60) presented the reasons for relying on the collective information conveyed by a stream of UGC:

social media posts that people share continuously as they observe the world around them are likely to be authentic and truthful, otherwise users stand to lose credibility within their offline social networks, which may include their friends, family and co-workers [Moreover] false information provided by a few is unlikely to lead to misleading inferences because of the truthful information shared by an overwhelming majority.

2. Specification

From the knowledge obtained in the previous stage, we could identify the main Beach Quality problems and then extract the most distinctive terms for each problem. Regarding problem identification, several decisions were made:

- a) Water Quality Index, Beach Crowding Index, Activities Index, Comfort Index, Natural Conditions Index, and Protection Index directly resulted in the problems

² Standard search operators are described at <https://developer.twitter.com/en/docs/twitter-api/v1/rules-and-filtering/search-operators>

³ <https://dev.twitter.com/streaming/public>

Seawater Quality, Beach Crowding, Beach Nuisance, Beach Soil, Beach Natural Conditions, and Beach Protection, respectively.

- b) Environmental Quality Index was split into seven problems (i.e. *Beach Jellyfish, Beach Seaweed, Marine Litter, Marine Spill, Seawater Colour, Seawater Odor, and Seawater Transparency*), Access and Parking Index into five problems (i.e. *Beach Pedestrian Access, Beach Parking, Beach Private Car, Beach Public Transport, and Beach Bicycle*) and, finally, Services and Facilities Index and Beach Safety Index into two problems (i.e. *Beach Amenities* and *Beach Facilities* for the former, and *Beach Safety* and *Beach Security* for the latter).
- c) Surrounding Area Quality Index and Physical Quality Index were discarded because they are irrelevant for social sensing, i.e. they do not consider whether natural conditions are good or bad for beach users. Water-Sand Pollution Index was also ignored, in this case because most of the community issues covered by this partial index can be found in *Marine Spill*.

The following step in this stage consisted in constructing a single statement in English for each of the above problems, including the most distinctive terms as problem descriptors. It is important to consider that problem statements, which aim to describe problem types in natural language, should be precise, concise, and complete. On the one hand, problem statements should be precise, presenting only true information about the corresponding problem types. On the other hand, problem statements should be concise, containing meaningful terms for the described problem, where semantically similar terms are discarded since they can be automatically retrieved from language resources. Finally, problem statements should be complete, covering all relevant terms for the problem type. To illustrate, *Marine Spill* was described with the statement “There is oil, spill, tar or dead fish on the sand or the sea”.

3. Conceptualisation

The previous stage returns unstructured knowledge in the form of problem statements. The conceptualisation stage is aimed at converting this knowledge into structured conceptual representations (i.e. problem schemas). We can say that DIAPASON understands a given UGC item when the system can detect whether or not the UGC reflects some social concern and, in case of a problem, recognise the problem type described by the user. Therefore, the interpretation of UGC involves mapping natural-language text into problem schemas, which will be subsequently used to perform social text analytics.

First, we converted the problem statements in the previous stage into problem types, as shown in Table 2. This task usually involves merging or splitting problem statements into one or several problem types, respectively.

Table 2. Beach-related problem types.

#	Problem type	#	Problem type
1	BEACHAMENITY	12	BEACHSAFETY
2	BEACHBICYCLEPARKING	13	BEACHSEAWEED
3	BEACHCARPARKING	14	BEACHSECURITY
4	BEACHCROWDING	15	BEACHSIGNPOSTING
5	BEACHFACILITIES	16	BEACHSOIL
6	BEACHJELLYFISH	17	MARINELITTER

7	BEACHNATURALCONDITION
8	BEACHNUISANCE
9	BEACHPEDESTRIANACCESS
10	BEACHPROTECTION
11	BEACHPUBLICTRANSPORT

18	MARINESPILL
19	SEAWATERCOLOUR
20	SEAWATERODOUR
21	SEAWATERQUALITY
22	SEAWATERTRANSPARENCY

Second, we conceptualised the semantics of problem types through problem schemas, i.e. conceptual representations for problem-type understanding. To illustrate the meaning representation of problem types, Example (1) shows the problem schema of MARINESPILL.

(1) (((oil-114980579 | spill-115049594 | tar-114911704) | (dead-300095280 & fish-102512053)) & (sand-115019030 | sea-109426788))

As can be seen, the main building blocks in problem schemas are concepts, which take the form of WordNet synsets (Fellbaum, 1998), and logical operators, which can connect concepts or expressions containing concepts. Expressions employ round brackets to determine their scope. It should be noted that we introduce each synset in (1) with an English word (e.g. sand-115019030) to facilitate the understanding of the meaning representation. However, such words play no role during the processing of problem schemas. As a given concept can be linguistically realised by several lexical units in different languages, we devised several methods for the linguistic projection of problem schemas. Indeed, this linguistic projection is a critical step when problem schemas are used with machine-learning models for text-mining tasks, i.e. topic detection and keyword recognition.

Third, we included the problem types in our ontological model, as in Figure 2, where they are represented as small numbered boxes.

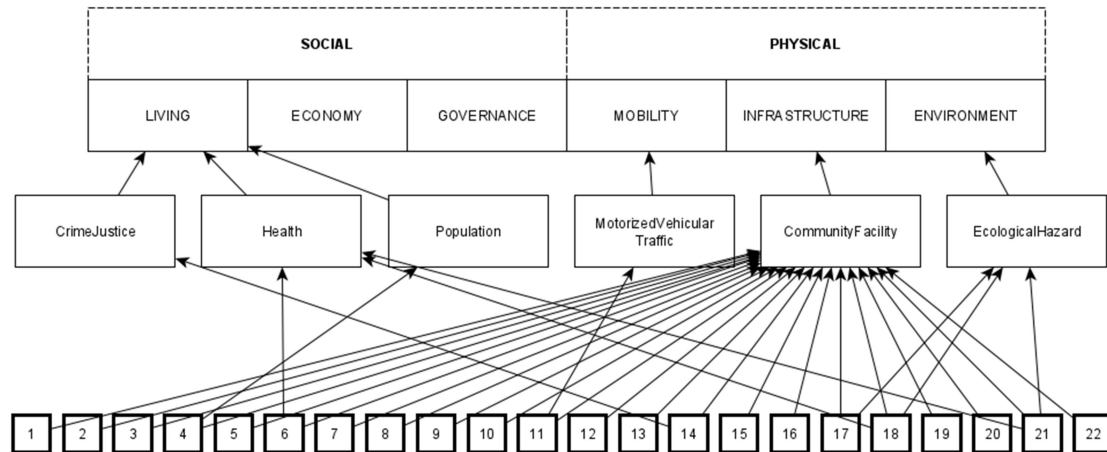


Figure 2. Modelling Beach Quality problems.

Here are some issues that deserve further attention:

- a) We categorise beaches as instances of community facilities, in line with the Community Tool Box (Fawcett et al. 2000),⁴ an Internet-based resource for practical information on community-building, providing training and technical

⁴ <http://ctb.ukans.edu>

assistance to both professionals and ordinary citizens about how to develop healthier communities.

- b) Our definition of *beach* is not restricted to “an accumulation of sand and gravel found at the landward margin of the sea or a lake” (Allaby, 2013: p. 59). Instead, we agree with Davis and FitzGerald (2004: p. 116), who stated that “a proper understanding of beaches must include the adjacent shallow marine waters, generally called the nearshore environment”. Therefore, the beach “extends from the low tide line landward across the generally unvegetated sediment to the next geomorphic feature in the landward direction, which maybe a dune (...), a cliff (...), or nowadays a seawall (...) or some other facet of human development” (idem). For this reason, problems with the seawater (e.g. SEAWATERCOLOUR, SEAWATERODOR, SEAWATERTRANSPARENCY, etc.) or taking place in the sea (e.g. MARINELITTER and MARINESPILL) are also related to the beach and thus pertain to the COMMUNITYFACILITY domain.
- c) Several beach-related problems are also linked to other domains different from COMMUNITYFACILITY. On the one hand, MARINELITTER, MARINESPILL, and SEAWATERQUALITY can be regarded as threats to the natural environment on account of the interactions of human beings (i.e. ECOLOGICALHAZARD in ENVIRONMENT), where the last two problem types can also be seen as risk factors for the individual's health (i.e. HEALTH in LIVING). Similarly, BEACHJELLYFISH is linked to HEALTH since jellyfish can cause pain or other serious health problems (e.g. difficulty breathing). On the other hand, if the number of law-enforcement authorities is insufficient, this can result in law-and-order problems, which is why BEACHSECURITY is connected with CRIMEJUSTICE in LIVING. Finally, BEACHCROWDING is linked to POPULATION in LIVING and BEACHPUBLICTRANSPORT to MOTORIZEDVEHICULARTRAFFIC in MOBILITY.

IV. CONCLUSIONS

In the last decade, social sensing has resulted in new research lines aimed at analysing and interpreting the overwhelming amount of UGC posted on social-media platforms. In this chapter, we briefly presented the ontological model of DIAPASON, a system that explores text-based UGC to detect problems in a given community. To illustrate, we described the knowledge acquisition, specification and conceptualisation stages that served to model Beach Quality problems. In this regard, the capability of DIAPASON to recognise problems from UGC can allow authorities to be informed about citizens' concerns and thus take corrective measures in the areas of interest.

Acknowledgements

Financial support for this research has been provided under grant PID2020-112827GB-I00, funded by MCIN/AEI/10.13039/501100011033, and grant number 101017861 [project SMARTLAGOON] funded by the European Union's Horizon 2020 research and innovation program.

References

Allaby, Michael, *A dictionary of geology and earth sciences*, 4th ed., Oxford, Oxford University Press, 2013.

- Appio, Francesco Paolo, Lima, Marcos, and Paroutis, Sotirios, "Understanding Smart Cities: Innovation ecosystems, technological advancements, and societal challenges", *Technological Forecasting & Social Change*, vol. 142, 2019, pp. 1-14.
- Ariza, Eduard, Jimenez, Jose A., Sarda, Rafael, Villares, Miriam, Pinto, Josep, Fraguell, Rosa, Roca, Elisabet, Marti, Carolina, Valdemoro, Herminia, Ballester, Ramon & Fluvia, Modest, "Proposal for an integral quality index for urban and urbanized beaches", *Environmental Management*, vol. 45, 2010, pp. 998-1013.
- Arthur, Rudy, Boulton, Chris A., Shotton, Humphrey, and Williams, Hywel T. P., "Social sensing of floods in the UK", *PLoS ONE*, vol. 13, 2018, No. 1. DOI: <https://doi.org/10.1371/journal.pone.0189327>
- Ayora, Veruska, Horita, Flavio, and Kamienski, Carlos, "Social networks as real-time data distribution platforms for smart cities", *Proceedings of the 10th Latin America Networking Conference*, New York, Association for Computing Machinery, 2018, pp. 2-9.
- Barbosa dos Santos, Marcelo Luis, "The 'so-called' UGC: An updated definition of user-generated content in the age of social media", *Online Information Review*, vol. 46, 2022, No. 1, pp. 95-113.
- Böhler-Baedeker, Susanne, & Durant, Tim, *SUMP glossary*, 2015. http://www.sump-challenges.eu/sites/www.sump-challenges.eu/files/3_ch4llenge_sump_glossary.pdf
- Bundesregierung, *Government report on wellbeing in Germany*, 2018. <http://www.gut-leben-in-deutschland.de/downloads/Government-Report-on-Wellbeing-in-Germany.pdf>
- Christensen, Henrik Serup, "All the same? Examining the link between three kinds of political dissatisfaction and protest", *Comparative European Politics*, vol. 14, 2016, No. 6, pp. 781-801.
- Crowe, Adam, *Disasters 2.0. The application of social media systems for modern emergency management*, Boca Raton, CRC Press, 2012.
- Davis Jr, Richard A. & FitzGerald, Duncan M., *Beaches and coasts*, Oxford, Blackwell, 2004.
- Doran, Dereka, Severin, Karlb, Gokhale, Swapnab, and Dagnino, Aldo, "Social media enabled human sensing for smart cities", *AI Communications*, vol. 29, 2016, pp. 57-75.
- Eitzen, Stanley, Baca Zinn, Maxine, and Eitzen Smith, Kelly, *Social problems*, 13th ed., Boston, Pearson, 2014.
- Eurostat, *Final report of the Expert Group on Quality of Life Indicators*, Publications Office of the European Union, 2017.
- Fawcett, Stephen B., Francisco, Vincent T., Schultz, Jerry A., Berkowitz, Bill, Wolff, Thomas J., and Nagy, Genevieve, "The Community Tool Box: A Web-based resource for building healthier communities", *Public Health Reports*, vol. 115, 2000, No. 2-3, pp. 274-278.
- Fellbaum, Christiane (ed.), *WordNet: An electronic lexical database*, Cambridge-Massachusetts, the MIT Press, 1998.
- Giffinger, Rudolf, Fertner, Christian, Kramar, Hans, Kalasek, Robert, Pichler-Milanovic, Nataša, and Meijers, Evert, *Smart cities - Ranking of European medium-sized cities*, 2007. http://www.smart-cities.eu/download/smart_cities_final_report.pdf
- Gomes, Rui (ed.), *Compass: Manual for human rights education with young people*, Strasbourg, Council of Europe, 2012.

- Goswami, Anuradha, and Kumar, Ajey, "A survey of event detection techniques in online social networks", *Social Network Analysis and Mining*, vol. 6, 2016, No. 1, pp. 1-25.
- Götz, Konrad, "Traffic mobility", in *Encyclopedia of quality of life and well-being research*, Michalos, Alex C. (ed.), Dordrecht, Springer, 2014, pp. 6705-6710.
- Govada, Sujata S., Spruijt, Widemar, and Rodgers, Timothy, "Smart city concept and framework", in *Smart economy in smart cities. Advances in 21st century human settlements*, Kumar, T. M. Vinod (ed.), Dordrecht, Springer, 2017, pp. 187-198.
- Haddow, G. D., Bullock, J. A., & Coppola, D. P., *Introduction to emergency management*, 4th ed., Elsevier, 2011.
- Harniss, Mark, "Accessibility", in *Encyclopedia of quality of life and well-being research*, Michalos, Alex C. (ed.), Dordrecht, Springer, 2014, pp. 9-10.
- Helliwell, John F., Layard, Richard, Sachs, Jeffrey D., and De Neve, Jan-Emmanuel (eds.), *World happiness report 2020*, New York, Sustainable Development Solutions Network, 2020.
- Howes, Rodney, and Robinson, Herbert, *Infrastructure for the built environment: Global procurement strategies*, London-New York, Routledge, 2005.
- Huang, Qunying, and Xiao, Yu, "Geographic situational awareness: Mining tweets for disaster preparedness, emergency response, impact, and recovery", *ISPRS International Journal of Geo-Information*, vol. 4, 2015, No. 3, pp. 1549-1568.
- Istituto Nazionale di Statistica, *Description of the BES domains and indicators selected by the Scientific Commission and launched on 22 June 2012*, 2012. <https://www.istat.it/it/files//2018/04/12-domains-scientific-commission.pdf>
- , *BES 2020: Il benessere equo e sostenibile in Italia*, 2020. https://www.istat.it/it/files//2021/03/BES_2020.pdf
- Lindell, Michael K., Prater, Carla, and Perry, Ronald W., *Introduction to emergency management*, Hoboken, Wiley, 2006.
- Litman, Todd, "Measuring transportation: Traffic, mobility and accessibility", *ITE Journal*, vol. 73, 2003, No. 10, 28-32.
- Marginean, Ioan, "Quality of Life Diagnosis (QoLD)", in *Encyclopedia of quality of life and well-being research*, Michalos, Alex C. (ed.), Dordrecht, Springer, 2014, pp. 5333-5339.
- McShane, Ian, *Bringing in the public: Community facilities and social value*, doctoral thesis, Swinburne University of Technology, 2008.
- Musto, Cataldo, Semeraro, Giovanni, Lops, Pasquale, and De Gemmis, Marco, "CrowdPulse: A framework for real-time semantic analysis of social streams", *Information Systems*, vol. 54, 2015, pp. 127-146.
- Neuman, Michael, "Infrastructure", in *Encyclopedia of the city*, Caves, Roger W. (ed.), New York, Routledge, 2005, pp. 385-389.
- OECD, *How's life? 2020: Measuring well-being*, OECD Publishing, 2020. https://www.oecd-ilibrary.org/economics/how-s-life/volume-/issue-_9870c393-en
- Office for National Statistics, *Measuring national well-being in the UK: International comparisons* 2019, 2019. <https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/articles/measuringnationalwellbeing/internationalcomparisons2019>
- Parrillo, Vincent N. (ed.), *Encyclopedia of social problems*, Los Angeles-London, Sage, 2008.
- Sakaki, Takeshi, Okazaki, Makoto, and Matsuo, Yutaka, "Tweet analysis for real-time event detection and earthquake reporting system development", *IEEE*

- Transactions on Knowledge and Data Engineering*, vol. 25, 2013, No. 4, pp. 919-931.
- Sardá, Rafael, Ariza, Eduard, Jiménez, José A., Valdemoro, Herminia, Villares, Miriam, Roca, Elisabet, Pintó, Josep, Martí, Carolina, Fraguell, Rosa, Ballester, Ramón, and Fluviá, Modest, “El índice de calidad de playas (BQI)”, in *Hacia un nuevo modelo integral de gestión de playas*, Sardá, Rafael, Pintó, Josep, and Francesc Valls, Josep (eds.), Girona, Documenta Universitaria, 2014, pp. 105-122.
- Schwab, Anna K., Eschelbach, Katherine, and Brower, David J., *Hazard mitigation and preparedness*, New Jersey, Wiley, 2007.
- Seccombe, Karen and Kornblum, William, *Social problems*, 16th ed., New York, Pearson, 2020.
- Sirgy, M. Joseph, “The psychology of material well-being”, *Applied Research Quality Life*, vol. 13, 2018, pp. 273-301.
- Stiglitz, Joseph E., Sen, Amartya, and Fitoussi, Jean-Paul, *Report by the Commission on the Measurement of Economic and Social Progress*, 2009. <https://ec.europa.eu/eurostat/documents/8131721/8131772/Stiglitz-Sen-Fitoussi-Commission-report.pdf>
- UN Committee on Economic, Social and Cultural Rights, *General comment No. 21 to Art. 15, Para. 1 (a), of the International Covenant on Economic, Social and Cultural Rights*, 2009. <https://www.refworld.org/docid/4ed35bae2.html>