



Article

AI Augmented Approach to Identify Shared Ideas from Large Format Public Consultation

Min-Hsien Weng ^{1,*} , Shaoqun Wu ²  and Mark Dyer ¹¹ School of Engineering, The University of Waikato, Hamilton 3216, New Zealand; mark.dyer@waikato.ac.nz² School of Computing and Mathematical Sciences, The University of Waikato, Hamilton 3216, New Zealand; shaoqun.wu@waikato.ac.nz

* Correspondence: min-hsien.weng@waikato.ac.nz

Abstract: Public data, contributed by citizens, stakeholders and other potentially affected parties, are becoming increasingly used to collect the shared ideas of a wider community. Having collected large quantities of text data from public consultation, the challenge is often how to interpret the dataset without resorting to lengthy time-consuming manual analysis. One approach gaining ground is the use of Natural Language Processing (NLP) technologies. Based on machine learning technology applied to analysis of human natural languages, NLP provides the opportunity to automate data analysis for large volumes of texts at a scale that would be virtually impossible to analyse manually. Using NLP toolkits, this paper presents a novel approach for identifying and visualising shared ideas from large format public consultation. The approach analyses grammatical structures of public texts to discover shared ideas from sentences comprising *subject + verb + object* and *verb + object* that express public options. In particular, the shared ideas are identified by extracting noun, verb, adjective phrases and clauses from subjects and objects, which are then categorised by urban infrastructure categories and terms. The results are visualised in a hierarchy chart and a word tree using cascade and tree views. The approach is illustrated using data collected from a public consultation exercise called “Share an Idea” undertaken in Christchurch, New Zealand, after the 2011 earthquake. The approach has the potential to upscale public participation to identify shared design values and associated qualities for a wide range of public initiatives including urban planning.

Keywords: natural language processing; data analysis; public participation; city; urban planning



Citation: Weng, M.-H.; Wu, S.; Dyer, M. AI Augmented Approach to Identify Shared Ideas from Large Format Public Consultation. *Sustainability* **2021**, *13*, 9310. <https://doi.org/10.3390/su13169310>

Academic Editor: Theodore Chadjipadelis

Received: 15 July 2021

Accepted: 10 August 2021

Published: 19 August 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Public data, contributed by citizens, stakeholders and other potentially affected parties, are becoming increasingly used to collect the shared ideas of a wider community. Online surveys with questions and answers powered by e-participation tools are being used to identify the public’s new ideas, preferences and opinions through public consultation exercises [1]. For example, “*The Quality of Life Survey*” undertaken recently in New Zealand asked more than 7000 residents for their opinions on a wide range of different aspects of urban life in four major cities [2]. Likewise, social media, such as Twitter, is another data source to observe public response to news events, such as an earthquake or presidential election, and collect public opinions [3–5]. Once ideas are collected from public data, the policy formulation process is undertaken to map the individual ideas of self-interests to the shared values of public interests, and then translate into action plans [6–8]. However, large format data from public consultation exercises are often organised and analysed manually for reporting to relevant parties or published online for public access. For example, this was the case for the “*Share an Idea*” public consultation exercise by the Christchurch City Council after the 2011 earthquake. The interpreted results were reported as “*Common Themes*” as a guide for rebuilding Christchurch Central City [9]. Not surprisingly, manual interpretation of large format data sets can be a slow and laborious process and even infeasible when large volumes of public data are involved [10].

Natural Language Processing (NLP) technologies provide the opportunity to automate a large proportion of data analysis when processing large volumes of text data at a scale that would be difficult to achieve manually, such as identifying shared ideas from online surveys. In practice, NLP comprises a set of analytical methods and tools that facilitates computational interpretation of written languages by being trained on large datasets [11]. NLP has been widely used in modern life in the area of machine translation [12,13], smart search engines accepting long-form and complicated queries [14], high-quality summary generation from newspapers, product or movie reviews and restaurant ratings analysis [15–18]. Not surprisingly, various applications involving NLP have increasingly appeared in public service domains, such as analysing Twitter messages to identify opinions about the quality of public services [19,20] or to communicate emergency alerts to citizens such as during Hurricane Sandy in the USA [21,22]. NLP-empowered analysis shines in examining the public's attitudes towards events (climate change, electricity price, political protest, etc.) and helping policy makers understand public concerns and interests for policy formation [23–25]. Similar examples can be found for healthcare applications, where NLP is applied to process health-related texts from social media and extract the medical insights, such as patient self-treatment experiences (symptoms and therapies) [26].

This paper describes a novel approach to augment large format public consultation analysis using NLP technologies in order to identify people's shared ideas about future needs of urban infrastructures. The use of NLP in this manner is still in its infancy, although there are examples where it has been used for topic modelling [27] or sentiment analysis [18], which provides a general overview of public opinions at coarse-grained level. In comparison, our aim was to achieve a greater textual understanding at fine-grained granularity of people's shared interests. In this instance, the shared interests were analysed, aggregated and interpreted using NLP to reveal preferences of soft and hard urban infrastructures [28]. This paper is part of ongoing development of the "Urban Narrative" project [29] that aims to aggregate individual contributions into common narratives to foster collaborative problem solving as advocated by Nabatchi [30].

The approach has been illustrated using a dataset obtained for Christchurch "Share an Idea" initiative following the 2011 earthquake when the City of Christchurch (NZ) sought citizens' views on priorities for the future redevelopment of the city to improve livelihood and liveability post-earthquake.

2. Literature Review

2.1. Shared Ideas in Public Consultation

"Ideas" expressed online as short written or spoken messages communicate people's thoughts and priorities. These are sometimes short-term issues or long-term concerns. These grassroots "ideas" can frequently offer valuable insights into "thorny" public problems that can yield innovative solutions grounded in underlying cultural beliefs and values systems [7]. Evidence shows that these "shared ideas" can play a crucial role in helping to shape public policymaking [30]. These "shared ideas" have at times gone on to create "world images" by determining political agendas and influencing future actions at times on a global scale [31]. Scholars [32,33] have highlighted the important role played by "shared ideas" when drawing attention to people's needs and priorities that in turn can be used by local or national governments to shape public policy. For example, "Share an Idea", conducted by Christchurch City Council, was used to collect grassroots ideas for rebuilding in the central city district following the 2010 and 2011 earthquakes. The community-wide conversation comprised a combination of online web-based forum together with face-to-face workshops with more than 100 community groups. In total, over 10,600 ideas were collected across a wide range of interests covering public space utilisation, transport choices, city life, marketplaces, earthquake memorials, etc. The shared ideas were manually analysed by the City Council and aggregated into four urban themes, namely *space*, *movement*, *market* and *life*. In turn, these broad themes were employed to produce high-level principles for rebuilding the central city district.

In recent years, similar digital platforms have been employed to upscale public participation, such as European eParticipation projects [1] that utilise Information Communication technologies to develop online surveys or e-petitions/e-voting systems for citizens to express their opinions and ideas online (such as systems named “*We the People*”, “*UK Parliament Petition*”, “*European Parliament Petition*”). Social media (Twitter, Facebook, Instagram, etc.) have been popular data sources to observe public opinions and perceptions. For example, the case study by A.P. Kirilenko and Stepchenkova [34] collected millions of Twitter messages to understand the public’s concerns on climate change. This interactivity of social media has enabled the public sector to have more frequent engagements with citizens to co-produce public services and public policy. The case studies by Criado and Villodre [19] and Mergel [20] demonstrated the strategies employing Twitter in increasing the satisfactory levels of public service delivery. Owing to the wide availability, user-friendly interactivity and fast growth of data, Twitter and other social media are widely considered by the research communities as a conducive source of public opinion collection alternative to survey-based public consultation exercises. However, the success of public participation ultimately depends on allowing shared decision-making as highlighted by Nabatchi [30].

2.2. Identifying Shared Ideas with Natural Language Processing

Data collected from social media or government sponsored e-participation platforms are being collected and analysed on a massive scale. However, the data often originates in unstructured format from several different sources (texts, web pages, pdf, Word documents, etc.) written in several different human languages (English, Chinese, Spain, etc.) often with numbers, dates, names of things, URLs, hashtags and quotes. This diverse, unstructured format poses serious challenges for analysis and interpretation using traditional analytic techniques. To address these challenges, Natural Language Processing (NLP) techniques have been increasingly used to process public data in order to gain insights into public opinions about public services [35]. The analysis has often been in the form of Topic Modelling and Sentiment Analysis, where Topic Modelling extracts a number of topics from the texts and Sentiment Analysis detects the public moods as positive or negative. Other techniques include Named Entity Recognition analytics that detects the names of public services from tweets to help improve the quality of public administration [10]. However, Topic Modelling and Sentiment Analysis are most relevant to our research.

Topic Modelling identifies the topics from a large body of texts which are then classified and organised by topic [27]. For example, Hagen et al. [36] employed the LDA (Latent Dirichlet Allocation) topic modelling technique to detect petition patterns in “*We the People*” e-petitioning system. In this case, the top 20 issues detected from public petitions concerned topics familiar to the public (e.g., freedom of religions) or related to important social events (e.g., gun control), whereas less familiar topics such as student visas and diversity of children health care issues (breastfeeding, abduction, baby cares) received less attentions from the public. Likewise, Lock and Pettit [37] employed LDA models to analyse online surveys in order to identify what factors influence customer satisfaction towards public transit systems. However, there are drawbacks to using LDA, as these mostly involve the significant effort required to fine-tune the model and the manual interpretation to select meaningful topics from automatically generated topic candidates.

In comparison, Sentiment Analysis assists the interpretation and understanding of public preferences on key issues ranging from awareness of climate change to protest movements by detecting moods of opinions as “positive”, “negative” or “neutral”. For example, Kirilenko et al. [38] demonstrated the concept of “people as sensors” by measuring the public attitudes towards climate change from Twitter messages. Likewise, Raza et al. [24] identified public sentiment towards the 2019 Azadi political protest in Pakistan. However, Sentiment Analysis is limited by the multiple interpretation of word order, slang, or the context e.g., “don’t like car”, which requires several sentiment analysis approaches (such as lexicon-based, rule-based or neural language models) inherently needed to achieve a balanced understanding of the sentiments [17,18,39].

To overcome some of these limitations, the two methods have been used in tandem to evaluate public satisfaction of large-scale infrastructure projects where the public satisfaction are measured with sentiment analysis and topic modelling is used to identify the causes of the different sentiments [25,37,40–42]. At the same time, the performance of NLP has been benefiting from ongoing advances of deep learning techniques [43].

Even so, there are still a number of challenges that researchers face when using NLP in analysing public data for opinion gathering. As reported by Kuflik et al. [44], analysing the transport-related data collected from Twitter requires a domain ontology that defines the domain specific terms (such as bus, car, traffic, road, park, etc. for transport) to identify and filter the relevant messages from a large amount of Twitter data. Building a comprehensive ontology is difficult because it requires the domain expertise to provide all the relevant terms. Twitter messages are short and informal in nature, containing hashtags, URL addresses and slang, which affect the accuracy of NLP tools. As a result, most NLP tools fail to produce a good quality of analysis results on Twitter data because they are designed and tuned for more formal written texts.

In this project, we use a different approach to identify shared ideas in public consultation compared with Topic Modelling that relies on identifying a set of topic words or Sentiment Analysis that captures public feeling towards an event or option. Instead, we use state-of-art NLP tools to analyse public texts and extract their linguistic features (part-of-speech tags, phrases, sentence structures) from which shared ideas are identified and grouped in terms of soft and hard urban infrastructures, or as shared interests compared to fixed positions. Two visualisation methods (Cascade View and Tree View) are developed to view shared ideas hierarchically or in fine details, which are different from the graphs used by Topic Modelling or Sentiment Analysis that provide a general overview or the trends of public options. The detailed methodology is described in the next section.

3. Methodology

As an alternative approach to Sentiment Analysis that offers coarse-grained perspectives on public attitudes or Topic Modelling that identifies individual topic words such as “bus”, “train” or “tram” to describe public transport and “travel”, “time” or “delay” to detect traveller complaints, our study focuses on the development of a methodology to provide a fine-grained analysis of public texts using grammatical sentence structures. Such analysis examines the subject, object and verb of the sentences in the form of *subject + verb + object* and *verb + object* within sentences, such as “*I wanted little mini electric buses for transport*” and “*I believe the river is an absolute asset to our city*”. The methodology involves six steps as shown in Figure 1. In Step 1, the public text is cleaned to make it suitable for NLP tools. In Step 2, the texts are split into individual sentences and from which linguistic features (such as part-of-speech of words, phrases, subjects and objects and their relations) are extracted. Step 3 selects sentences that follow grammatical patterns *subject + verb + object* and *verb + object* and subjects, objects and verbs are extracted to generate people’s ideas. Step 4 summarises the generated ideas by urban infrastructure category (such as Transport, Space, Building, People). Step 5 groups the ideas into shared interests and fixed positions by using the text analysis. Step 6 presents the grouped ideas in a set of visual and interactive charts and graphs to facilitate communication and exploration.

The methodology was developed using the public texts extracted from Christchurch City Council’s “*Share an Idea*” public consultation exercise that took place during six weeks following earthquakes in 2010 and 2011. The community wide conversation involved more than 1000 people with over 106,000 ideas about how to rebuild Christchurch City Centre collected from a variety of online and face-to-face public engagement activities, including community workshops, forums, “*Share an Idea*” website, surveys and social media. The results were published in a report entitled “*Christchurch Common Themes*” [9]. The report itself reproduced approximately 2795 quotes (4636 sentences and 64,981 words) as a distillation from the original database. For the purpose of the article, the Christchurch Common Themes Dataset is labelled (CCTD) throughout this paper.

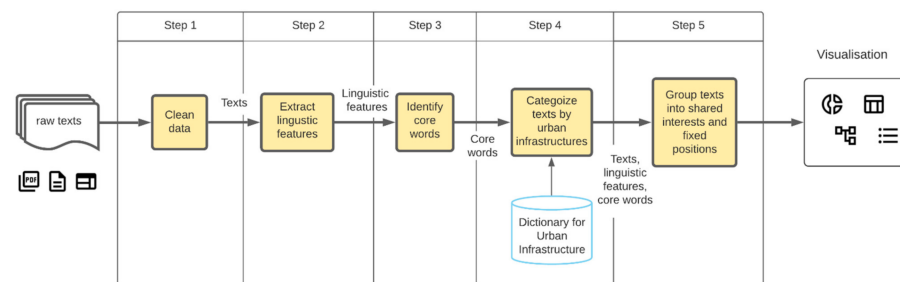


Figure 1. A workflow diagram depicts the five steps of our AI-augmented approach from raw texts to result visualisation.

3.1. Step 1: Data Cleaning

Step 1 removes the repeated texts in CCTD and cleans non-uniform formatting or use of symbols and punctuation that would otherwise reduce the accuracy of text processing. In particular, the tabs and duplicated white spaces are placed with a single space, back (') and double quotes (") with single quote marks and hyphens (-) with commas to unify the symbols.

3.2. Step 2: Linguistic Feature Extraction

In Step 2, Stanford CoreNLP toolkits [45], a deep learning-based natural language processing toolkits, are used to analyse and extract linguistic data. First, CCTD texts are split into individual sentences. The text "I want more public transport. Pedestrians take predominance." becomes two sentences: "I want more public transport." and "Pedestrians take predominance." Then sentences are tokenised with Part-Of-Speech (POS) tags assigned to individual words. Figure 2 shows the tagged version of the sentence "I want more public transport." with POS tags at the top where PRP stands for personal pronoun, VBP for verb, JJR for comparative adjective, JJ for adjective, and NN for noun. Finally, four types of linguistic data are extracted and stored as the metadata (i.e., data describes the data) of a sentence.

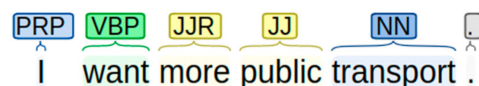


Figure 2. An example of Part-Of-Speech tagged sentence where PRP indicates a Pronoun, VBP a verb and NN a noun. (Source: Stanford CoreNLP Toolkit v4.2.0 [45]).

The linguistic metadata employed includes the following types of POS, clauses, phrases and sentences:

- **Verbs, adjectives, nouns and pronouns.** As shown in Figure 2, *want* is a verb and *transport* is a noun, which are identified based on the POS tags associated with each word. Through POS tags, *more* or *public* is an adjective that modifies the noun (*transport*), and *I* is the pronoun that refers to persons or things (*she, we, they, you, it, me, her, us, them, that, those, these, etc.*).
- **Noun, verb and adjective phrases.** Phrases are extracted using the Stanford Constituency parser [46]. A basic noun phrase comprises at least one noun word (e.g., *transport*), which often are modified by other nouns or adjectives (e.g., *more public* in Figure 2). A noun phrase itself can contain other noun phrases as shown in the following sentence where *green spaces and beautifully colourful gardens* is a noun phrase made up of two noun phrases *green spaces* and *beautifully colourful gardens*.

I want green spaces and beautifully colourful gardens.

A verb phrase constitutes a verb and its dependents that can be a noun, adjective or verb phrase. As shown in the following text, the main verb is *have*, followed by the noun phrase

free Wi-Fi and the verb phrase *to encourage tourists and travellers to linger* that in turn contains the main verb *encourage*, the noun phrase *tourists and travellers* and the verb phrase *to linger*.

Central cities should have free Wi-Fi to encourage tourists and travellers to linger.

An adjective phrase is a sequence of words with a head adjective. In the first example below, the adjective phrase is *very eco-friendly* with *eco-friendly* as the head adjective. In the second example, *totally accessible for bikes* is an adjective phrase with *accessible* as the head adjective.

Every building must be very eco-friendly.

The street layout should be totally accessible for bikes.

- **Object clause.** In English, an object clause acts as the object of a verb. In the following sentences, *that the clubs should be more classy* and *that CBD should be of high density with medium rise buildings* are the object clause of the verb *think* and *believe* respectively.

I think that the clubs should be more classy.

I believe that CBD should be of high density with medium rise buildings.

- **Sentence structure.** A complete sentence typically has a grammatical structure of *Subject + Verb + Object* as illustrated in Figure 3. The subject is usually a noun phrase or a pronoun (e.g., *I, we, you*) and the object can be a noun, verb, adjective phrase or a clause. This includes sentences that convey a command, as illustrated in Figure 2, called an imperative sentence. This type of sentence always takes the second person (*you*) for the subject but most of the time the subject remains hidden. The Stanford NLP Open Information Extractor [47–49] makes it easy to identify sentence structures.

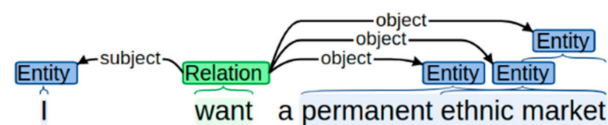


Figure 3. An example of a subject, object and relation graph where “I” is the subject and “a permanent ethnic market” the object, and “want” the relation. (Source: Stanford CoreNLP Toolkit v4.2.0 [45]).

However, not all sentences in the Christchurch Common Themes Dataset (CCTD) were complete sentences with the grammatical structures comprising *subject + verb + object* or *verb + object*. Instead, incomplete sentences were submitted by the public that included suggestions such as *quality and affordability, great central city apartment living, a safe environment, car-free shopping streets, were quite intimidating, don't want, too good to lose, close second, not so stiflingly conservative, sustainable*. These suggestions evidently represented public ideas on design values or design attributes wanted for the central city development plan for Christchurch, however, they were not considered in this study because the focus was on compiling automated methods of analysis. The unstructured suggestions would have required manual interpretation at this stage. Nevertheless, the incomplete sentences will be the focus of a later study, given their obvious importance.

3.3. Step 3: Identifying Core Words to Characterise Thematic Urban Infrastructures of Importance to the Public Using Linguistic Metadata

In Step 3, the grammatical structure of sentences from Step 2 are further examined to investigate the relationships between *subject + verb + object* or *verb + object*. In the examples given below from the CCTD dataset, the words *I* and *CBD* are subjects connected to the objects ‘*green spaces and beautifully colourful gardens*’ and ‘*of high density with medium rise buildings*’ by the verbs *want* and *be*, respectively. This allows us to begin formulating a public narrative based on ideas from public consultation. In this case, the corresponding common or proper nouns used to label different components of urban infrastructure (*spaces, gardens, CBD and buildings*) are compiled as core words to characterise thematic urban infrastructures of importance to the public.

I want green spaces and beautifully colourful gardens

CBD should be of high density with medium rise buildings

3.4. Step 4: Categorising Ideas from PUBLIC consultation in Terms of Soft and Hard Urban Infrastructures

In Step 4, a framework for thematic urban infrastructures is adopted based on the notion of soft and hard urban infrastructures in Table 1 proposed by Dyer et al. [28]. The term “soft” refers to public administrative, organisational and social structures present in a city, whereas the term “hard” describes the physical components of a city that enable the soft infrastructure to function. The framework allows core words extracted from the text analysis in Step 3 to identify the type(s) of soft and hard urban infrastructures highlighted by public consultation. This cataloguing of public “ideas” as soft or hard urban infrastructures enables the desired design attributes or design qualities to be further explored by examining relationships expressed between *subjects* and *objects* commonly using *adjectives*, e.g., *green spaces*, *colourful gardens*, *medium rise buildings*.

Table 1. Soft and hard urban infrastructures after Dyer et al. [28].

Hard Infrastructure	
Utilities	Utilities are considered to be physical services such as transportation, water and waste systems, ICT, etc. These utilities connect and operate equally across all urban scales, including national and international interconnectivity.
Urban Space	Urban spaces are considered to be largely as bounded space, in the form of streets, urban plazas or local squares, playgrounds, parks, etc. Urban space is typically identifiable at the neighbourhood or district scale, depending on the nature of the open space and pattern of land ownership.
Buildings	Buildings are considered to be architectural space defined as single or grouped buildings forming part of an urban block. This will include dwellings, educational buildings, healthcare buildings, etc.
Soft Infrastructure	
Institutional	Institutional infrastructure refers to public and private systems which provide certain services within the city such as local government, legal frameworks including land ownership, healthcare services, or educational services. It may also include sporting, art and culture, or official community support organisations. These institutions are typically top-down and more formal in nature.
Community	Community infrastructure refers to formal and informal networks, community or local business groups that occur within neighbourhoods or districts. These infrastructures rely on bridging and linking social capital. While “Communities of Interest” or online communities may not be location specific, many community organisations will relate to a specific physical community delineated by political, parish or physical boundaries (a river, large street, etc.). In this regard community infrastructures will often operate within the district scale and arguably at a more identifiable level at the neighbourhood scale.
Personal	Personal infrastructure refers to the support systems a person will have at an individual, family, or friendship level. This will often involve bonding social capital where membership of a family or social group is critical to a sense of belonging. It will also include educational attainment and other support systems that occur at an individual level.

3.5. Step 5: Categorising Ideas as Shared Interests Compared to Fixed Positions

A parallel step in the analysis is differentiation between shared interests compared to fixed positions as expressed through public consultation. As highlighted by Nabatchi [30],

successful public consultation depends on creating collaborative environments whilst avoiding adversarial ones. Collaborative environments are where participants willingly work together towards a common goal based on shared interests (often based on shared values) as opposed to fixed positions that potentially lead to conflict. The design of the public consultation exercise itself can influence how participants frame contributions. As noticed for the Christchurch “*Share an Idea*” public consultation, it produced a combination of demands that could be labelled as fixed positions compared to expressions of belief, appreciation and fondness, which would fall into the category of shared interests. With those aims in mind, the text analysis allows the verbs *want*, *like*, *love* and *believe* to be identified as early indicators of fixed positions (*want*) compared to shared interests (*like*, *love*, *believe*). Examples of the different categories of verbs and implied fixed positions or shared interests are given below from the CCTD Dataset.

“I would **like** to use some of the bricks and material from buildings lost in the quake in new buildings.”

“I would **love** to see the inner city car free with a lot of cycle ways, bus lanes and pedestrian only areas.”

“I **believe** that more green alternatives should be utilised to improve essential services.”

“I **want** recycling bins and more trees.”

“I **want** to see more rainwater reused for irrigating green spaces.”

3.6. Step 6: Visualisation of a Common Narrative Evolving from Public Consultation

Having undertaken text analysis of the large volumes of words collected from public consultation and interpreted the meaning in terms of urban infrastructures and shared interests, a final step is to communicate the results ideally in real time. Data visualisation tools were chosen to communicate a common narrative that could ultimately lead to design personas being created by sorting shared interest in relation to demographic information about contributors. At this early stage in the study, the visualisation techniques chosen offer a *Cascade View* and *Tree View*. The *Cascade View* comprises a combination of hierarchy plots to indicate overall trends towards different soft and hard infrastructures linked to frequency plots of specific infrastructure terms within these broad categories. Ultimately, the frequency plots provide access to individual sentences within a single display that reveal individual contributions for the reader to view. In comparison, the *Tree View* illustrates a pattern of dialogue around a common subject, verb or object such as *I want* as a collection of fixed positions or *composting* as an object of interest. Ultimately, the visualisation techniques are intended to create a common narrative that ideally reveals shared interests that can better inform future decision-making around a more sustainable and resilient future.

In this project, we used two programming languages, Python and JavaScript, along with a set of publicly accessible software libraries. We developed two Python programs and one JavaScript program. The first Python program cleans up the public text in Step 1 and uses the Stanford CoreNLP Toolkit to extract linguistic features from public texts and stores them with the original text together (Step 2). The second Python program is used in Step 3, 4 and 5 to identify shared ideas and group them in terms of soft and hard urban infrastructures or shared interests and fixed positions. The JavaScript program uses visualisation libraries to display the grouped shared ideas in *Cascade* and *Tree View*. Table 2 provides an overview of the libraries used, including the name, version, the purpose of use and in which step (please refer to the detailed documentation on how to use these libraries on their website). For example, Step 2 uses the Stanford CoreNLP toolkit to extract the linguistic data. In Step 6, D3.js and Plotly.js are used to produce *Cascade View* and Google chart for *Tree View*. Note that the code implementation of our project is included in the Supplementary Materials.

Table 2. A list of libraries and tools used in our implementation.

Library	Purposes	In Which Step
Stanford CoreNLP Toolkit (v4.2.0) ¹	Extract linguistic features from public texts	Linguistic feature extraction (Step 2)
D3.js (v5) ² and Plotly.js (v2.3) ³	Produce Cascade View	Visualisation (Step 6)
Google Chart (v50) ⁴	Produce Tree View	Visualisation (step 6)

¹ Stanford CoreNLP Toolkit: <https://stanfordnlp.github.io/CoreNLP> (accessed on 1 July 2021); ² D3.js (Data-Driven Documents): <https://d3js.org> (accessed on 1 July 2021); ³ Plotly.js (Plotly JavaScript Open Source Graphing Library): <https://plotly.com/javascript> (accessed on 1 July 2021); ⁴ Google Charts: <https://developers.google.com/chart> (accessed on 1 July 2021).

4. Result

4.1. Topics of Public Importance Highlighted by Text Analysis of Christchurch Common Themes Dataset

The Christchurch Common Themes Dataset (CCTD) used for the pilot study contains 3969 sentences; of which 53% (2113 sentences) had the grammatical structure *subject + verb + object* and 21% (824 sentences) comprised *verb + object*. The residual 26% (1032 sentences) were incomplete sentences and excluded from the pilot study. The remaining 2937 complete sentences were used to explore topics of interest raised by the public in the Christchurch “Share an Idea” public engagement initiative. The topics of interest were identified as *subjects* and *objects* and categorised in relation to soft and hard urban infrastructures using the dictionary in Table 3.

Table 3. A dictionary of urban infrastructures, categories and associated terms.

Infrastructure	Category	Terms
Hard	Public Space	<i>area, CBD, centre, city, courtyard, park, piazza, place, river, space, square, venue</i>
	Transport	<i>avenue, transport, bike, bus, car, cycle, lane, monorail, parking, pedestrian, rail, road, shuttle, station, street, traffic, tram, truck, vehicle, walkway</i>
	Building	<i>apartment, bar, building, cafe, garden, hospital, housing, library, mall, market, office, quarter, restaurant, retail, school, shop, store, university</i>
Soft	People and Community	<i>activity, art, child, community, disability, family, health, neighbourhood, owner, parent, people, resident, safety, shopping, sport, tourist</i>

Table 3 gives the urban infrastructure dictionary with four categories and a number of associated terms. The key infrastructure categories were selected based on four common themes that emerged from the “Christchurch Common Themes” report [9] and grouped into soft and hard urban infrastructure [28]. For each category, we collected a list of associated terms after manually reviewing the summaries and topics in the “Common Themes” report. The Public Space category describes people’s strong desires for people-centred and green spaces in the new city centre. The Building category is for the development of safe buildings and new marketplaces that encourage businesses and retailers to bring people back to the city. The Transport category aims to provide people with better transport choices (*bus, cycle, tram, etc.*) to easily commute from/to the new city centre. The People and Communities category transforms the city into a safe and vibrant city accessible for all people (*families, ethnicities, communities, etc.*).

The sentences in the CCTD dataset were clustered into these four categories based on the terms in the dictionary (Table 3). For example, the sentence “*I want social spaces*” is categorised under Public Space because of the occurrence of the term *space*. This dictionary was built on the Common Theme report, and the Christchurch City Council used their own approach to summarise the people’s ideas. Our approach is based on this dictionary

to categorise the texts in the CCTD dataset, and also identify subjects and objects in the sentences. The results are presented in Tables 4 and 5 for subjects and objects, respectively.

Table 4. Categories of *subjects* raised by the Christchurch “Share an Idea” Public Consultation, their percentages and some examples.

Category	Percentage	Examples
Public Space	7.9%	<i>A green sustainable city, industrial area, rooftop garden area, more green area, multi-ethnic area, market square, more green space, more retail space, safe place, international quality convention centre, dog park, quality venue, inner courtyard</i>
Transport	6.3%	<i>A smart integrated reliable public transport system, low energy transport, small electric bus, bus lane, bus exchange, too many cars, electric cars, free tram system, expanded tram, separate cycle paths, cycle lanes, dual walk/cycle way, communal bikes, a safer bike way, secure bicycle parking, convenient and secure cycle lockup, light rail, a tree-lined ring road</i>
Buildings	4.2%	<i>low-rise building, safe building, new architecturally designed building, heritage building, affordable apartment, more residential apartments, ethnic restaurant, metal bar, hobby/craft-related shop, suburban mall, large farmer market, mixing up social types of housing, high quality primary and secondary schools, public library</i>
People and Communities	4.8%	<i>young people, older people, less drunk people, local and tourist, land owners, supermarket owners, young parents, teenage children, stronger community</i>
Miscellaneous	17.3%	<i>industrial businesses, outdoor amphitheatre, alternate energy systems, new waste system, high density developments, the council, public drinking fountain, huge musical fountain, pure water, all rain water, underground power lines, wind power, the use of specific precincts, better lighting</i>
Pronouns	29.3%	The pronoun <i>you</i> is treated as the subject in <i>verb + object</i> sentences
	30.2%	<i>I, we, it, this, they, that, there, one, these, those, etc.</i>

Analysis of topics expressed as a subject in Table 4 showed that 7.9% of sentences were concerned with Public Space, compared with 6.3% for Utilities (in the form of Transport), 4.2% for Building and 4.8% for People. A remaining 17.3% (such as *industrial businesses, amphitheatre, alternative energy system, the council*) fell outside of the dictionary of terms compiled for soft and hard urban infrastructures. By examining the results in terms of infrastructure, it was found that the subject phrases mostly relate to specific urban spaces (*city centre, market square, green space, dog park, inner courtyard, etc.*), private dwelling/public buildings (*safe building, affordable apartment, restaurant, school, public library, etc.*), transport services (*public transport system, bus exchange, free tram system, cycle lane, light rail, etc.*) or a social group of people (*older people, family, young parents, community, etc.*). In comparison, subjects in the miscellaneous category covered topics such as *wind power, water, waste system, lighting*.

Lastly, the study showed that 59.5% of subjects were pronouns rather than common or proper nouns. The pronoun *you* was the most frequently cited in *verb + object* sentences (29.3%), followed by the first person singular *I* (15.6%), *we* (5.2%) and other pronouns (9.4%) such as *it, this, they, that, there, one, these, those*.

Table 5. Categories of *objects* raised by Christchurch “Share an Idea” Public Consultation, their percentages and some examples.

Category	Percentage	Examples
Public Space	31.4%	<i>no cars inner city, wheelchair-friendly city, family-friendly city, large open spaces for community and cultural activity, public space for all ages to enjoy, retail areas, more areas for children to play in the CBD, central dog park, a place for people to gather and relax, to preserve the Maori and English inheritance of our city</i>
Transport	12.0%	<i>allow for more disabled/short term car parking, less cars and more emphasis on walking and cycling and public transport, electric buses, integrated bus light rail tram and bicycle transport, to see more dedicated cycleways and safer intersections for cyclists, to slow traffic to make it safe for everyone, to separate bike lanes with bike and pedestrian priority</i>
Buildings	9.2%	<i>all new buildings to have rooftop gardens, a variety of more modern buildings, 5 storey building height limit, to build safe tall buildings in an earthquake prone location, a certificate on every commercial building that says what level of earthquake code this building is now at, more buildings converted to apartments, most of the old heritage buildings to be rebuilt, to use some of the bricks and material from buildings lost in the quake in new buildings</i>
People and Communities	5.7%	<i>people safety, public safety and peace of mind, lots of fun activities for families like ice skating in winter and big playgrounds, to become a global community, a new deep sense of community, able to include all people regardless of their physical abilities</i>
Miscellaneous	34.6%	<i>publicly owned business incubators with low rent and leases limited to three years, more drinking fountains, lots of free buskers festival events/firework shows/outdoor cinema, more recognition of Maori culture, more foreign investment, services such as banks/government departments/clusters of dentists and lawyers, low rise ecologically sound structures</i>
Pronouns	7.1%	<i>it, them, us, you, me, there, one, that, this, those, these</i>

In comparison, the objects identified in complete sentences were expressed as either a noun, verb, adjective, adverb phrases, object clauses or pronoun. The results are presented in Table 5 where objects are grouped in relation to different categories of soft and hard urban infrastructures. The analysis revealed that 31.4% of objects were concerned with Public Space, followed by Transport (12.0%), Buildings (9.2%) and People and Communities (5.7%), leaving a miscellaneous category (34.6%). The results indicate the public’s desire for (1) a people-friendly space where families and communities can gather and hold the cultural activities to celebrate the heritage of Maori and English culture in New Zealand, (2) more transport options to travel to around the city centre, such as biking, light rail, tram, etc., (3) stronger and safer buildings that follow the low-rise and earthquake-resilient building standard, and (4) public safety, fun family activities and inclusive community. The miscellaneous category covered topics ranging from business incubators, cultural festivals, building structures and cultural recognition.

Lastly, a variety of verbs were used in the *subject + verb + object* and *verb + object* sentences. The *be* verb in different forms (*be, is, are, was, were, been, etc.*) was the most common verb, used in about 15% of sentences. Other frequently appearing verbs (occurring more than 20 times) included *want, like, love, believe, have, need, make, use, encourage, like, keep, create, get, do, bring, think, walk, move, live, love, take, give, go*. The verbs *want, like, love, believe* offered some indication of shared interests compared with fixed positions as explained

earlier. The influences of these verbs are best illustrated using the *Tree Views* presented in the following section.

4.2. Visualisation of Topics Highlighted by Text Analysis of Christchurch Common Themes Dataset (CCTD)

Having summarised the main topics raised by text analysis in Tables 4 and 5, the next step is to visualise the results in a more direct and easily digested manner for public engagement. The aim has ultimately been to create a data story that can be updated in real time to promote community conversation online. With this aim in mind, illustrations based on Cascade View and Tree View have been developed as follows.

4.2.1. Cascade View

Cascade View (Figure 4a,b) has been developed to depict a cascading stream of information beginning with a hierarchy chart (Figure 4a) to illustrate the broad categories of interests captured by public consultation. For the Christchurch “Share an Idea” case study, the hierarchy chart focuses on different categories of soft and hard urban infrastructures, encompassing utilities (*transport, energy, water, telecommunications*), public spaces (*garden parks, rivers, plazas, playgrounds*) and buildings (*housing, schools, hospitals, libraries, etc.*), people and communities. Having depicted relative interests, the next stage illustrates the relative frequency of responses within different categories of soft and hard urban infrastructures. For example, urban utilities encompassed different modes of urban transportation. The third and fourth level of visualisation provides fine-grained information about individual terms visualised in a word cloud and individual sentences. At each stage, the screen provides an interactive link to the access information from the next stage, i.e., the word cloud provides access to individual sentences containing that word.

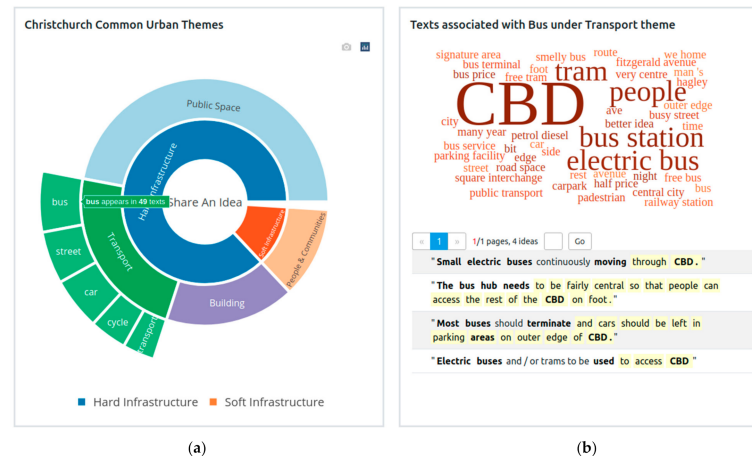


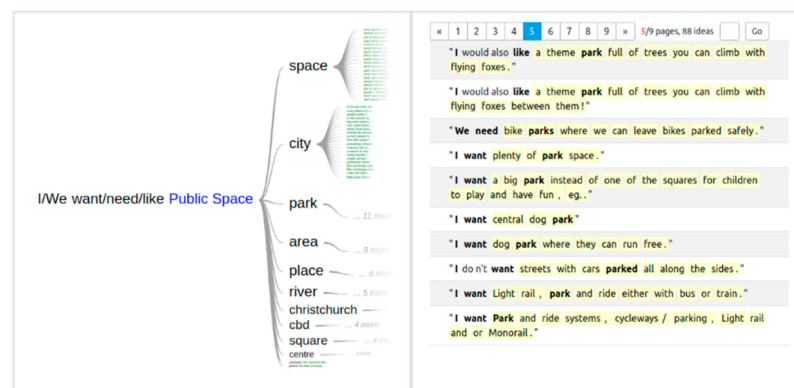
Figure 4. Cascade View of Christchurch Common Themes Data “Share an Idea”. (a) The hierarchy chart that illustrates the broad categories of interests captured by public consultation. (b) The texts associated with “Bus” under the Transport theme visualised in a word cloud and individual sentences.

The generation of Cascade View is based on information extracted by NLP text analysis of Christchurch Common Themes Dataset (CCTD). The illustration reveals public interest in transportation with a particular interest shown for improved public transportation with more buses and fewer cars in the city centre. The information was extracted from the noun phrases of the *subject + verb + object* and *verb + object* sentences. The size of each term under a category reflects the number of times a term was mentioned. In this case, the term *bus* is the most frequently mentioned term under Transport (49 times), followed by *street, car, cycle, transport, etc.* By “clicking” on a term *bus* (shown in green), a word cloud (Figure 4b) is displayed corresponding to *objects* expressed as a bus in the sentences. The relative font sizes for the word cloud reflect citation rates. Further “clicking” on individual words in the

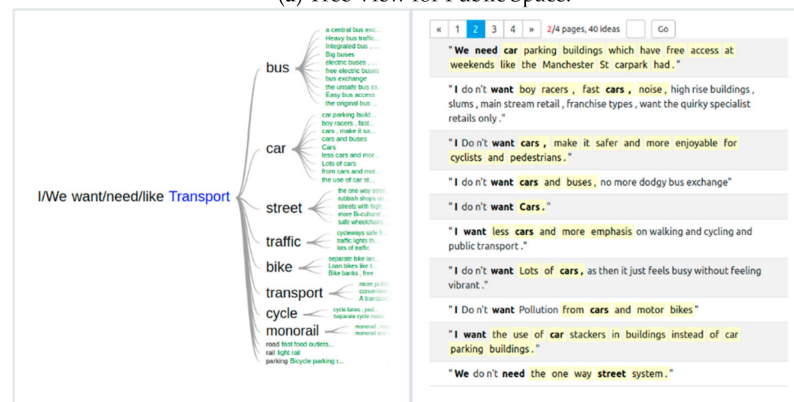
word cloud (such as CBD) allows original sentences from the CCTD dataset to be displayed such as “use of electric buses in the CBD”.

4.2.2. Tree View

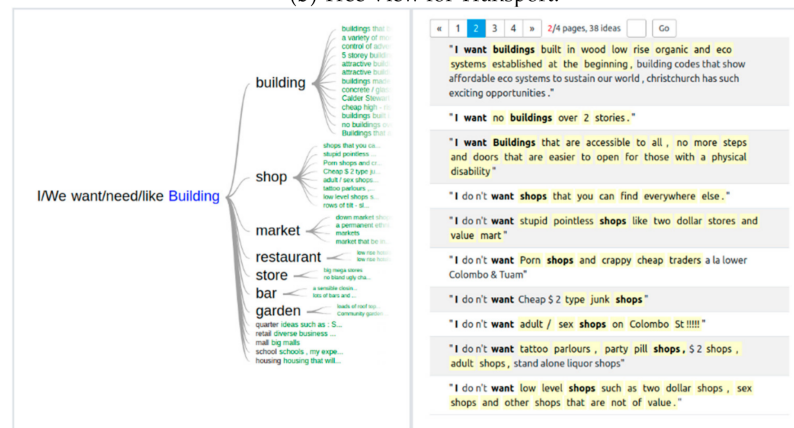
To complement *Cascade View*, a parallel mode of visualisation was developed using *Tree Views*. Tree Views (Figure 5a–c) provide an opportunity to cluster public suggestions together into urban themes using word trees. The word trees can be constructed based on grammatical structures (i.e., *noun phrase*, *verb phrase*, *object clauses*) or verbs indicating fixed positions or shared interests (i.e., *want*, *need*, *like*, *love*, *think*, *believe*). In each case, the sentences are ordered under themes for soft and hard infrastructures. For example, three Tree Views are presented in Figure 5 plotted from a total of 327 sentences (8.2% of the total sentences).



(a) Tree View for Public Space.



(b) Tree View for Transport.



(c) Tree View for Building.

Figure 5. Tree Views cluster public suggestions together into urban themes using word trees.

These Tree Views catalogue specific public suggestions about improvements for Public Space (Figure 5a), Transportation (Figure 5b) and Buildings (Figure 5c) by collating individual recommendations into a network of word trees. There is also an option to construct the word trees based on grammatical structures that highlight objects as *noun phrases* or recommended actions as *verb phrases* or combinations of highlighted objects with associated actions as *object clauses*. For the purpose of this article, the Tree View in Figure 5 has been created based solely on the urban themes of public space, transport and buildings. As such, it offers a clear insight into public's preferences and dislikes in relation to transport (e.g., *wanting less cars in central city*), public space (e.g., *green parks for cycling and dog walking plus park and ride facilities for multi-mode transport*) and buildings (e.g., *attractive timber buildings not glass skyscrapers*).

5. Discussion

5.1. Novel Approach for the Analysis, Interpretation and Visualisation of Large Format Dataset from Public Consultation

This paper presents an innovative approach for the analysis, interpretation and visualisation of large format dataset from public consultation. The objective is to facilitate large-scale, community-wide conversations about shared values and priorities. At its core, the approach uses NLP tools to augment sentence-based dataset into usable phrases that express public insight and ideas visualised through Cascade and Tree Views. The terms Cascade and Tree View refer to a combination of hierarchical visualisation methods that have been used to gradually reveal ever greater details of data as a part of a data storytelling approach. More information about this data visualisation is given below.

This novel technique has been trialled using CCTD dataset from the Christchurch "Share an Idea" public consultation exercise undertaken in 2011 following a series of highly destructive earthquakes [9]. In comparison with the manual analysis undertaken by Christchurch City Council, the NLP-driven analysis provides a semi-automated process to readily analyse the 106k public contributions in a matter of minutes. Perhaps more importantly than a rapid analysis of large single datasets, the novel approach provides an opportunity to analyse and visualise public contributions in real time. This capability enables deliberative public discussion to take place in the here and now, as advocated by Nabatchi [30] in her seminal paper on designing public consultation. The aim being to use a deliberative model of communication to support structured problem solving based on shared knowledge and insights from both the user (public) and provider (local authority or private service provider). In comparison, one-way communication is often used for information-sharing purposes, which provides little opportunity to discuss public values.

In a similar vein, the semi-automated NLP-driven process offers an opportunity to explore public "interests" compared with "positions". As highlighted again by Nabatchi [30] and supported by Fung [50,51], interest-based processes are more likely than position-based processes to generate a level of cooperation needed to help public administrators identify and understand public values and priorities in contentious situations. In this instance, NLP analysis of the CCTD dataset allowed combinations of pronouns and verbs to provide an approximate indication of interest-based suggestions compared with position-based ones, such that the pronoun and verb *I want* would tend to indicate a public position compared with the combination of pronoun and verbs *I believe/think/love*, etc., being more inclined to indicate a public interest. The Tree View visualisation tool enabled these combinations of pronouns and verbs to be grouped together and classified in terms of urban infrastructure, such as transportation.

The term semi-automated process is used because the analysis of augmented sentence data depends on use of a predefined dictionary. In the Christchurch "Share an Idea" case study, the dictionary characterised a framework for soft and hard urban infrastructure that classified elements of cities into six categories, namely institutions, community and personal for soft urban infrastructure compared to public space, utilities and buildings for hard urban infrastructure. For the Christchurch case study, the approach enables the dataset to be readily interpreted into these key elements of urban infrastructure. A similar

approach has been used to interpret the data using a dictionary for the Circular Economy CE [52], where greater emphasis was placed on recycling, reuse and reduction of waste.

5.2. Technical Details of the NLP Driven Process

Returning to the technical details of the NLP driven process, the case study restricted itself to 74% of the original Christchurch dataset. This was because the study focused only on analysing complete sentences comprising *subject + verb + object* and *verb + object*. The remaining 26% involved incomplete sentences, as shown below. However, the incomplete sentences did convey public ideas, but NLP tools failed to identify grammatical components due to misuse of punctuations and incorrect grammar. It was possible to define a set of grammatical rules to capture some irregular structures, but these rules were handcrafted after manual scanning of raw texts, which becomes infeasible when dealing with large formats.

*Model on treatment of Seine: Art, cafes, gardens all the way: The Avon Walk!
for a great Central City? People-bringing community together.*

Dining experiences, especially evening on/by the Avon River, Oxford Terrace, The Arts Centre, culture.

Similarly, there were a number of short passages that were hard to understand even by manual analysis, as shown below.

small park like areas.

Education centre, displays.

Ballantynes plus a 3?

Leaving aside difficulties dealing with incomplete sentences, the analysis of negative sentiments also proved challenging. For example, negative sentiments are expressed in the following sentences without using the word “not”. Yet, the word “not” was commonly used to detect negative sentiments. To avoid overlooking these negative sentiments, a more subtle set of adjectives and adverbs are needed to capture these indirect views (such as detecting words *less, avoid, limited, deter*).

I want less souvenir shops and more high streets and one off businesses to show NZ and locally made products to new Zealanders and tourists.

Avoid visual pollution created by cheap loud signage.

Parking is very limited in town and traffic congestion makes roads difficult and dangerous to cross when shopping.

It deters footpath life, such as cafe’s spilling out onto footpaths.

5.3. Technical Guidelines for Large Format Public Data Collection

The performance of our fine-grain analysis approach lies in the accuracy of the Stanford CoreNLP toolkits we used in the project. The toolkits have been widely recognised as one of the state-of-art natural language processing tools that can robustly and efficiently extract linguistic data from large volumes of free-formatted texts with a low error rate, above 80% of accuracy reported [46,48,49]. The accuracy of the toolkits decreases when the text input contains incomplete sentences (discussed in Section 5.2) or sentences using irregular punctuations or grammar. To achieve a satisfying result and to utilise the full potential of NLP tools, we recommend adopting certain guidelines for large format public data collection as follows, which is similar to the template proposed by Rubin et al. [53], offering additional guidance on closed and open questions.

1. Encourage complete sentence contributions by providing a default set of pronouns and verbs (*I/we . . . etc.*).
2. Differentiate between interest-based and position-based suggestions using a default set of verbs (*We must have . . . /We would like . . . etc.*).

3. Link contributions where possible to underlying cultural values through drop down menu to understand better public motives and priorities (*sense of community, sense of connections, sense of identity, etc.*).
4. Encourage ranking of cultural values and translation into associated qualities when implementing those cultural values.
5. For topic-specific public engagement, structure discussion around existing frameworks such as soft and hard urban infrastructures for urban planning and design [28] or 9R Framework for Circular Economy [54].

6. Conclusions

The pilot study has demonstrated both the potential and limitations of using NLP tools to analyse large format public data from public consultation exercises. At the policy level, it provides a novel digital platform to enable greater public engagement at grassroots to bridge the perennial gap between top-down and bottom-up approaches. This challenges the status quo where local and national governments prefer the top-down approaches. However, at a time of significant social and political changes, these traditional approaches of policy-making are being challenged. Interestingly, the EU sustainable goal advocates for a greater social inclusion, which is in line with the motivation for digital platforms. However, significant technical and political limitations still exist for sharing the power of political decision-making.

One notable technical limitation was the dependency on availability of complete sentences for the text analysis to be carried out. In the study, this limitation resulted in 26% of data being discarded. To overcome this constraint, guidelines have been proposed to format the collection of text data by encouraging sentence structures and attempting to differentiate between shared interests compared with fixed positions, ideally encouraging the format to facilitate collaboration.

It should also be noted that interpretation of NLP analysis required a thematic dictionary which in this pilot study was based on terms used to characterise soft and hard urban infrastructure. Hence, the novel method is a semi-automated technique reliant on compiling thematic dictionaries by theme experts. Even so, the technique eases the burden of manual interpretation by experts of large format data and offers the prospect of greater transparency and instantaneous two-mode of communication with the public.

On the positive side, NLP has been shown to provide a rapid means of objectively analysing relatively large amounts of text data which, when coupled with a thematic dictionary, allows an insight into public ideas and insights for transformational change. The choice of visualisations is still under development with numerous visualisation tools available to illustrate hierarchies, frequencies and networks from using sunburst plots, chord charts to word clouds and word trees. All of these methods provide a means of communicating complex data as a data story for public engagement to enable a purposeful, deliberative community discussion based on shared interests rather than fixed positions. However, its uptake depends on political will to share power and decision-making between traditional top-down methods and grassroots engagement.

Supplementary Materials: The following are available online at https://github.com/samminweng/urban_narratives. Computer code and software was released and deposited in the public GitHub repository and uploaded as the Supplementary Materials of this paper.

Author Contributions: Conceptualisation, M.D., S.W. and M.-H.W.; methodology, S.W. and M.-H.W.; software, M.-H.W.; validation, M.D., S.W. and M.-H.W.; formal analysis, M.-H.W.; investigation, M.-H.W.; resources, M.-H.W.; data curation, M.-H.W.; writing—original draft preparation, S.W. and M.-H.W.; writing—review and editing, M.D. and M.-H.W.; visualisation, M.D. and M.-H.W.; supervision, S.W. and M.D.; project administration, M.D. and S.W.; funding acquisition, M.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available in a publicly accessible repository. The data presented in this study are openly available in FigShare at (10.6084/m9.figshare.14977464), reference number (Weng, Min-Hsien (2021): Urban Narratives. figshare. Journal contribution. <https://doi.org/10.6084/m9.figshare.14977464.v1>, accessed on 14 July 2021).

Acknowledgments: The authors acknowledge CHCH City Council access to the “Common Themes” report and “Share an Idea”.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tambouris, E. eParticipation in Europe: Current state and practical recommendations. In *E-Government Success around the World: Cases, Empirical Studies, and Practical Recommendations*; IGI Global: Hershey, PA, USA, 2013; pp. 341–357.
2. Nielsen. *Quality of Life Survey 2018: Topline Report, A Report Prepared on Behalf of Auckland Council, Wellington City Council, Christchurch City Council, and Dunedin City Council*; Nielsen: Wellington, New Zealand, 2018.
3. Enli, G. Twitter as arena for the authentic outsider: Exploring the social media campaigns of Trump and Clinton in the 2016 US presidential election. *Eur. J. Commun.* **2017**, *32*, 50–61. [[CrossRef](#)]
4. Sakaki, T.; Okazaki, M.; Matsuo, Y. Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors. In *Proceedings of the 19th International Conference on World Wide Web, New York, NY, USA, 26–30 April 2010*; pp. 851–860. [[CrossRef](#)]
5. Wang, H.; Can, D.; Kazemzadeh, A.; Bar, F.; Narayanan, S. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 System Demonstrations, Jeju Island, Korea, 8–14 July 2012*; pp. 115–120.
6. Dalton, T.; Draper, M.; Weeks, W.; Wiseman, J. *Making Social Policy in Australia: An Introduction*; Taylor and Francis: Oxford, UK, 2020; p. 243. [[CrossRef](#)]
7. Fischer, F. *Reframing Public Policy: Discursive Politics and Deliberative Practices*; Oxford University Press: Oxford, UK, 2003.
8. Sam, M.P. What’s the big idea? Reading the rhetoric of a national sport policy process. *Sociol. Sport J.* **2003**, *20*, 189–213. [[CrossRef](#)]
9. Christchurch City Council. *Common Themes from Public Ideas Christchurch Central City Plan*; Christchurch City Council: Christchurch, New Zealand, 2011.
10. Porwol, L.; Hassan, I.; Ojo, A.; Breslin, J. A knowledge extraction and management component to support spontaneous participation. In *Proceedings of the International Conference on Electronic Participation, Thessaloniki, Greece, 30 August–2 September 2015*; pp. 68–80.
11. Eisenstein, J. *Introduction to Natural Language Processing*; MIT Press: Cambridge, MA, USA, 2019.
12. Awadalla, H.H. Achieving Human Parity on Automatic Chinese to English News Translation. 2018. Available online: <https://www.microsoft.com/en-us/research/publication/achieving-human-parity-on-automatic-chinese-to-english-news-translation/> (accessed on 1 July 2021).
13. Wu, Y. Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. *arXiv* **2016**, arXiv:1609.08144.
14. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2019**, arXiv:1810.04805. Available online: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85083815650&partnerID=40&md5=4986c6d6076c0c91df84d17216b47216> (accessed on 1 July 2021).
15. Fang, X.; Zhan, J. Sentiment analysis using product review data. *J. Big Data* **2015**, *2*. [[CrossRef](#)]
16. Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I. Language models are unsupervised multitask learners. *OpenAI Blog* **2019**, *1*, 9.
17. Snyder, B.; Barzilay, R. Multiple aspect ranking using the good grief algorithm. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics, Proceedings of the Main Conference, Rochester, NY, USA, 22–27 April 2007*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2007; pp. 300–307.
18. Socher, R. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, WA, USA, 18–21 October 2013*; pp. 1631–1642. Available online: <https://www.aclweb.org/anthology/D13-1170> (accessed on 1 July 2021).
19. Criado, J.I.; Villodre, J. Delivering public services through social media in European local governments. An interpretative framework using semantic algorithms. *Local Gov. Stud.* **2020**, *47*, 253–275. [[CrossRef](#)]
20. Mergel, I. A framework for interpreting social media interactions in the public sector. *Gov. Inf. Q.* **2013**, *30*, 327–334. [[CrossRef](#)]
21. Chatfield, A.T.; Reddick, C.G. All hands on deck to tweet #sandy: Networked governance of citizen coproduction in turbulent times. *Gov. Inf. Q.* **2018**, *35*, 259–272. [[CrossRef](#)]
22. Panagiotopoulos, P.; Barnett, J.; Bigdeli, A.Z.; Sams, S. Social media in emergency management: Twitter as a tool for communicating risks to the public. *Technol. Forecast. Soc. Chang.* **2016**, *111*, 86–96. [[CrossRef](#)]
23. Kavanaugh, A.; Song, Z.; Li, L.; Fox, E. Communication Behavior in an Emerging Democracy. In *Proceedings of the 20th Annual International Conference on Digital Government Research, New York, NY, USA, 18–20 June 2019*; pp. 445–455. [[CrossRef](#)]

24. Raza, A.A.; Habib, A.; Ashraf, J.; Javed, M. Semantic orientation based decision making framework for big data analysis of sporadic news events. *J. Grid Comput.* **2019**, *17*, 367–383. [[CrossRef](#)]
25. Sun, Y. Residents' sentiments towards electricity price policy: Evidence from text mining in social media. *Resour. Conserv. Recycl.* **2020**, *160*, 104903. [[CrossRef](#)]
26. Okon, E.; Rachakonda, V.; Hong, H.J.; Callison-Burch, C.; Lipoff, J.B. Natural language processing of Reddit data to evaluate dermatology patient experiences and therapeutics. *J. Am. Acad. Dermatol.* **2020**, *83*, 803–808. [[CrossRef](#)] [[PubMed](#)]
27. Blei, D.M. Probabilistic Topic Models. *Commun. ACM* **2012**, *55*, 77–84. [[CrossRef](#)]
28. Dyer, M. Framework for soft and hard city infrastructures. *Proc. Inst. Civ. Eng.* **2019**, *172*, 219–227. [[CrossRef](#)]
29. Dyer, M.; Weng, M.-H.; Wu, S.; Ferrari, T.; Dyer, R. Urban Narrative: Computational Linguistic Interpretation of Large Format Public Participation for Urban Infrastructure. *Urban Plan.* **2020**, *5*, 20–32. [[CrossRef](#)]
30. Nabatchi, T. *A Manager's Guide to Evaluating Citizen Participation*; IBM Center for the Business of Government: Washington, DC, USA, 2012.
31. Weber, M. The Social Psychology of the World Religions. In *From Max Weber: Essays in Sociology*; Oxford University Press: New York, NY, USA, 1946.
32. Schmidt, V.A. Discursive institutionalism: The explanatory power of ideas and discourse. *Annu. Rev. Political Sci.* **2008**, *11*, 303–326. [[CrossRef](#)]
33. Schmidt, V.A. Taking ideas and discourse seriously: Explaining change through discursive institutionalism as the fourth 'new institutionalism'. *Eur. Political Sci. Rev.* **2010**, *2*, 1–25. [[CrossRef](#)]
34. Kirilenko, A.P.; Stepchenkova, S.O. Public microblogging on climate change: One year of Twitter worldwide. *Glob. Environ. Chang.* **2014**, *26*, 171–182. [[CrossRef](#)]
35. Sahu, G.P.; Dwivedi, Y.K.; Weerakkody, V. Application of natural language processing (NLP) techniques in e-governance. In *E-Government Development and Diffusion: Inhibitors and Facilitators of Digital Democracy*; IGI Global: Hershey, PA, USA, 2009; pp. 122–132. [[CrossRef](#)]
36. Hagen, L.; Harrison, T.M.; Uzuner, Ö.; May, W.; Fake, T.; Katragadda, S. E-petition popularity: Do linguistic and semantic factors matter? *Gov. Inf. Q.* **2016**, *33*, 783–795. [[CrossRef](#)]
37. Lock, O.; Pettit, C. Social media as passive geo-participation in transportation planning—how effective are topic modeling & sentiment analysis in comparison with citizen surveys? *Geo-Spat. Inf. Sci.* **2020**, *23*, 275–292. [[CrossRef](#)]
38. Kirilenko, A.P.; Molodtsova, T.; Stepchenkova, S.O. People as sensors: Mass media and local temperature influence climate change discussion on Twitter. *Glob. Environ. Chang.* **2015**, *30*, 92–100. [[CrossRef](#)]
39. Gilbert, C.; Hutto, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media (ICWSM-14), Palo Alto, CA, USA, 1–4 June 2014; Volume 81, p. 82. Available online: <http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf> (accessed on 1 July 2021).
40. Jiang, H.; Qiang, M.; Lin, P. Assessment of online public opinions on large infrastructure projects: A case study of the Three Gorges Project in China. *Environ. Impact Assess. Rev.* **2016**, *61*, 38–51. [[CrossRef](#)]
41. Cao, J. Web-Based Traffic Sentiment Analysis: Methods and Applications. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 844–853. [[CrossRef](#)]
42. Méndez, J.T.; Lobel, H.; Parra, D.; Herrera, J.C. Using Twitter to Infer User Satisfaction With Public Transport: The Case of Santiago, Chile. *IEEE Access* **2019**, *7*, 60255–60263. [[CrossRef](#)]
43. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
44. Kuflik, T.; Minkov, E.; Nocera, S.; Grant-Muller, S.; Gal-Tzur, A.; Shoor, I. Automating a framework to extract and analyse transport related social media content: The potential and the challenges. *Transp. Res. Part C Emerg. Technol.* **2017**, *77*, 275–291. [[CrossRef](#)]
45. Manning, C.D.; Surdeanu, M.; Bauer, J.; Finkel, J.; Bethard, S.J.; McClosky, D. The Stanford CoreNLP Natural Language Processing Toolkit. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Baltimore, MD, USA, 23–25 June 2014; pp. 55–60. Available online: <http://www.aclweb.org/anthology/P/P14/P14-5010> (accessed on 1 July 2021).
46. Socher, R.; Bauer, J.; Manning, C.D.; Ng, A.Y. Parsing with Compositional Vector Grammars. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria, 4–9 August 2013; Volume 1, pp. 455–465. Available online: <https://www.aclweb.org/anthology/P13-1045> (accessed on 1 July 2021).
47. Angeli, G.; Premkumar, M.J.J.; Manning, C.D. Leveraging Linguistic Structure For Open Domain Information Extraction. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China, 26–31 July 2015; Volume 1, pp. 344–354. [[CrossRef](#)]
48. Chen, D.; Manning, C. A Fast and Accurate Dependency Parser using Neural Networks. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 740–750. [[CrossRef](#)]
49. Schuster, S.; Manning, C.D. Enhanced English Universal Dependencies: An Improved Representation for Natural Language Understanding Tasks. In Proceedings of the Tenth International Conference on Language Resources and Evaluation, Portorož, Slovenia, 23–28 May 2016.
50. Fung, A. Survey article: Recipes for public spheres: Eight institutional design choices and their consequences. *J. Political Philos.* **2003**, *11*, 338–367. [[CrossRef](#)]

51. Fung, A. Varieties of Participation in Complex Governance. *Public Adm. Rev.* **2006**, *66*, 66–75. [[CrossRef](#)]
52. Dyer, M.; Wu, S.; Weng, M.-H. Convergence of Public Participation, Participatory Design and NLP to Co-Develop Circular Economy. *Circ. Econ. Sustain.* **2021**. [[CrossRef](#)]
53. Rubin, G.J.; Potts, H.W.W.; Michie, S. The impact of communications about swine flu (influenza A H1N1v) on public responses to the outbreak: Results from 36 national telephone surveys in the UK. *Health Technol. Assess.* **2010**, *14*, 183–266. [[CrossRef](#)] [[PubMed](#)]
54. Potting, J.; Hekkert, M.; Worrell, E.; Hanemaaijer, A. *Circular Economy: Measuring Innovation in the Product Chain*; PBL Publishers: The Hague, The Netherlands, 2017.