Mapping Ordinances and Tweets using Smart City Characteristics to Aid Opinion Mining

Manish Puri
Department of Computer Science
Montclair State University
Montclair, NJ, USA
purim1@montclair.edu

Xu Du
Environmental Management PhD Program
Montclair State University
Montclair, NJ, USA
dux3@montclair.edu

Aparna S. Varde
Department of Computer Science /
Environmental Management PhD Program
Montclair State University
Montclair, NJ, USA
vardea@montclair.edu

Gerard de Melo
Department of Computer Science
Rutgers University
Piscataway, NJ, USA
gdm@demelo.org

ABSTRACT

This research focuses on mining ordinances (local laws) and public reactions to them expressed on social media. We place particular emphasis on ordinances and tweets relating to Smart City Characteristics (SCCs), since an important aim of our work is to assess how well a given region heads towards a Smart City. We rely on SCCs as a nexus between a seemingly infinite number of ordinances and tweets to be able to map them, and also to facilitate SCC-based opinion mining later for providing feedback to urban agencies based on public reactions. Common sense knowledge is harnessed in our approach to reflect human judgment in mapping. This paper presents our research in ordinance and tweet mapping with SCCs, including the proposed mapping approach, our initial experiments, related discussion, and future work emerging therein. To the best of our knowledge, ours is among the first works to conduct mining on ordinances and tweets for Smart Cities. This work has a broader impact with a vision to enhance Smart City growth.

CCS CONCEPTS

• **Information systems** → **Data mining**; *Content analysis and feature selection*; *Clustering and classification*;

KEYWORDS

Social media, Enterprise Intelligence, Knowledge bases, Local laws, NLP, Sentiment analysis, Text mining

ACM Reference Format:
Manish Puri, Xu Du, Aparna S. Varde, and Gerard de Melo. 2018. Mapping Ordinances and Tweets using Smart City Characteristics to Aid Opinion Mining. In *WWW '18 Companion: The 2018 Web Conference Companion, April 23–27, 2018, Lyon, France.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3184558.3191632

ACM ISBN 978-1-4503-5640-4/18/04.
https://doi.org/10.1145/3184558.3191632

1 INTRODUCTION

This research addresses the task of mining urban policy. Our vision is to analyze ordinances or local laws from websites with respect to the public reaction to them expressed on social media. This enables tangential surveys to assess opinions of residents, reflecting their satisfaction and views on urban policies. An important focus in our work is to determine to what extent such ordinances contribute to establishing the relevant urban region as a Smart City. Hence, we aim to categorize these ordinances based on their pertinent Smart City Characteristics (SCCs), of which a small snapshot with highlights is shown in Figure 1 (image source [19]). Public opinion is gathered from Twitter, given its role as a micro-blogging site with over 330 million active users. The specific objective of the present research is to relate the ordinances to the respective tweets on Twitter that express the public reaction to them.

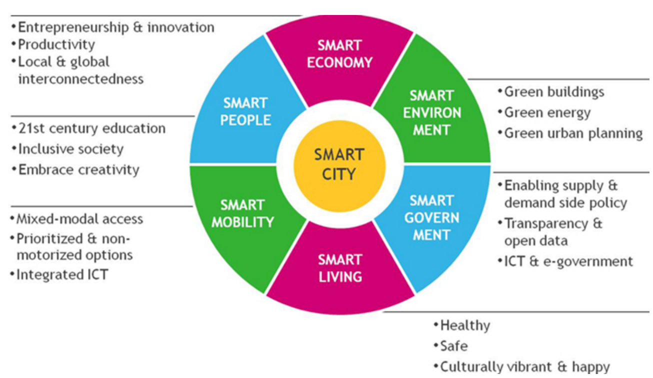

Figure 1: Smart City Characteristics – Highlights

We aim to connect ordinances to relevant tweets by drawing on their semantic relatedness. This is non-trivial, as ordinances and tweets both involve highly intricate and rather heterogeneous natural language, so simple keyword matching does not suffice. Traditional machine learning techniques [36] and related advances are not found suitable for learning this sort of mapping, as they require vast amounts of training data. Since ours is pioneering work in ordinance mining, we do not have such prior training data.

To overcome these challenges, we propose a two-step approach for mapping that exploits the transitive nature of the connection between ordinances and tweets considering their relationship with SCCs. Specifically, the transitive property we invoke is that: *if the ordinance relates to a given SCC and any tweet relates to the same SCC, then the ordinance bears a connection to the tweet.* This approach is proposed because classical sources of SCC data, e.g. [16, 19] are finite and are restricted to a limited set of identifying features that can be relied upon for mapping (see Figure 1). Thus, this transitive approach is more feasible than attempting to directly relate a seemingly infinite number of tweets to ordinances from various websites.

As a first step, we discover connections between SCCs and ordinances using classical SCC sources guided by common sense knowledge (CSK) from web-based repositories. In a second step, we consider the mapping of tweets to SCCs, again drawing on such CSK. This approach then enables us to directly relate ordinances and the tweets to the pertinent aspects of Smart Cities and also sets the stage for sentiment polarity classification [10, 23] and sentiment aspect analysis [34] of pertinent tweets using suitable methods to assess public opinion.

This work aims for broader impact by contributing to the development towards Smart Cities. If we identify which SCCs are being addressed by the local laws or ordinances passed by urban agencies, we are able to provide feedback on how well their urban policies head towards Smart City development across various categories.

Moreover, this work relates to the theme of Social Sensing. Public reactions inferred from opinion mining (to be conducted after connecting ordinances to tweets using SCCs) can further enable involved urban councils and management agencies to judge public satisfaction. This can allow for assessing the appeal of Smart City ordinances from a public opinion standpoint, thus providing useful feedback to the agencies that may enable them to enhance their policies for Smart City development. To achieve this sort of analysis, we draw on artificial intelligence aspects of text mining, natural language processing, and common sense knowledge.

The rest of this paper is organized as follows. Section 2 describes pertinent related work. Section 3 explains our proposed mapping approach to connect ordinances and social media postings. Section 4 summarizes its evaluation through experiments and discussion. Section 5 gives the conclusions, including our findings and a description of ongoing research.

2 RELATED WORK

While there has been ample work on mining social media, most previous work differs substantially from the task we consider here.

There is a long history of research on link prediction in social networks [2, 35]. These methods, however, are geared towards creating links between homogeneous sorts of nodes, such as predicting friendship connections between pairs of social network users. The same applies to most of the research on the even longer standing problems of entity resolution [7] and alignment between resources [8]. Only few approaches have targeted open-domain linking between arbitrary entities and concepts [1, 4, 9, 25, 26]. However, these typically assume structured data as input, i.e. entities with a series of attributes. In our case, we are attempting to connect two forms of unstructured natural language text. On the one side, we have public ordinances expressed using highly formal language, replete with legal terminology. On the other side, we have social media posts consisting of text that is typically very informal in nature, including embedded hashtags, URLs, etc.

For social media text, one important line of inquiry has focused on unsupervised topic modeling and trend detection in social media [15]. In [38], a fuzzy-based approach is used to preprocess and analyze hashtags in Twitter with the resulting fuzzy clusters being studied to investigate temporal trends on hashtag popularity. Such works however cannot easily be applied to the task of mapping tweets to a pre-existing set of ordinances, which we consider in our research. Neural vector-based representations of documents [5] also fail when the two items are as heterogeneous as in our case.

Some recent approaches on linking social media text have relied on supervised classification. While standard methods can be applied to predict links between heterogeneous items [36], an important challenge is that large training sets are required to accurately cope with the short length (leading to data sparsity) and variability of tweets. To overcome this, the TweetSift system [18] classifies tweets by topic while exploiting external entity knowledge and topic-enhanced word embeddings. The latter leads to topic-specific word embeddings such that the different senses of ambiguous words obtain different representations. However, this assumes that the knowledge base can provide highly pertinent signals about entities such as specific Twitter users. Our model in contrast exploits generic common sense knowledge and does not require a detailed labeled training set.

Furthermore, previous work has not considered the setting of ordinances (with tweets) and Smart City Characteristics, along with their challenging use of language. To the best of our knowledge, our work is therefore among the pioneering research in this area.

Much attention is being given to Smart Cities in recent years. Buses in Barcelona are designed to run on routes optimal for power consumption [19]. Canal lights in Amsterdam automatically brighten and dim based on pedestrian usage [19]. The work in [22] addresses the potential enhancement of automated vehicles by embedding them with common sense knowledge. Such initiatives contribute mainly to the *Smart Mobility* characteristic. There is also significant research on making use of technology in fighting crime, e.g. the monitoring system to identify and categorize crime-related events in text documents [24] that was developed within the EU ePOOLICE project. Such research contributes to the *Smart Living* characteristic. The work in [21] targets the *Smart Environment* characteristic through cloud computing solutions for data centers (instead of on-premise servers). They analyze scenarios where cloud models provide greater energy efficiency, yet meeting productivity targets. Security, privacy, and availability issues are discussed for cloud usage in the greening of data centers. Free cooling for data centers as addressed in [20] by considering temperature, humidity and other parameters, also contributes to *Smart Environment.* The work in [37] has a tangential influence on *Smart Economy.* The authors propose a mathematical model to minimize trips in scheduled pickups and deliveries by cooperation. This is a cost-effective method useful in urban delivery systems to reduce operational expenses in a cooperative mode. Likewise, the research conducted in [33] has an indirect impact on the *Smart People* characteristic by addressing an aspect of 21st century education through collocation-based writing aids for second language learners of English, as they constitute a large part of the population in cities worldwide. The work of [12] while primarily impacting *Smart Environment* through its estimation of air quality by analyzing pollutant data, also has a secondary impact on *Smart Living* since it addresses issues from a health standpoint. Thus, several researchers are conducting studies to augment the characteristics of Smart Cities.

Our work in this paper seeks to make a notable impact here, by advocating for the deployment of common sense knowledge in the realm of Smart Cities. While works such as [11, 17] motivate the need for common sense in the areas of *Smart Mobility* and *Economy*, respectively, the actual use of such knowledge in these paradigms remains at the stage of inception, e.g. [22]. As addressed in several works on common sense in machine intelligence (acquisition, representation, and application) surveyed in [31], the increased usage of CSK in many areas would promote much smarter machines. Our research in this paper aims to take a significant step along this avenue, with the overall goal of enhancing Smart Cities.

3 PROPOSED MAPPING APPROACH

The approach we propose for ordinance to tweet mapping through Smart City Characteristics (SCCs) is illustrated in Figure 2. It is described in detail in the following subsections.

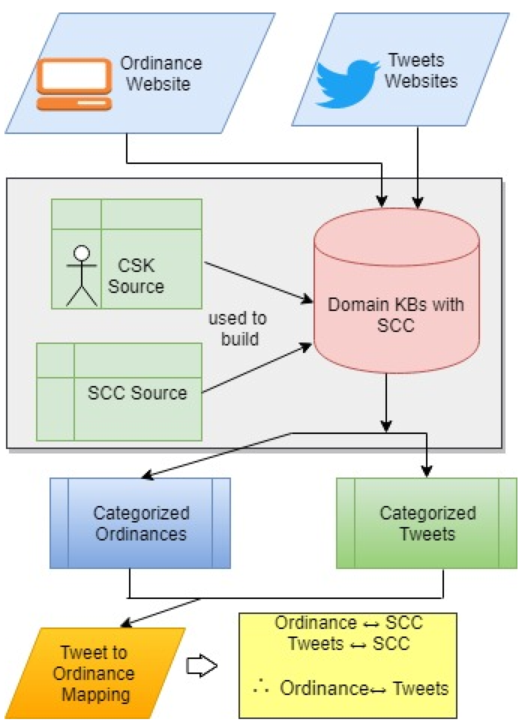

Figure 2: Proposed approach for SCC mapping

3.1 CSK — SCC based KB Development

The *SCC source* used in our approach is derived from the widely accepted technical report from TU Wien [19], which enumerates six SCCs. These are *Smart Governance* (or *Government*), *Smart Economy*, *Smart Mobility*, *Smart Environment*, *Smart People*, and *Smart Living*, respectively.

Consider, for instance, the SCC *Smart Governance*. This encompasses the features listed next, some of which are also included among the highlights listed in Figure 1.

- Transparency in government
- Optimizing public service and administration
- Direct involvement in public policies
- Citizen participation
- Positive and open communication channel with citizens
- More informed decisions by feedback and engagement

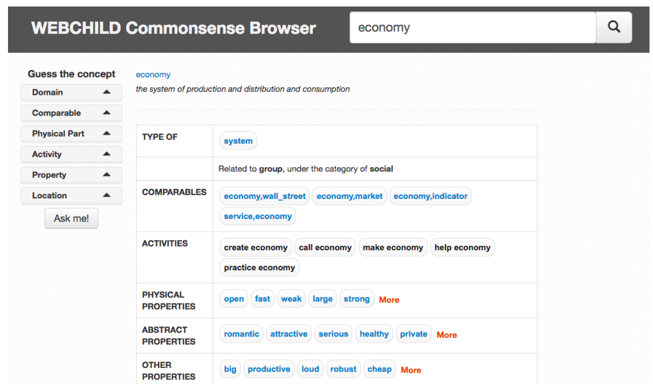

Figure 3: Relevant partial screenshot of WebChild

Thus, if ordinances reference any of the above features, we infer that they likely relate to Smart Governance. However, these expressions are not particularly likely to be observed in the ordinances literally. If human users were to inspect these ordinances, they could draw relevant connections, which are often quite subtle, by relying on linguistic knowledge and common sense. To automate this process, we draw on common sense knowledge (CSK) web sources, specifically, the large WebChild repository [28, 30] with common sense concepts mined from vast amounts of data on the Web along with their *properties* and *relationships*. A partial screenshot of the WebChild browser appears in Figure 3. This depicts a relevant concept *economy*, which pertains to a specific SCC.

Using WebChild as the main *CSK source* along with requisite information for knowledge base development [32] and other common sense related sources such as the lexical database WordNet [14], we build domain-specific knowledge bases on Smart City Characteristics (Domain KBs with SCC). These KBs are text-based and contain terms relevant to specific Smart City Characteristics derived from CSK repositories and SCC sources, using NLP and semantic matching. Note that one could also apply techniques such as knowledge base extraction from text [27, 29] and rule mining [6] to increase the size of these domain KBs. Figure 4 shows a subset of our Domain KBs with terms relevant to the characteristics of *Smart Environment* and *Smart Mobility*.

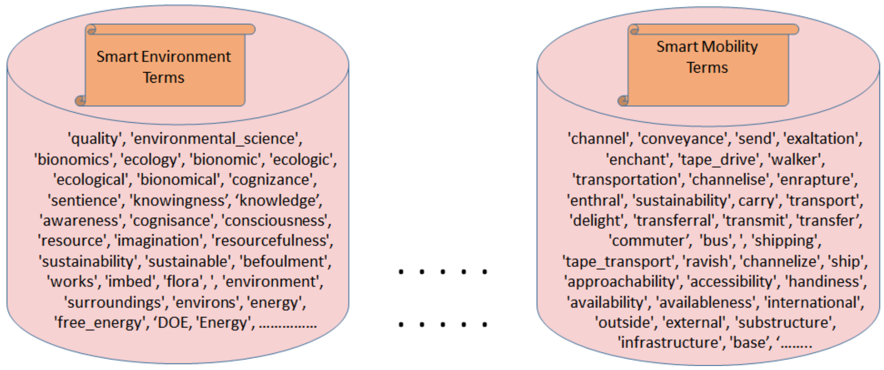

Figure 4: Part of Domain KBs with SCC (Subset of Smart Environment and Smart Mobility terms)

Algorithm 1 Linking algorithm

for each SCC S_j **do**
  Build domain knowledge base K_j
for each ordinance O_i **do** ▷ ordinance linking
  for each SCC S_j **do**
    $L_{i,j} \leftarrow \sum_{x \in K_j} \mathrm{C}(O_i, x)$
  Assign O_i to the S_j with $j = \mathrm{argmax}_j\, L_{i,j}$
for each social media posting T_i **do** ▷ social media linking
  for each SCC S_j **do**
    $M_{i,j} \leftarrow \sum_{x \in K_j} \mathrm{C}(T_i, x)$
  Assign T_i to the S_j with $j = \mathrm{argmax}_j\, M_{i,j}$
$\mathcal{O} \leftarrow \{(O_i, T_k) \mid \exists\, S_j : (O_i \text{ assigned to } S_j) \wedge (T_k \text{ assigned to } S_j)\}$
return $\mathcal{O}$ ▷ Links between ordinances and social media

3.2 Linking using SCCs and CSK

Using these domain KBs, CSK concepts are deployed to semantically relate terms x in ordinance text T to SCCs. We denote this as $C(T, x)$. For example, if the ordinance text includes the term *smoke detector*, then CSK concepts help to semantically relate this with the SCC *Smart Environment* through the CSK properties of *smoke detector* that have features relevant to this SCC. This information is found in the domain KBs derived from the SCC and CSK sources.

The same ordinance can also have features that relate to other SCCs. It is possible that some terms in ordinances may overlap with multiple SCCs. In that case, they would be observed in the KBs of each of those SCCs. If such concept terms are discovered in the ordinances, their occurrences are counted towards multiple categories. For example, if a term such as *sustainability* occurs in an ordinance, then that ordinance would be counted under the characteristics of *Smart Mobility* as well as *Smart Environment* (see Figure 4). Thus, the counts for both of these SCCs would be updated in this particular example. Finally, all the aggregate SCC counts are examined and each ordinance is accordingly linked to the SCC with the maximum number of relevant features. CSK plays a crucial role in finding semantic relatedness for this mapping through concepts, properties, etc. Likewise, we map tweets to SCCs following a similar CSK-guided procedure. Using this, we finally aim to output the linkages between ordinances and tweets via mutual SCC connections. Thus, we emphasize that: *an ordinance broadly links to a particular tweet if they both map to the same SCC.*

This mapping approach used for linking them is summarized in Algorithm 1 herewith. As of now, for simplicity, we emit only the closest matching SCC for the ordinances and tweets as output.

4 EVALUATION OF THE MAPPING

We conduct an evaluation of mapping ordinances and tweets with SCCs using large amounts of real data from publicly accessible websites on ordinances and tweets. A summary of our experimental evaluation is presented in the following.

4.1 Ordinance to SCC Mapping

Large amounts of historical data on ordinances are gathered from the website of the NYC council [3], which is openly available to the public. A small portion of this is shown in the screenshot that appears in Figure 5.

These ordinances are first extracted into a machine-readable form and then subjected to a preprocessing step such that only their textual content is retained. The other attributes such as "Prime Sponsor", "Council Member Sponsor", etc. (see Figure 5) are filtered out during this preprocessing phase. The textual content of the

Figure 5: Sample of NYC Council website

Table 1: A Sample Ordinance and its SCC Mapping

Smart City Characteristic	Count of Terms
Economy	1
Environment	0
Governance	15
People	0
Mobility	0
Living	0

ordinances then serves as input to our algorithm that conducts the ordinance to SCC mapping. The algorithm interfaces with the SCC KB and uses the relevant terms for mapping. This inking procedure is formalized within Algorithm 1. It accordingly counts all such matches to output the SCC with the maximum counts as the closest matching one.

Shown herewith is an excerpt from an ordinance (Ord. 1) from the aforementioned NYC council website, along with its closest matching SCC (Table 1) based on quantifying relevant ordinance terms with SCC features.

> Ord. 1: *A Local Law to amend the administrative code of the city of New York, in relation to amending the district plan of the Downtown – Lower Manhattan business improvement district to change...*

With reference to this ordinance excerpt, our algorithm relies on the SCC Domain KBs and comes to the conclusion that the only term relevant to the *Smart Economy* characteristic is *business*, while many terms are relevant to *Smart Governance*, including, among others, *law*, *administrative*, *district plan*, *improvement*, etc., as summarized in Table 1. Thus, the SCC that is returned as the closest matching one in this example is *Smart Governance*.

Numerous further ordinances are analyzed following the same pattern. Note that in our execution so far, only the closest matching SCC is offered as the ordinance mapping output, for simplicity. The same holds for the mapping of tweets to SCCs.

We evaluate of a subset of NYC council data that encompasses two recent ordinance sessions, namely, 2006 to 2009 and 2010 to 2013. Based on this evaluation, we obtain a summary plot of ordinance to SCC mappings given in Figure 6.

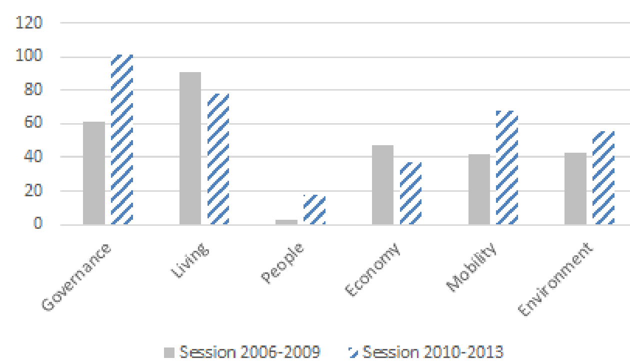

Figure 6: Summary plot of ordinance SCC mapping

The observations in this summary plot are useful to provide some feedback to urban management agencies on the extent to which their ordinances cater to various aspects of Smart Cities. For example, from the results, one can conclude that the Smart City Characteristic receiving the greatest attention is *Smart Living* in the first session and *Smart Governance* in the second session. In both of the sessions, the SCC supposedly receiving the least attention is *Smart People*. This may help the urban agencies to plan their future policies such that they make progress on policies pertaining also to those characteristics that have been received comparably little attention so far, in this case the *Smart People* characteristic. Details on various aspects of urban legislation impacts with respect to such analysis appear in [13] catering mainly to a domain-specific angle. This is an important motivation for our current research with ordinances, tweets and SCCs.

4.2 Tweet to SCC Mapping

We extract thousands of tweets posted by the public on Twitter pertaining to NYC location-specific data. The Twitter Streaming API feature labeled *Filter Realtime Tweets* is used for conducting the extraction. The tweets are extracted to a text file and further processed using NLP techniques such as regular expressions. The relevant parts of the tweets such as their textual content and hyperlinks are retained. These are stored as cleaned tweets. The SCC mapping is then performed on the cleaned tweets using the concerned part of our approach as depicted in Algorithm 1. In Figure 7, we show only a small subset of cleaned tweets used among over 1,000 tweets extracted in our experiments.

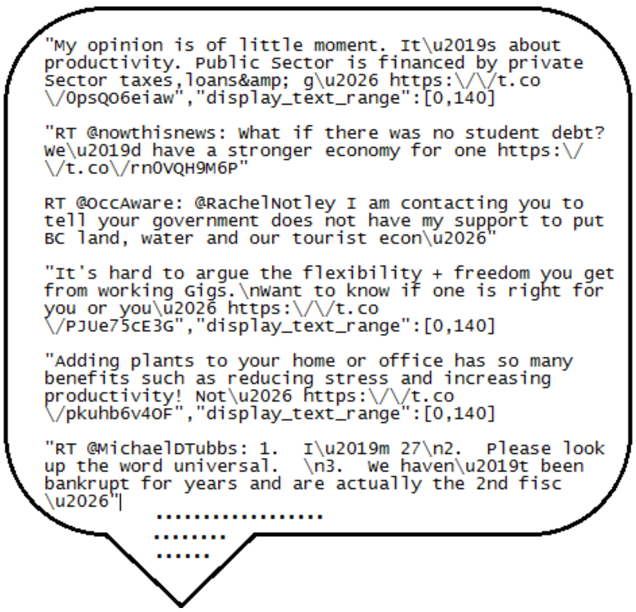

Figure 7: Subset of tweets analyzed from NYC sites

Based on these tweets, Figure 8 depicts a partial snapshot of our program mapping these cleaned tweets to their most relevant SCC, with reference to the relevant part of the process in Algorithm 1. This is interpreted as follows. Among tweets processed herewith, the overall mapping indicates that 37 of them are on *Smart Economy*, 25 are on *Smart Environment*, 208 on *Smart Living*, etc. These are obtained by the processing shown in the figure, e.g., features of the *Smart Living* SCC include the terms: *home*, *benefits*, *tourist*, *building*, etc., while those of *Smart Environment* include: *energy*, *sustainable*, etc. (The terms are obtained from KBs built using CSK and SCC sources). It is observed in this figure that, overall, 352 tweets are mapped to SCCs (37+25+…+208). Hence, many tweets among approximately 1000 cleaned ones analyzed in these experiments are not mapped to any SCC. This could be due to the fact that not all tweets published by users pertain to SCCs. It could also be that some mappings are not precisely identified in the initial experiments conducted herewith.

```
('Smart City Characteristic      Occurences')
('  Smart Economy                  ', 37)
('  Smart Environment              ', 25)
('  Smart Governance               ', 45)
('  Smart People                   ', 19)
('  Smart Mobility                 ', 38)
('  Smart Living                   ', 208)

('SMART LIVING FEATURE: TERM ', 'home', 1)
('SMART GOVERANCE FEATURE: TERM ', 'office', 1)
('SMART LIVING FEATURE: TERM ', 'benefits', 2)
('SMART ENVIRONMENT FEATURE: TERM ', 'reducing', 1)
('SMART ECONOMY FEATURE: TERM ', 'economy', 1)
('SMART ECONOMY FEATURE: TERM ', 'economy', 2)
('SMART LIVING FEATURE: TERM ', 'tourist', 3)
('SMART ECONOMY FEATURE: TERM ', 'economy', 3)
('SMART ECONOMY FEATURE: TERM ', 'economy', 4)
('SMART LIVING FEATURE: TERM ', 'tourist', 4)
('SMART LIVING FEATURE: TERM ', 'building', 5)
('SMART MOBILITY FEATURE: TERM ', 'car', 1)
('SMART ENVIRONMENT FEATURE: TERM ', 'energy', 2)
('SMART ENVIRONMENT FEATURE: TERM ', 'sustainable', 3)
...........................
......................
```

Figure 8: Partial snapshot of tweet to SCC mapping

4.3 Assessment and Discussion

In order to facilitate judging the correctness of the mappings, we have developed very simple GUIs in our initial execution. We illustrate a relevant part of our *Tweet Mapping GUI* next. This accepts a tweet as the input and emits the closest matching SCC from the user as its output, or "*No matches*" if none gets matched. Figure 9 shows an example of a tweet and its SCC identified as *Smart Environment*, while an example of a non-matching tweet is given in Figure 10. Both of these are partial GUI screenshots.

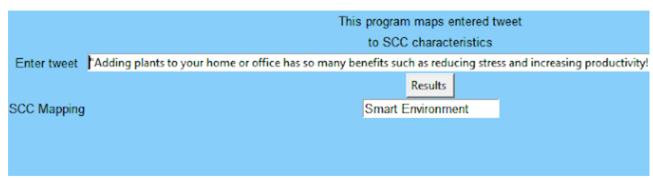

Figure 9: Example of SCC mapping identified

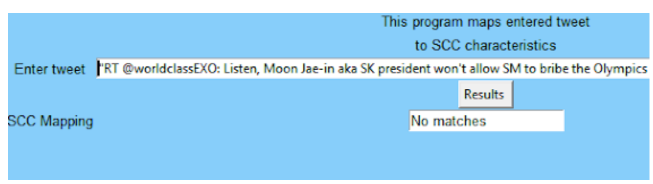

Figure 10: Example of no matches for any SCC

Considering several tweets entered and SCCs identified through this GUI, the correctness of these mappings is assessed by domain experts from Earth and Environmental Studies. A similar *Ordinance Mapping GUI* is provided for the ordinance to SCC mappings for enabling at-a-glance displays. These mappings are also assessed by the domain experts.

The actual calculation of accuracy is done using an *Accuracy* metric as follows. Considering the judgment provided by domain

Table 2: Accuracy of Ordinance and Tweet Mapping

	Ordinances	Tweets
Expert 1	84%	72%
Expert 2	86%	69%
Expert 3	81%	70%

experts, if an ordinance or tweet is mapped to a given SCC by our proposed approach (or if it returns a *No Match*) and this is verified as correct by the expert, it is considered a *True Mapping (TM)*. If the expert labels this mapping as incorrect, it is a *False Mapping (FM)*. For example, if the approach indicates that the SCC is *Smart Governance*, but the expert states that it is *Smart Economy* or that it is a *No Match*, it would be a *False Mapping*. Also, if the approach indicates a *No Match*, but the expert states that it maps to a given SCC, it is still considered a *False Mapping*. In other words, the ground truth is defined by experts for the data analyzed herewith.

With this justification, we proceed to calculate

$$Accuracy = \frac{TM}{TM + FM}, \quad (1)$$

and this is used for measuring the effectiveness of our proposed mapping approach. This is analogous to the classical notion of true positives and false positives in data mining and machine learning techniques [36]. (We do not consider true negatives and false negatives at this point in our research, since their appropriate definition needs further insights and discussions with domain experts. This is an aspect of future work). Based on the given definition of Accuracy herewith, we obtain the evaluation scores as listed in Table 2.

Thus, the ordinance to SCC mapping, as verified by domain experts, is found to be accurate for around 85% of the ordinances. This is considered satisfactory on the whole, although there is scope for improvement. The main reasons for the difference are that some ordinances can actually map almost equally to multiple SCCs, and hence it is possible that our approach identifies one particular SCC as the top match, while an expert identifies another.

The accuracy of the tweet to SCC mapping is in the range of around 70%, which seems fairly reasonable for a start. However, it is much lower than that of the ordinance to SCC mapping. We present a few examples of tweets below that are classified incorrectly or return no match, thereby adversely affecting the performance of the tweet to SCC mapping in our approach.

- "Wind 0.0 mph N. Barometer 30.134 in, Falling slowly. Temperature 26.1 °F. Rain today 0.00in. Humidity 94%"
- "RT @worldclassEXO: Listen, Moon Jae-in aka SK president won't allow SM to bribe the Olympics bcs that's gonna ruin the country's reputation…"
- "@FoxNews No DACA until Wall is built."
- "Our February STEM Hero is... #STEMed #STEM #SciEd #ScienceEd @polyprep https://9O7Gf5sZP7…"

Inspecting such examples, an important observation is that the problem of inaccurate mapping (or that of no matches being found) occurs mainly due to challenges such as ambiguity, informal language, excessive use of acronyms and hashtags, etc. These issues pose significant challenges in the execution of the mapping. This calls for further research on the tweet to SCC mapping process. We have encountered the challenges listed next in the tweet mapping part of our research.

(1) Tweets use informal language, which makes their extraction and analysis difficult.
(2) The length restrictions imposed on tweets results in users resorting to an excessive use of acronyms.
(3) There is limited coverage, e.g., 1/3 of mentions on the web cannot be linked to Wikipedia (around 30% loss).
(4) NEE (Named Entity Extraction) and NED (Named Entity Disambiguation) involve many degrees of uncertainty.

Addressing these non-trivial challenges, while also aiming to improve the ordinance to SCC mapping accuracy and considering the semantic proximity to multiple SCCs via rankings, constitutes our ongoing work. We further aim to drill down to a finer level in the mapping, which would entail identifying fine-grained aspects of the individual features in the SCCs as opposed to the entire SCC per se. This is likely to yield better performance.

5 CONCLUSION

This paper proposes an approach to map ordinances to tweets expressing public opinion, based on Smart City Characteristics (SCCs) relevant to both of them. The execution of our approach with initial experiments yields an accuracy of ordinance to SCC mapping of around 85%, while that of the tweet to SCC mapping is approximately 70%, both confirmed by domain experts.

Ongoing research includes addressing challenges in the tweet to SCC mapping, improving the accuracy of the ordinance to SCC mapping, considering the mapping of both ordinances and tweets to multiple SCCs with ranking, and also attaining a finer granularity in the mappings besides the broad categorizations considered here. This would enable us to draw more specific conclusions from the results, particularly when they are used for polarity classification of tweets to assess the public reaction.

The long term vision of our research is to provide urban agencies useful feedback on how well they are doing in policy decisions (based on this mining) and hence indicate how closely the given urban region heads towards a Smart City. In summary, our research makes the following contributions:

(1) addressing the mining of local laws or ordinances and their public reaction through tweets to give urban agencies useful feedback, which is pioneering work in the area;
(2) proposing an approach for ordinance to tweet mapping using Smart City Characteristics as a nexus, deploying the transitive property of semantic relatedness between them;
(3) conducting a study with genuine ordinance data from the NYC council, with mapping accuracy of around 85% (ordinance to SCC) and 70% (tweet to SCC) respectively;
(4) motivating the need for mapping ordinances and tweets to multiple SCCs with ranking, and dealing with finer levels of granularity in SCC features for enhanced performance

Ultimately, we envision our work as contributing to the development of Smart Cities on the whole as a broader impact.

ACKNOWLEDGMENTS

We acknowledge Niket Tandon from the Allen Institute for Artificial Intelligence (Seattle, WA) for his valuable inputs. We thank Robert Taylor and Clement Alo from Earth and Environmental Studies at Montclair State University for their feedback. Early work on this project started while Aparna Varde was a visiting researcher at the Max Planck Institute for Informatics (Saarbrücken, Germany) in the group of Gerhard Weikum in August 2015. Gerard de Melo's research at Rutgers University is funded in part by ARO grant no. W911NF-17-C-0098 as part of the DARPA SocialSim program. Manish Puri and Xu Du are funded by a Research Assistantship (Computer Science) and a Doctoral Assistantship (Environmental Management) respectively at Montclair State University.

REFERENCES

[1] Christoph Böhm, Gerard de Melo, Felix Naumann, and Gerhard Weikum. 2012. LINDA: Distributed Web-of-Data-Scale Entity Matching. In *Proceedings of the 21st ACM Conference on Information and Knowledge Management (CIKM 2012)*, Xue-wen Chen, Guy Lebanon, Haixun Wang, and Mohammed J. Zaki (Eds.). ACM, New York, NY, USA.

[2] Zhu Cao, Linlin Wang, and Gerard de Melo. 2018. Link Prediction via Subgraph Embedding-Based Convex Matrix Completion. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI 2018)*. AAAI Press.

[3] The New York City Council. 2018. Legislative Research Center Web Page. (2018). Retrieved March 4, 2018 from http://legistar.council.nyc.gov/

[4] Gerard de Melo. 2013. Not Quite the Same: Identity Constraints for the Web of Linked Data. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence (AAAI 2013)*, Marie desJardins and Michael L. Littman (Eds.). AAAI Press, Menlo Park, CA, USA, 1092–1098. http://www.aaai.org/ocs/index.php/AAAI/AAAI13/paper/view/6491

[5] Gerard de Melo. 2017. Multilingual Vector Representations of Words, Sentences, and Documents. In *Proceedings of IJCNLP 2017*. http://www.aclweb.org/anthology/I17-5002

[6] Gerard de Melo, Mouna Kacimi, and Aparna Varde. 2015. Dissertation Research Problems in Data Management and Related Areas. *SIGMOD Record* 44, 4 (December 2015). http://sigmod.org/publications/sigmodRecord/1512/pdfs/09_reports_Melo.pdf

[7] Gerard de Melo and Gerhard Weikum. 2008. Language as a Foundation of the Semantic Web. In *Proceedings of the Poster and Demonstration Session at the 7th International Semantic Web Conference (ISWC 2008) (CEUR WS)*, Christian Bizer and Anupam Joshi (Eds.), Vol. 401. CEUR, Karlsruhe, Germany.

[8] Gerard de Melo and Gerhard Weikum. 2008. Mapping Roget's Thesaurus and WordNet to French. In *Proceedings of the 6th Language Resources and Evaluation Conference (LREC 2008)*. ELRA, Paris, France, 3306–3313.

[9] Gerard de Melo and Gerhard Weikum. 2010. Untangling the Cross-Lingual Link Structure of Wikipedia. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Stroudsburg, PA, USA.

[10] Xin Dong and Gerard de Melo. 2018. Cross-Lingual Propagation for Deep Sentiment Analysis. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI 2018)*. AAAI Press.

[11] Katherine Rose Driggs-Campbell, Victor Shia, and Ruzena Bajcsy. 2014. Decisions for Autonomous Vehicles: Integrating Sensors, Communication, and Control. In *Proceedings of the 3rd International Conference on High Confidence Networked Systems (HiCoNS '14)*. ACM, New York, NY, USA, 59–60. https://doi.org/10.1145/2566468.2576850

[12] Xu Du, Onyeka Emebo, Aparna Varde, Niket Tandon, Sreyasi Nag Chowdhury, and Gerard Weikum. 2016. Air Quality Assessment from Social Media and Structured Data: Pollutants and Health Impacts in Urban Planning. In *IEEE International Conference on Data Engineering (ICDE) - Workshops*. 54–59.

[13] Xu Du, Diane Liporace, and Aparna Varde. 2017. Urban Legislation Assessment by Data Analytics with Smart City Characteristics. In *IEEE Ubiquitous Computing, Electronics and Mobile Communications Conference (UEMCON)*. 20–25.

[14] Christiane Fellbaum (Ed.). 1998. *WordNet: An Electronic Lexical Database*. The MIT Press. http://www.amazon.ca/exec/obidos/redirect?tag=citeulike04-20{&}path=ASIN/026206197X

[15] Liangjie Hong and Brian D. Davison. 2010. Empirical Study of Topic Modeling in Twitter. In *Proceedings of the First Workshop on Social Media Analytics (SOMA '10)*. ACM, New York, NY, USA, 80–88. https://doi.org/10.1145/1964858.1964870

[16] IEEE. 2018. IEEE Smart Cities Technical Community. (2018). Retrieved March 4, 2018 from https://smartcities.ieee.org/

[17] G. Leef. 2007. *Smart Economics: Commonsense Answers to 50 Questions about Government, Taxes, Business and Households*. Foundation for Economic Education.

[18] Quanzhi Li, Sameena Shah, Xiaomo Liu, Armineh Nourbakhsh, and Rui Fang. 2016. TweetSift: Tweet Topic Classification Based on Entity Knowledge Base and Topic Enhanced Word Embedding. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (CIKM '16)*. ACM, New York, NY, USA, 2429–2432. https://doi.org/10.1145/2983323.2983325

[19] TU Wien (Vienna University of Technology). 2015. *European Smart Cities, Technical Report*. Technical Report. Vienna, Austria.

[20] Michael Pawlish and Aparna Varde. 2010. Free Cooling: A Paradigm Shift in Data Centers. In *IEEE International Conference on Information and Automation for Sustainability (ICIAfS)*. 347–352.

[21] Michael J. Pawlish, Aparna S. Varde, and Stefan A. Robila. 2015. The Greening of Data Centers with Cloud Technology. *Int. J. Cloud Appl. Comput.* 5, 4 (Oct. 2015), 1–23. https://doi.org/10.4018/IJCAC.2015100101

[22] Priya Persaud, Aparna Varde, and Stefan Robila. 2017. Enhancing Autonomous Vehicles with Commonsense: Smart Mobility in Smart Cities. In *IEEE International Conference on Tools with Artificial Intelligence (ICTAI) - Smart Cities Workshop*.

[23] Andi Rexha, Mark Kröll, Mauro Dragoni, and Roman Kern. 2016. Polarity Classification for Target Phrases in Tweets: A Word2Vec Approach. In *The Semantic Web - ESWC 2016 Satellite Events, Heraklion, Crete, Greece, May 29 - June 2, 2016, Revised Selected Papers (Lecture Notes in Computer Science)*, Harald Sack, Giuseppe Rizzo, Nadine Steinmetz, Dunja Mladenic, Sören Auer, and Christoph Lange (Eds.), Vol. 9989. 217–223. https://doi.org/10.1007/978-3-319-47602-5_40

[24] Jacobo Rouces, Gerard de Melo, and Katja Hose. 2015. Representing Specialized Events with FrameBase. In *Proceedings of the 4th International Workshop on Detection, Representation, and Exploitation of Events in the Semantic Web (DeRiVE 2015) at ESWC 2015*. http://ceur-ws.org/Vol-1363/paper_7.pdf

[25] Jacobo Rouces, Gerard de Melo, and Katja Hose. 2016. Complex Schema Mapping and Linking Data: Beyond Binary Predicates. In *Proceedings of the WWW 2016 Workshop on Linked Data on the Web (LDOW 2016)*.

[26] Jacobo Rouces, Gerard de Melo, and Katja Hose. 2016. Heuristics for Connecting Heterogeneous Knowledge via FrameBase. In *Proceedings of ESWC 2016 (Lecture Notes in Computer Science)*. Springer.

[27] Niket Tandon and Gerard de Melo. 2010. Information Extraction from Web-Scale N-Gram Data. In *Web N-gram Workshop. Workshop of the 33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Chengxiang Zhai, David Yarowsky, Evelyne Viegas, Kuansan Wang, and Stephan Vogel (Eds.), Vol. 5803. ACM, 8–15.

[28] Niket Tandon, Gerard de Melo, Fabian Suchanek, and Gerhard Weikum. 2014. WebChild: Harvesting and Organizing Commonsense Knowledge from the Web. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining (WSDM '14)*. ACM, New York, NY, USA, 523–532. https://doi.org/10.1145/2556195.2556245

[29] Niket Tandon, Gerard de Melo, and Gerhard Weikum. 2011. Deriving a Web-Scale Common Sense Fact Database. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence (AAAI 2011)*. AAAI Press, Palo Alto, CA, USA, 152–157.

[30] Niket Tandon, Gerard de Melo, and Gerhard Weikum. 2017. WebChild 2.0 : Fine-Grained Commonsense Knowledge Distillation. In *Proceedings of ACL 2017, System Demonstrations*. Association for Computational Linguistics, Vancouver, Canada, 115–120. http://aclweb.org/anthology/P17-4020

[31] Niket Tandon, Aparna S. Varde, and Gerard de Melo. 2018. Commonsense Knowledge in Machine Intelligence. *SIGMOD Rec.* 46, 4 (Feb. 2018), 49–52. https://doi.org/10.1145/3186549.3186562

[32] Aparna Varde, Niket Tandon, Sreyasi Nag Chowdhury, and Gerhard Weikum. 2015. *Common Sense Knowledge in Domain-Specific Knowledge Bases*. Technical Report. Max-Planck-Institut für Informatik, Saarbrücken, Germany.

[33] Alan Varghese, Aparna Varde, Jing Peng, and Eileen Fitzpatrick. 2015. A Framework for Collocation Erroe Correction in Web Pages and Text Documents. *ACM SIGKDD Explorations* 17, 1 (June 2015), 14–23.

[34] Linlin Wang, Kang Liu, Zhu Cao, Jun Zhao, and Gerard de Melo. 2015. Sentiment-Aspect Extraction based on Restricted Boltzmann Machines. In *Proceedings of ACL 2015*. 616–625.

[35] Liqiang Wang, Yafang Wang, Bin Liu, Lirong He, Shijun Liu, Gerard de Melo, and Zenglin Xu. 2017. Link Prediction by Exploiting Network Formation Games in Exchangeable Graphs. In *Proceedings of IJCNN 2017*.

[36] Ian H. Witten, Eibe Frank, and Mark A. Hall. 2011. *Data Mining: Practical Machine Learning Tools and Techniques* (3rd ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[37] Yang Yu, Qi Lou, Jiafu Tang, Junwei Wang, and XiaoHang Yue. 2017. An exact decomposition method to save trips in cooperative pickup and delivery based on scheduled trips and profit distribution. *Computers Operations Research* 87 (2017), 245 – 257. https://doi.org/10.1016/j.cor.2017.02.015

[38] L. A. Zadeh, A. M. Abbasov, and S. N. Shahbazova. 2015. Analysis of Twitter hashtags: Fuzzy clustering approach. In *2015 Annual Conference of the North American Fuzzy Information Processing Society (NAFIPS) held jointly with 2015 5th World Conference on Soft Computing (WConSC)*. 1–6. https://doi.org/10.1109/NAFIPS-WConSC.2015.7284196