

Feasibility and Framing of Interventions Based on Public Support: Leveraging Text Analytics for Policymakers

Philippe J. Giabbanelli^{1,2}(✉), Jean Adams², and Venkata Sai Pillutla¹

¹ Department of Computer Science, Northern Illinois University, DeKalb, IL, USA
giabba@cs.niu.edu, z1780791@students.niu.edu

² UKCRC Centre for Diet and Activity Research (CEDAR),
MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine,
Institute of Metabolic Science, Cambridge CB2 0QQ, UK
jma79@medschl.cam.ac.uk

Abstract. Public opinions play an important role in planning policies. A beneficial population intervention may not be publicly acceptable, or policymakers may be over-cautious and believe their constituents do not sufficiently support it. Understanding the feasibility and framing of interventions based on public support is thus an important endeavor for public health. While surveys or qualitative analyses are a typical approach, they can require significant time or manpower. In contrast, algorithms for text analytics are now available that could be readily used by policymakers. As a case study, this paper used the debate that surrounded taxes on sugar sweetened beverages (SSB) in California. Our main contribution lies in detailing the process of automatizing the analysis of public health opinions, particularly using off-the-shelf software that policymakers can use, and exemplify the types of policy questions that can be investigated.

1 Introduction

Close to two thirds of US adults are currently overweight or obese [1]. This has major consequences on quality of life as well as on healthcare costs, which are projected to reach 860 to 960 billion US dollars by 2030 [2] in the absence of long-term solutions. Several policies have been proposed to tackle the obesity epidemic [3–5]. These include economic measure, such as taxes on healthy food items [6–8] or subsidies for healthier ones [9]. There is growing evidence from both modelling and pragmatic studies [10,11], that taxes on less healthy foods can reduce sales and consumption of those foods. However, many of these more structural, population interventions may not be publicly acceptable, meaning that policymakers choose not to implement them and their potential is not realized [12]. Furthermore, policymakers may be over-cautious, thinking that some measures may not

Research funded by the Department of Computer Science and the College of Liberal Arts and Sciences at Northern Illinois University, and the Centre for Diet and Activity Research (CEDAR), a UKCRC Public Health Research Centre of Excellence.

be acceptable to their constituents without having strong evidence of this. Consequently, there is a need to better understand public opinions regarding these sorts of public health policies. This can help not only decide which policies to pursue, but also how to frame them in a more publicly acceptable way.

Surveys have long been used to assess public opinions about public health policies [13]. However, they need time to be deployed and analyzed, and may only provide limited insight into policy debates. Qualitative analyses of documents that are readily available (e.g., news stories, blogs, editorials) provide an alternative source of information on public opinions concerning public health policies. Information includes the topics or sentiments expressed by constituents. For example, Nixon *et al.* performed an ethnographic qualitative analysis (i.e. a type of qualitative analysis) of news reports on soda tax initiatives in 3 US cities [14]. However, content analyses typically require a significant amount of time and/or manpower as each article is read and coded by humans for sentiment, topic, or other variables of interest [15]. Some of these tasks can actually be performed by computers using text analytic algorithms, many of which are now part of off-the-shelf software. As recently highlighted by Hamad *et al.* [16], the automated analysis of news media in obesity is still in its infancy. Given the importance of addressing obesity, and the complexity of the public policy debates surrounding it, understanding how to use mature techniques from text analytics in this domain could offer important insights for public health researchers and policymakers.

The use of text analytics in public health research is part of the growing area of ‘infodemiology’, which studies the distribution and determinants of information in an electronic medium in order to inform public health and public policy [17]. Twitter has been the electronic medium of choice for many studies, as it provides publicly available data together with a social network and sometimes geo-coding [18–21]. However, Twitter messages (‘Tweets’) can be at most 140 characters long, which limits the information that they contain regarding a public health debate. In contrast, news articles allow arguments to be more fully developed. Newspapers are also one of the most trusted information sources, encountered by 65 % of US adults each week [22] with most readers being registered voters [23]. In addition, some US newspapers are still considered to clearly impact policy agenda [24]. While written news provides depth in arguments, it can be narrow in the breadth of opinions represented. For example, analyses of public debates regarding soda taxes have found that they mostly received positive news coverage even though the public ultimately voted against these taxes [14, 25]. One way to capture both the breadth of opinion included in social media such as Twitter, and the depth of opinion included in written news media, is to include both news reports and on-line reader responses to these. Consequently, we perform text analytics of news articles supplemented by readers’ comments. As a case study, we used the debate that surrounded taxes on sugar sweetened beverages (SSB) in California in 2014–15. While the *results* are thus most informative to that specific debate, the main contribution of this paper resides in detailing the *process* of automatizing the analysis of public opinions and exemplifying the types of public health questions that it supports.

This paper is organized as follows. In Sect. 2, we provide a brief background on policies regarding SSB taxes in California. Section 3 describes how we performed data collection and cleaning. After preparing the corpus, we analyze it in Sect. 4 using a variety of algorithms from text analytics, and we explain how these analyses provide different types of information in order to understand the public debate. Finally, we address technical limitations in Sect. 5 and provide a brief discussion on future work.

2 Public Policy Background

Despite evidence from modelling studies for a beneficial health effect of SSB taxes [10, 11], concerns over public acceptability of such policies are one reason why policy-makers and politicians appear reluctant to publicly consider them. The two SSB taxes in our case study were put to the vote in 2014 in Berkeley and San Francisco, California. An earlier 2011 survey in nearby Santa Clara county found that 67% would support a SSB tax and 37% would oppose it [26].

In Berkeley, CA, the tax was \$0.01 per fluid ounce on the distributors of sugar sweetened beverages (SSB), and syrups operating within the city. The proposal was for a tax that would apply to sugary soda, energy drinks, juice with added sugar, and syrups that go into sugary drinks. 100% juice and drinks with milk as the first (primary) ingredient were exempt because of their nutritional value. Diet soda and alcoholic drinks were also exempt. The tax revenue was designed to go into the city's general fund, and a Sugar-Sweetened Beverage Product Panel of Experts had to publish an annual report forming recommendations on how to allocate the funds to "reduce the consumption of sugar sweetened beverages in Berkeley and to address the results of such consumption" [27]. During the run-up to the November 2014 ballot, there was significant campaigning from those both opposed to, and in favour of, the SSB tax. Support for the tax focused around the campaign group *Berkeley vs Big Soda* funded primarily by contributions from local residents, public health organisations, and former New York City Mayor Michael R. Bloomberg. Opposition to the tax was driven by *Californians for Food & Beverage Choice* in association with the American Beverage Association, who reportedly spent around \$2.4m on campaigning [28]. This intense level of local campaigning generated substantial media coverage. Ultimately, the tax was put to the vote and received support from 76.16% of voters. A simple majority was required. By adopting this tax, Berkeley became the first city in the USA to introduce an SSB tax and one of the first jurisdictions to do so world-wide.

The same day and less than a 30 mins drive away, citizens of San Francisco, CA voted on a proposition for a tax of \$0.02 per fluid ounce payable, as in Berkeley, by SSB distributors. This was an 'hypothecated' tax with generated funds (estimated at \$31 million per annum) earmarked for health, nutrition and physical activity programmes in public schools and parks, at the direction of a Healthy Nutrition and Physical Activity Access Fund Committee. The proposition was sponsored by six local Supervisors. The official opponents were the Libertarian Party of San Francisco. However, the American Beverage Association also funded opposition through both the Coalition for an Affordable City,

a local pressure group formed specifically to oppose the ballot, and, as in Berkeley, *Californians for Food & Beverage Choice*. The proposition received 55.59% support from voters. However, because the proceeds from the tax were dedicated to specific purposes, approval of this measure required a 2/3 super-majority. Thus, the tax passed in Berkeley while it did not pass in San Francisco.

3 Creating a Corpus: Data Collection and Cleaning

3.1 Data Collection

The time period of interest and the main events are summarized in Fig. 1. We included news reports published between 1 January 2014 and 31 January 2015. On 11 February and 4 February, the decisions to vote on SSB tax through a formal ballot were made in Berkeley and San Francisco respectively. This represents the first formal step towards a decision to hold a ballot. By extending the data collection period back to 1 January 2014 we captured any coverage related to the run-up to this decision. The Berkeley tax was implemented on 1st January 2015. Extending the inclusion period to 31 January 2015 gave the opportunity to capture short, but not long, term reflections on the process of implementation.

In order to explore any changes in reporting over the time-periods before the ballot, after the ballot but before implementation, and after implementation, reports were considered in three groups. Reports published between 1 January 2014 and 4 November 2014 were *pre-ballot*; reports published between 5 November 2014 and 31 December 2014 were *post-ballot* but *pre-implementation*; and reports published between 1 January 2014 and 31 January 2015 were *post-implementation*.

We performed text analytics on all types of newspaper text articles (including news, features, editorials and other comment) as well as readers' comments. Infographics, videos or raw poll results were not considered as text articles. This

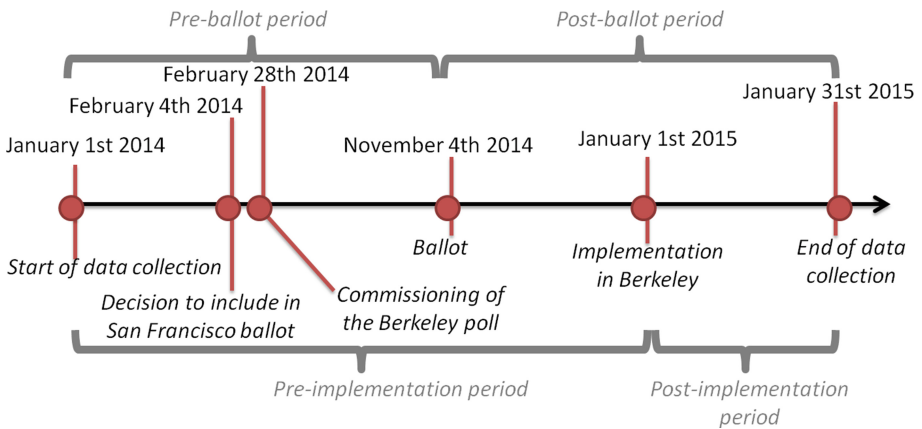


Fig. 1. Timescale for the data collection and main periods used in the analysis.

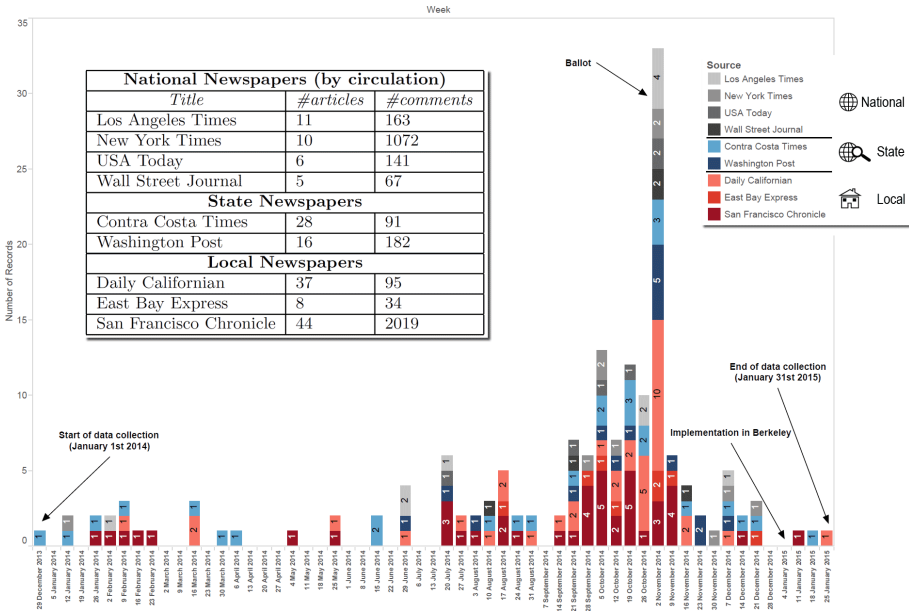


Fig. 2. Collected news articles by source and time. The inset summarizes selected newspapers with the number of articles (as found per selection criteria in Box 1) and corresponding comments.

strategy was chosen to capture comment articles written by leading members of the pro- and anti- lobbies, which go beyond news articles written by journalists. In addition, reader responses to these articles provide an insight into public perception and can potentially show a divide between the arguments used in newspapers and those held by the readers.

Newspapers were selected if they published at least 4 articles between 1 January 2014 and 31 January 2015 that matched our target content, per the rules summarized in Box 1. Candidate newspapers included local and national American newspapers, as well as international English-language newspapers. Four successive approaches were used to identify candidate newspapers, resulting in a total of 9 newspapers with 165 articles and 3,864 comments (Fig. 2 inset). The times at which these articles were written can be seen in Fig. 2, showing that articles often appeared around the elections as witnessed in previous policy research [14].

Box 1. Selection Criteria.

*(Berkeley OR San Francisco) AND tax AND (soda OR
 sweetened beverage OR sugary drink)*
 OR
(Berkeley AND measure D)
 OR
(San Francisco AND Proposition E)

First, we applied the search criteria via the LexisNexis database, which has a wide reach, particularly of American content. Using this database as the first step is a common approach [14]. This resulted in including the *Contra Costa Times*, *Los Angeles Times*, *New York Times*, and the *Washington Post*.

Second, we repeated the search criteria on each of the 5 largest daily newspapers in the USA (measured by the 2013 combined circulation and on-line viewing data compiled by the Alliance for Audited Media) via their own search facility. This resulted in adding *USA Today* and the *Wall Street Journal*. Third, we repeated the search criteria for newspapers that had a significant readership in either Berkeley or San Francisco. This resulted in inclusion of the *Daily Californian* (local Berkeley newspaper), the *East Bay Express*¹, and the *San Francisco Chronicle*².

Finally, we also applied the search criteria to the top 5 English-language newspapers outside the USA, by circulation figures. None were retained for analysis since our search criteria did not find enough documents in these newspapers.

3.2 Data Cleaning and Wrangling

Each document was separated into the news article and the readers' comments. For each article, we removed all parts that were not part of the article itself (e.g., advertisements, links to other articles). Meta-data about the article was kept in a separate database and contained the article's title, author(s), publication date, newspaper, type of newspaper (i.e., local/state/national), number of readers' comments, and search terms that led to finding the article. As there were only 165 articles following 9 different formats, cleaning of the article and preparation of the database was done manually, rather than investing in developing cleaning scripts tailored to each newspapers' format.

In contrast to the articles, having 3,864 comments made it necessary to process them using scripts. Our scripts can be accessed online at <https://osf.io/3x6av/>. Writing such cleaning scripts can be time consuming, partly because of the wide differences in functionality and formats across newspapers (Table 1). For example, most allowed users to 'like'/'recommend' a comment but only two allowed users to 'dislike' a comment. Comments sometimes include conversations, but tracking who a user was answering heavily depended on how it was managed by a newspaper's website. Most had a straightforward structure (e.g., indenting answers using spaces, or embedding an answer within the HTML block of a comment) but some had no structure and it was up to the users to correctly write @name at the beginning of their comment (thus leaving room for errors).

¹ The *Daily Californian* reports a press run of 10,000, while the *East Bay Express* reports a press run in Berkeley of 13,442. The average number of persons who read a copy (i.e., the pass-along rate) is estimated at 2.3 by the National Newspaper Association. As Berkeley had an estimated 116,768 inhabitants in 2013 according to the US Census Bureau, these two newspapers may be read by approximately 20% and 26% of the population.

² The combined circulation and on-line viewing data reported 20% of the market in Alameda county (where Berkeley is located) as well as 12% in San Francisco.

Table 1. Differences in comments across newspapers

Newspaper	Comment functionality				Comment structure	Encoding
	Likes	Dislikes	Reply	Full date		
Contra Costa Times	No	No	Yes	No	Set spaces	Text
Daily Californian	Yes	No	Yes	No	Set spaces	Text
East Bay Express	Yes	Yes	No	Yes	N/A	Text
Los Angeles Times	Yes	Yes	Yes	No	Depends on people’s use of @	Text
New York Times	Yes	No	Yes	Yes	HTML Structure	HTML
San Francisco C.	Yes	No	Yes	No	Depends on people’s use of @	Text
USA Today	Yes	No	Yes	Yes	Set spaces	Text
Wall Street Journal	No	No	Yes	Yes	HTML Structure	HTML
Washington Post	Yes	No	Yes	Yes	HTML Structure	HTML

Newspapers’ formats were sufficiently different that we recommend taking this into account when writing data collection scripts (e.g., by driving Selenium from Python): some formats can best be cleaned when copy/pasted from the user display, while for others the HTML page is best.

4 Applying Text Analytics to the Corpus

4.1 Solutions Most Readily Available to Policymakers

This section shows what policymakers *could* typically have access to in order to analyze news articles and associated reader comments. Thus, we focus on well-established text analytics software such as **Jigsaw** or **IN-SPIRE**; more functionalities could be obtained using newer, more specialized, or research software (e.g., **Luminoso** from the MIT Media Lab [29] or **TopicNets** [30] from the University of California). Additional examples can be found in [31].

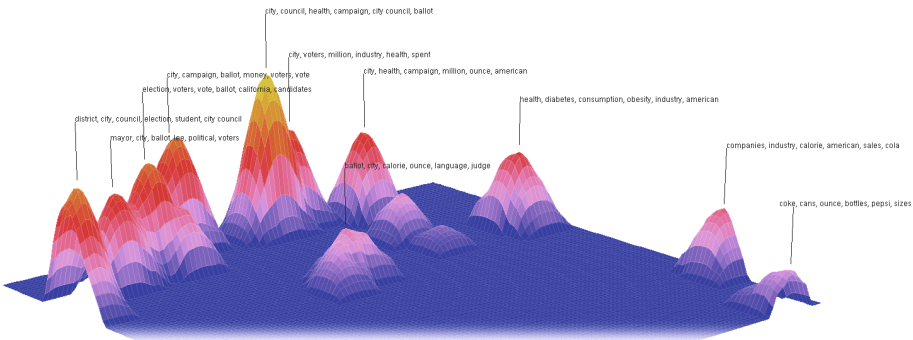


Fig. 3. Using the Galaxy view from **IN-SPIRE** [32] on the articles.

Documents are typically coded for themes. Both *Jigsaw* or *IN-SPIRE* allow themes to emerge (through clustering), as illustrated in Fig. 3. The height of each theme indicates the number of documents in it, while the words above the theme show the main keywords used by the algorithm to define that theme. The distance between themes indicates how they relate. In this example, it is immediately apparent that there were three broad categories: elections and ballots (left), health (center), company regulations (right). The specific themes within the last category include company sales (which may get impacted by the tax) and changes in company sizes (to compensate for the tax). The software allows correlations between themes or how they change over time (available as supplemental material at <https://osf.io/3x6av/>), which are other typical tasks for qualitative analysis.

When trying to assess public opinion, one may seek to examine the context in which specific words are used. An example of a motivating question would be: ‘what do people say about *tax*?’ One way to do this is to build a word tree using *Jigsaw*. Figure 4 shows such a word tree using readers’ comments, and several of the main arguments already appear: some consumers see it as an attack against their freedom to eat a wide range of foods (e.g., a ‘regressive sin tax’ along the lines of taxing cookies or ‘everything that can kill you’), have doubts about the use of proceeds (e.g., ‘revenue to fund other projects or even their own generous pay raises’), fear a disproportional impact on the poor, or even on jobs (‘corporations absorb the hit and reduce jobs’). A similar word tree built on the news articles (provided at <https://osf.io/3x6av/>), rather than the readers’ comments, depicts a different picture with benefits on childhood obesity and funding health programs more prominently featured.

Finally, a policymaker may be interested in knowing who is behind certain arguments, or how organizations are associated. This can be achieved by entity

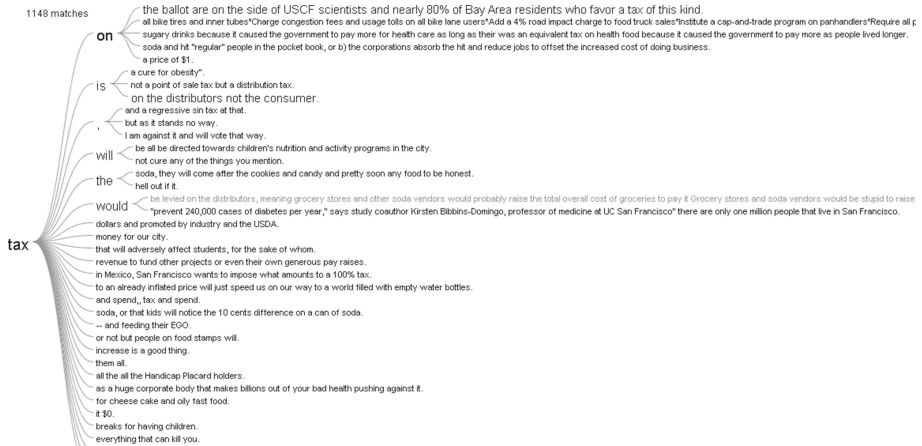


Fig. 4. Finding readers’ arguments that followed the word ‘tax’ using *Jigsaw*. The full-sized figure is available at <https://osf.io/3x6av/> together with the wordtree based on the articles.

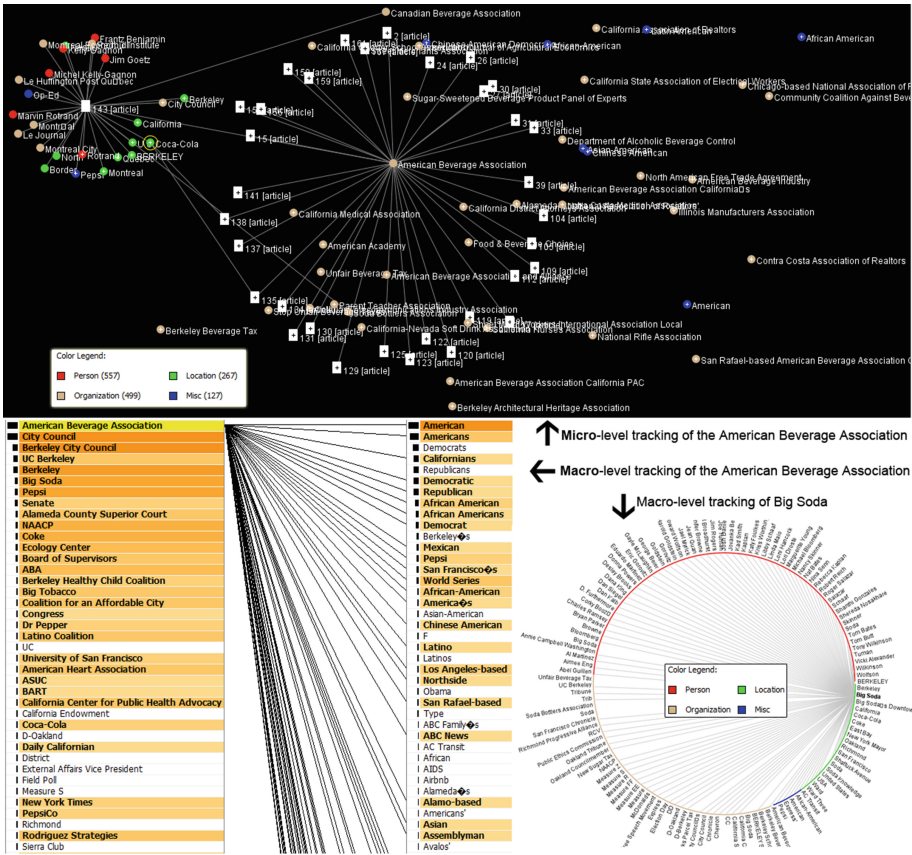


Fig. 5. Using entity tracking in Jigsaw on the articles.

tracking. Jigsaw automatically processes the documents to identify organizations, persons, locations, and other types of entities. Entity tracking can be done at the micro-level by following links (displayed as lines in Fig. 5 top): for example, one can start with the American Beverage Association, pick a document in which it is mentioned, and see what else is being mentioned. Alternatively, it can be done at the macro-level by finding the entities that co-appear the most with the American Beverage Association (e.g., showing that African-Americans are particularly co-mentioned), across all documents (Fig. 5 bottom left).

4.2 Advanced Solutions

The previous section emphasized techniques that policymakers could access using off-the-shelf software. There is still a variety of analyses that can be useful but may be less accessible to policymakers through current software. We provide additional examples of quantitative analyses on readers' comments at

<https://osf.io/3x6av/>. Text summarization is of particular interest when policy-makers are faced with a large corpus that they need to quickly condense.

There are two broad ways of summarizing a text: extraction (also known as the shallow technique) and abstraction (also known as the deep technique). Extraction uses the words and phrases of the actual text and applies smoothing techniques to address any incoherence. Abstraction may not contain the explicit words of the text. As an example of extraction, we implemented the (graph-based ranking) LexRank algorithm [33]. One issue with this algorithm is that policy-makers cannot simply pass it the text and get a summary: they need to choose the value of a parameter, which has a large impact on results. For example, one value can produce an irrelevant summary (“the issue with Fructose is way it is metabolized only by the liver”) while another value produces a very relevant one (“the proposed law benefits San Francisco bureaucrats like Scott Wiener who would like to get their hands on that expected \$30 million a year by taxing”). By carefully choosing the right value, it is possible to generate summaries and compare how they change depending on the source (news vs readers’ comments) or time (pre-ballot, post-ballot, etc.). The temporal difference is clear. For example, the readers’ comments post-implementation are summarized as “While some local businesses have felt the effects of the citywide ordinance, sales of sugary drinks sold on campus have not and will not be affected because the UC system is not bound by city laws.” In contrast, their comments pre-implementation were (perhaps sarcastically) summarized as “If this passes, please continue on and tax red meats, ice cream, donuts, fast food [etc.]”. The difference between the article summary and the readers’ comments summary echoes observations from word trees (Fig. 4).

5 Discussion

In this paper, we exemplified how current methods and software in text analytics could answer questions of interest to policymakers. We used SSB taxes as a guiding example to detail the process of collecting, cleaning, and analyzing data. We emphasized software that policymakers could directly use, while highlighting that there are other relevant analyses such as text summarization.

As we noted in the creation of the corpus, preparing the data is time consuming. Thus, one should not simplify text analytics as being immediate while qualitative analyses takes time: *both* approaches need time and manpower (particularly in the set-up phase) but they do not scale the same way. In text analytics, time is spent in writing scripts (e.g., to collect or clean data), and the time it takes is proportional to the number of different data sources (e.g., one has to decode the format specific to each newspaper). Passed this set-up cost, the cost of analyzing one additional article is negligible. In qualitative analyses, substantial time would be spent in developing the coding framework but analyzing each additional article will also require a small amount of additional time. As an example, in the supervised approach used by Hamad and colleagues, three people were involved in manually coding 354 articles to set-up the system, but it was then able to process 14,302 articles within a few days [16].

The off-the-shelf solutions that we presented have in common that they see all documents as being equally relevant. However, this is not the case in reality. For example, the *San Francisco Chronicle* (which had the most articles per our search criteria) also had articles of which the SSB tax was not the subject; instead, the tax may have been briefly mentioned as part of a subject's past records of political endorsements. This could lead to finding irrelevant themes or entities that are not truly connected. Consequently, a more accurate analysis would have to ensure that only relevant parts of the article (if any) are used. Similarly, when assessing public opinions via readers' comments, we need to ensure that the article they comment on strongly relates to the SSB tax. As the software capabilities evolve over time [34], such issues of relevance should gradually be addressed.

While the emphasis of this paper was on the generic process needed to perform text analytics, and on the type of questions that can be addressed, we note several limitations affecting the results specific to the guiding example. First, our search procedure cannot claim to have found all articles relevant to the SSB tax debate in California. While it is common to use only one database for text analytics (e.g., [16]), we searched within the LexisNexis database as well as daily newspapers with a large readership either locally or nationally, and the top 5 English-language newspapers outside the USA. This procedure is skewed towards large newspapers, and could be complemented by other online databases such as Access World News (<http://infoweb.newsbank.com>). In addition, using the largest newspapers does not guarantee that their articles have been at the core of the debates. The Topsy database can be used to find such 'hot' articles, by identifying the ones that are linked to retweets on Twitter. Using this procedure to identify articles has recently been shown to lead to different results in terms of sentiment or themes [15], although further research is needed to identify the procedure that selects the articles most representative of public opinions.

References

1. Flegal, K., Carroll, M., Ogden, C., Curtin, L.: Prevalence and trends in obesity among us adults, 1999–2008. *JAMA* **303**, 235–241 (2010)
2. Wang, Y., Beydoun, M., Liang, L., Caballero, B., Kumanyika, S.: Will all americans become overweight or obese? estimating the progression and cost of the us obesity epidemic. *Obesity* **16**, 2323–2330 (2008)
3. Zhang, D., Giabbanelli, P.J., et al.: Impact of different policies on unhealthy dietary behaviors in an urban adult population: An agent-based simulation model. *Am. J. Public Health* **104**(7), 1217–1222 (2014)
4. Giabbanelli, P.J., Jackson, P.J., Finegood, D.T.: Modelling the joint effect of social determinants and peers on obesity among canadian adults. In: *Theories and Simulations of Complex Social Systems*, pp. 145–160 (2014)
5. Verigin, T., Giabbanelli, P.J., Davidsen, P.: Supporting a systems approach to healthy weight interventions in british columbia by modeling weight and well-being. In: *Proceedings of the 2016 Annual Simulation Symposium (ANSS)* (2016)
6. Mytton, O., Clarke, D., Rayner, M.: Taxing unhealthy food and drinks to improve health. *BMJ* **344**, e2931 (2012)

7. Powell, L., Chaloupka, F.: Food prices and obesity: evidence and policy implications for taxes and subsidies. *Milbank Q* **87**(1), 229–257 (2009)
8. Brownell, K., Farley, T., et al.: The public health and economic benefits of taxing sugar-sweetened beverages. *NEJM* **361**, 1599–1605 (2009)
9. An, R.: Effectiveness of subsidies in promoting healthy food purchases and consumption: a review of field experiments. *Public Health Nutr.* **16**(7), 1215–1228 (2013)
10. Eyles, H., Mhurchu, C.N., Nghiem, N., Blakely, T.: Food pricing strategies, population diets, and non-communicable disease: A systematic review of simulation studies. *PLoS Med* **9**(12), e1001353 (2012)
11. Briggs, A., Mytton, O., Kehlbacher, A., Tiffin, R., Rayner, M., Scarborough, P.: Overall and income specific effect on prevalence of overweight and obesity of 20% sugar sweetened drink tax in uk: econometric and comparative risk assessment modelling study. *BMJ* **347**, f6189 (2013)
12. Pope, T.M.: Limiting liberty to prevent obesity: Justifiability of strong hard paternalism in public health regulation. *Conn. Law Rev.* **46**, 1859–1876 (2013)
13. Rivard, C., Smith, D., McCann, S.E., Hyland, A.: Taxing sugar-sweetened beverages: a survey of knowledge, attitudes and behaviours. *Publ. Health Nutr.* **15**(8), 1355–1361 (2012)
14. Nixon, L., Mejia, P., Cheyne, A., Dorfman, L.: Big soda's long shadow: news coverage of local proposals to tax sugar-sweetened beverages in richmond, el monte and telluride. *Crit. Publ. Health* **25**(3), 333–347 (2015)
15. Mahoney, L., Tang, T., et al.: The digital distribution of public health news surrounding the human papillomavirus vaccination: a longitudinal infodemiology study. *JMIR Publ. Health Surveill.* **1**(1), e2 (2015)
16. Hamad, R., Pomeranz, J., Siddiqi, A., Basu, S.: Large-scale automated analysis of news media: a novel computational method for obesity policy research. *Obesity* **23**, 296–300 (2015)
17. Eysenbach, G.: Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. *J. Med. Internet Res.* **11**(1), e11 (2009)
18. Salathe, M., Khandelwal, S.: Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLoS Comput. Biol.* **7**(10), e1002199 (2011)
19. Bodnar, T., Salathe, M.: Validating models for disease detection using twitter. In: *Proceedings of the 22nd International Conference on World Wide Web Companion*, pp. 699–702 (2013)
20. Ji, X., Chun, S.A., Geller, J.: Monitoring public health concerns using twitter sentiment classifications. In: *IEEE International Conference on Healthcare Informatics (ICHI) 2013*, pp. 335–344 (2013)
21. Dredze, M., Cheng, R., Paul, M., Broniatowski, D.: Healthtweets.org: A platform for public health surveillance using twitter. In: *AAAI Workshop on the World Wide Web and Public Health Intelligence* (2014)
22. The Media Insight Project: *The personal news cycle* (2013)
23. Sasseen, J., Olmstead, K., Mitchell, A.: Digital: as mobile grows rapidly, the pressures of news intensify. *Pew Center Project for Excellence in Journalism* (2013)
24. Dolsak, N., Houston, K.: Newspaper coverage and climate change legislative activity across us states. *Global Policy* **5**(3), 286–297 (2014)
25. Niederdeppe, J., Gollust, S., Jarlenski, M., Nathanson, A., Barry, C.: News coverage of sugar-sweetened beverage taxes: pro- and antitax arguments in public discourse. *Am. J. Publ. Health* **103**, e92–e98 (2013)

26. Stoddard, P., van Erp, B., Induni, M., Reedy, A., Broderick, B.: Sugar-sweetened beverage tax acceptability in the u.s. versus foreign-born population: Results from a large northern california county. In: 142nd Annual Meeting & Expo of the American Public Health Association (APHA), p. 303809 (2014)
27. The People of the City of Berkeley: Imposing a general tax on the distribution of sugar-sweetened beverage products. Ordinance (2014)
28. Dinkelspiel, F.: A record \$3.6 million spent in berkeley campaigns, November 2014
29. Speer, R., Havasi, C., Treadway, N., Lieberman, H.: Visualizing common sense connections with luminoso. In: Proceedings of the First International Workshop on Intelligent Visual Interfaces for Text Analysis, pp. 9–12 (2010)
30. Gretarsson, B., O'Donovan, J., et al.: Topicnets: Visual analysis of large text corpora with topic modeling. *ACM Trans. Intell. Syst. Technol.* **3**(2), 23 (2014)
31. Giabbanelli, P.J., Jackson, P.J.: Using visual analytics to support the integration of expert knowledge in the design of medical models and simulations. In: Proceedings of the 8th Workshop on Biomedical and Bioinformatics Challenges for Computer Science (2015)
32. Wong, P., Hetzler, B., et al.: In-spire infovis 2004 contest entry. In: IEEE Symposium on Information Visualization (2004)
33. Otterbacher, J., Gunes, E., et al.: Biased lexrank: passage retrieval using random walks with question-based priors. *Inf. Process. Manage.* **1**, 42–54 (2009)
34. Gorg, C., Liu, Z., Stasko, J.: Reflections of the evolution of the jigsaw visual analytics system. *Inf. Vis.* **13**(4), 336–345 (2014)