Integrated Economy and Society: Diversity, Creativity, and Technology
16–18 May 2018
Naples • Italy
make learn TIIM
Management, Knowledge and Learning International Conference 2018
Technology, Innovation and Industrial Management

TRACKING SOCIAL PERCEPTION ON HEALTHCARE SERVICE QUALITY USING SOCIAL MEDIA

Hong Joo Lee
Catholic University of Korea, South Korea
hongjoo@catholic.ac.kr

Minsik Lee
Yonsei University, South Korea
salvia0413@yonsei.ac.kr

Habin Lee
Brunel University London, United Kingdom
habin.lee@brunel.ac.uk

Abstract :
Despite the opportunities and demands to use social media to support public policy-making processes, a systematic approach to reflect social media sentiments in policy making processes is yet to be proposed in the literature. This paper suggests a systematic method to assign tweets into one of SERVQUAL dimensions to identify sentiments and to track perceived service quality for policy makers in national health services (NHS). In this study, we propose a methodology to (1) identify more reliable topic sets through repeated LDA and clustering and (2) classify tweets with the topics based on an existing theory in service quality. To show the applicability of our method, we selected healthcare as our target area and picked the NHS of U.K. for measuring the service quality of public policy. We collected tweets about NHS for about 4 years and created dictionaries related to the domain of healthcare with user reviews on hospitals and general practitioners in U.K. We applied the suggested methodology to track social perceptions and compared the applicability among different methods.

Keywords: Social perceptions, SERVQUAL, Healthcare, NHS, Sentiment analysis, Topic modeling

1. INTRODUCTION

The quality management of healthcare service is one of the major policy issues of many central governments. Traditional methods to evaluate the quality of healthcare services requires significant time and cost for monitoring mechanisms primarily based on survey from relevant stakeholders (Greaves et al., 2014). Social media provides healthcare policy makers with an opportunity to collect and analyze data on the quality of national healthcare services but also poses challenges due to the nature of data including huge volume, cleaning, natural language processing, and relating contents with constructs of a well-established theoretical quality model (King et al., 2013). This paper presents a novel method to translate social media sentiment into SERVQUAL that is widely used for measuring health care service quality overcoming the challenges. The translated SERVQUAL allows real-time monitoring of citizens satisfaction of health care services and can be used as an early-warning indicator.

With the proliferation of social media, extensive amount of online user-generated content or word of mouth have been produced and the surge of its volume is getting accelerated. Business companies are on the lookout to understand and monitor social perceptions on their brands and services (Swani, Brown, & Milne, 2014). Firms are doing this by collecting and analyzing user reviews and similar digital traces, including social media, to understand how they are perceived by their communities (Chen, Kou, Shang, & Chen, 2015; Goh, Heng, & Lin, 2013). In addition, social perceptions and moods of people on certain events can be collected through social media and utilized to predict outcomes of collective behavior.

Many government agencies are experimenting electronic participation with the use of social media and they hope to increase government transparency, participation, and collaboration (Mergel, 2013)[6]. Policy makers in public authorities are facing challenges around the development and implementation of evidence-based policy making (Habin Lee, Tsohou, & Choi, 2017). This citizen-inclusive approach is increasingly favored by legislators and ever more robust underpinned by significant amounts of data sets often harvested and available through the internet and social media. Patient experiences shared through social media or online communities include real people talking about what they have experienced and how they feel, in their own words (De Silva, 2013). By listening to online voices, public service design can be significantly improved (Criado, Sandoval-Almazan, & Gil-Garcia, 2013). However, due to the large volume of online voices available, it is a challenge to measure social perceptions manually. Nonetheless, there is a big benefit to unravelling the value contained in big data to improve existing public services. We need intelligent augmentation for measuring and improving public services based on computational intelligence. Tracking the service quality of National Health Service (NHS) with social media can help us identify dimensions to be investigated for further improvements reducing the number of survey that requires more costs and time. (Greaves et al., 2012) provides an evidence that patient web-based ratings on service experience are associated with hospital ratings derived from a national paper-based patient survey. The analysis of patient stories can be integrated with more quantitative surveys or other technical approaches to provide a comprehensive picture (De Silva, 2013). Despite the opportunities to use social media to support public policy-making processes, existing empirical studies on compiling social media sentiments into service quality measurements for public policy have some limitations in analyzing data and unraveling meanings.

In this study, we propose a systematic method to track social sentiments on service quality using social media data. To show the applicability of our proposed method, we select healthcare as our target area and pick NHS of U.K. to measure service quality of public policy. Since social media data have lots of noisy data, we applied Doc2Vec and machine learning algorithms to identify relevant data on service experience. With the relevant data, we utilize Latent Dirichlet Allocation (LDA) for topic modeling and try to get more reliable topic sets by repeating LDA and clustering topic sets for lessening subjective bias. We classify tweets with the topics based on an existing framework in service quality – SERVQUAL (Parasuraman, Zeithaml, & Berry, 1985; 1988). SERVQUAL has been widely used for assessing the quality of health services in the literature as its five service dimensions provide policy makers with specific implications for intervention (Al-Borie & Sheikh Damanhouri, 2013; Altuntas, Dereli, & Yilmaz, 2012; Purcărea, Gheorghe, & Petrescu, 2013; Teshnizi, Aghamolaei, Kahnouji, Teshnizi, & Ghani, 2018). Words belonging to each topic are matched with terms from the pre-classified data and survey questionnaire for each construct of SERVQUAL. In doing so, we can measure similarity values between a tweet and the topic sets. These similarity values are input data for machine learning algorithms to classify a tweet into one of SERVQUAL dimensions and other. Then, a dictionary for the healthcare domain is built to measure sentiments of each construct of service quality. Several dictionaries are built through different methods and their performances are compared to choose the most appropriate one. We collect tweets about NHS for about 4 years, patient reviews on hospitals, and reviews on general practitioners (GPs) to build a sentiment dictionary. In addition, we collect survey questionnaires from studies related to healthcare service quality to match the constructs of SERVQUAL to the topics. As a result, we track social sentiments relating to each dimension of service quality of NHS systematically.

2. CONCEPTUAL BACKGROUND

Service quality is considered as one of the critical success factors in an organization's effort to differentiate itself from its competitors (Ladhari, 2009). Zeithaml (2000) summarized the evidences of the relationship between service quality and profits in six categories: direct (increase profitability), offensive (obtaining new customers), defensive (customer retention), increase purchase intention, distinguishing customer segments, and identifying the key drivers of service quality. Thus, diverse perspectives of measuring service quality were suggested to track the perceptions of customers on their services (Cronin & Taylor, 1992; Grönroos, 1984; U. Lehtinen & Lehtinen, 1982; Parasuraman et al., 1985). SERVQUAL is the best known and most commonly used measure (Ladhari, 2009) and seminal contribution for measuring service quality (Palese & Piccoli, 2016). After several updates, SERVQUAL consolidated into five dimensions: tangibles (the appearance of physical facilities, equipment, and personnel); reliability (the ability to perform the promised service reliably and accurately); responsiveness (the willingness to help customers and provide prompt service); assurance (the knowledge and courtesy of employees and their ability to inspire trust and confidence); and empathy (the level of caring and individualized attention the firm provides to its customers). SERVQUAL has been utilized to measure service quality in various service settings, industries, and countries (Ladhari, 2009). In healthcare, there has been many studies on measuring service quality through SERVQUAL. Some studies proposed the extension of SERVQUAL for healthcare industry (Lee, Delene, Bunda, & Kim, 2000) and proposed different measures (Park, Kim, Park, & Agarwal, 2016; Sofaer & Firminger, 2005). There have been debates about the various aspects of SERVQUAL including validity, reliability, and applicability (Ladhari, 2009). One of our contributions is to utilize the SERVQUAL framework as a guide to train classifier of machine learning algorithms with a labeled data and survey questionnaire and classify data from social media.

3. METHODOLOGY

3.1. Data

For this study, we suggest a systematic method of tracking social perceptions on service quality from social media data. Our target public service is the NHS of the U.K. since we can collect sufficient social media data and relevant patient reviews from its website, NHS Choices (http://www.nhs.uk/). We collect 50,716 tweets that contain NHS in their posts from January 1, 2013 to October 31, 2016. We use tweets uploaded in the U.K. and written in English. We pre-process all tweets by removing URLs, numbers, punctuation marks, stop words, and other languages. Then we extract all words from the tweets and stem the words since one word can have different forms (e.g., pay and paid). We build a term-document matrix with the stemmed words and remove terms with a sparsity greater than 0.9999 to reduce complexity. In addition, the term NHS is removed from the matrix since every tweet contains it therefore meaningless. The cell values of the term-document matrix are term frequencies.

3.2. Steps

Step 1: Excluding non-relevant tweets

Most of the tweets are not relevant to service quality for patients of hospitals or GPs but arguing about healthcare reform, political discussions, NHS budget and so on. Thus, we need to identify tweets about service quality for the further analysis. We apply a machine learning approach to identifying non-relevant tweets. Since we need a training dataset, two graduate school students who are aware of the concepts of SERVQUAL are recruited to classify randomly selected 600 tweets. The purpose of this study and the dimensions of SERVQUAL are introduced to the recruited raters. Two raters first individually classify the tweets into one of SERVQUAL dimensions and then discuss together to agree on their classification results. The agreed classifications are used as training and test data in the following steps. In the data set, there are more tweets related to Reliability and Tangibles dimensions than other SERVQUAL dimensions though the largest number of the tweets are classified as other.
We also use the survey items of SEVRQUAL studies for the training and test datasets to expand related word lists in corresponding dimensions. SERVQUAL has been adopted for measuring patient service

satisfaction in many healthcare studies and the approach of utilizing survey items of SERVQUAL is also adopted in text mining studies (Kim, 2015). We collect survey items of SERVQUAL in healthcare and pre-process them as we did for the tweets in our study.
This study uses Doc2Vec to represent words in semantic spaces and applies diverse machine learning algorithms to classify the tweets into either service-experience related or not. We conduct 5-fold cross validation with the 600 tweets and the survey items. Table 1 shows the average accuracy values of the classifications and neural network method shows the highest accuracy value. We use the neural network predictor to classify all tweets into relevant or non-relevant tweets and only the relevant tweets are used in the next step. By doing this, noises from non-relevant tweets can be reduced for performing topic modeling.

Table 1: Accuracy of classifications

Methods	Accuracy	F1
Support Vector Machine	0.5850	0.5879
Naïve Bayes	0.6334	0.6292
Decision tree	0.5614	0.5622
K-Nearest Neighbours	0.6182	0.5173
Logistic regression	0.6066	0.6062
Neural Network	0.6482	0.6424

Step 2: Repeated Topic Modeling and Clustering

The assumption of LDA is that “documents are represented as a random mixture over latent topics - where each topic is characterized by a distribution over words” (Blei, Ng, & Jordan, 2003) and that LDA extracts latent topics among documents. LDA is based on Gibbs sampling which attempts to collect samples from the posterior to approximate it with an empirical distribution. Random selection of the distribution over topics and vocabulary is required to perform LDA. Due to these random selection procedures, the results of LDA vary in different implementations. Researchers choose one set of topics which can explain their data well after repeated trials. (Palese & Piccoli, 2016) use seed words for the five constructs of SERVQUAL to identify corresponding topics through LDA – called weakly supervised LDA (Lin, Lu, Xiong, & Zhu, 2012). Though they use the seed words to guide their topic selection, it is still grounded on sampling-based algorithm and the selection of the words is done manually.
Unlike the weakly supervised LDA approach, this study runs LDA multiple times and applies clustering to the results of the multiple LDAs for extracting more reliable results of topic modeling and for reducing human interventions. This study uses the tweets predicted as relevant in step 1 and performs LDA to have 30 topics with 30 words per topic at one trial and repeat it 1,000 times with varying delta values from 0.1 to 10 (Grün & Hornik, 2011). Number of words per topic is usually selected from 20 to 30 and we chose large enough number topics for applying clustering. Delta values are used for initializing Gibbs sampling procedures in functions of the topicmodel package (Grün & Hornik, 2011). We used default values of topicmodel packages for other settings of topic modeling.
As a result of running LDA once, we get 30 topics with thirty words belonging to a topic and their probabilities. With a total of 30,000 topics from 1,000 repetitions, we apply clustering algorithms to have similar subsets of topics by calculating the distances of topics with the probabilities of words in a topic. Hierarchical clustering is applied to have 30 topic clusters.

Step 3: Dimension classification

We, then, assign each tweet to one of SERVQUAL constructs or to other dimension. Similarity values of each tweet presented in the term-document matrix to the 30 topic clusters described in Step 2 are measured with the Jaccard index. The Jaccard coefficient calculates similarity between finite sets and is defined as the size of the intersection divided by the size of the union of the comparing sets. These 30 similarity values of a tweet to the 30 topic clusters are input values for machine learning algorithms to classify each tweet into one of SERVQUAL or other dimensions. We apply diverse machine learning algorithms and conduct 5-fold cross validation with the labeled 600 tweets as explained in Step 1. Table 2 shows the average accuracy values of the diverse algorithms and again neural network shows the highest accuracy value.

Table 2: Accuracy of classifications into SERVQUAL

Methods	Accuracy	F1
Support Vector Machine	0.6301	0.4993
Decision tree	0.4416	0.4516
K-Nearest Neighbours	0.3966	0.4241
Neural Network	0.6317	0.5120

Step 4: Dictionary building

To use the dictionary-based matching approach for measuring sentiments of a statement, we need a dictionary which has sentiment values of words. AFINN (Nielsen, 2011) assigns words with negative scores for negative sentiment and positive scores for positive sentiment. Bing (Liu, 2012) and NRC (Mohammad & Turney, 2013) categorize words in a binary fashion into positive or negative category. Since the widely used dictionaries are for general purposes, we need to build our own dictionary for healthcare service domain.

To build our own dictionary (NHSdict), we collect user reviews on medical services from NHS Choices. We collect 2,163 reviews from 136 hospitals by randomly selecting reviews from the website. Each review has a star rating from 1 to 5. We classify reviews with 1 or 2 stars as negative and those with 4 or 5 stars as positive. Since there is small number of positive reviews, we use 408 negative reviews and 408 positive reviews to calculate the effect of a word on the classification of its review. We pre-process the reviews as we did for the tweets in this study. We apply logistic, lasso, ridge, and elastic regression (James, Witten, Hastie, & Tibshirani, 2013) to build a model for classifying reviews into positive or negative. The independent variables of the regressions are words from the reviews and the coefficient values are their sentiment scores.

We perform 10-fold cross validation on 916 reviews to compare the accuracy of the sentiment scores from the regressions. We simply summate sentiment scores of words which are contained in a review. Then we classify the review as negative if the summated score is less than zero, otherwise we classify it as a positive review. The classifications of the sentiment scores from the ridge regression outperform those from other regressions. Thus, we use the sentiment scores of words from the ridge regression for NHSdict.

Step 5: Sentiment analysis

We measure sentiments of tweets by utilizing AFINN and NHSdict. We simply summate the sentiment scores of words contained in a tweet and then take the average of the sentiment scores according to the SERVQUAL constructs. Though AFINN assigned sentiment scores range between -5 to 5, the sentiment scores of NHSdict range between 0.1 to -0.1. To compare sentiment scores using two dictionaries, we standardize the sentiment scores by transforming the scores to z distribution. Then we incorporate AFINN and NHSdict together to expand the dictionaries since the number of words in both are limited. The sentiment scores of words in the incorporated dictionary are from the standardized scores described above. For words present in both dictionaries, we choose the sentiment score from NHSdict. We apply all three dictionaries (AFINN, NHSdict, and the integrated dictionary) to measure the sentiments of the dimensions of the service quality.

4. FINDINGS AND DISCUSSIONS

While most tweets are classified as neutral by AFINN dictionary, most tweets are classified as negative by NHSdict. We randomly select 3,822 tweets which are not classified as neutral by both dictionaries. Among 3,822 tweets, 2,621 tweets (69%) are classified as the same sentiment by both dictionaries. AFINN has 339 words and NHSdict has 3,998 words. The number of overlapped words in the both dictionaries is 31.

We measure sentiments of tweets as described in the above section and track the sentiments according to the SERVQUAL constructs. Table 3 shows average sentiments scores of the 600 tweets explained in Step 1 measuring using the three dictionaries we built in Step 4 and 5. The sentiments scores are standardized to compare the corresponding values of each dimension among the three dictionaries.

The sentiment scores for Empathy are all positive and the other scores for other SERVQUAL dimensions were all negative except Tangibles with the integrated dictionary. Among the negative dimensions, Responsiveness and Assurance have the lower sentiment scores, and the sentiment scores for Reliability and Tangibles are near zero in NHSdict and the integrated dictionary. Based on these scores, people complain about Responsiveness and Assurance more than other dimensions.

Table 3: Sentiment scores

Dimensions	Average sentiment score – AFINN (standard deviation)	Average sentiment score – NHSdict (standard deviation)	Average sentiment score – Integrated dictionary (standard deviation)
Assurance	-0.6718 (1.1667)	-0.1221 (0.6807)	-0.7446 (1.1672)
Empathy	0.2789 (0.8267)	0.3474 (1.0991)	0.6742 (1.3644)
Reliability	-0.1799 (1.0508)	-0.0278 (0.9977)	-0.0868 (1.4030)
Tangibles	-0.0731 (0.9211)	-0.0862 (1.0704)	0.0164 (1.2430)
Responsiveness	-0.0236 (0.8581)	-0.2366 (0.7769)	-0.6478 (1.0594)
Other	0.0674 (1.0096)	0.0354 (1.0098)	0.0705 (0.7444)

Ham, Berwick, & Dixon (2016) suggests a coherent and integrated strategy for improving quality of NHS including systematic, transparent measurement and reporting of progress in improving quality, and commitment to listening to and learning from the experiences of patients. Data on patient experience, systematic methods to analyze and feedback to improve patient care are needed to accomplish the strategies. Currently, NHS monitors over 300 indicators to measure quality of health services including mortality rate, diagnostic waiting times, results of inpatient surveys, NHS staff survey and clinical audits. NHS Patient experience on social media can be input data for improving healthcare quality and the systematic way of unravelling their experience into dimensions of service quality in this paper can provide a way of analyzing the data. A nation-wide implementation of surveys on patient satisfaction of NHS implies a significant cost therefore the frequency and timing of a survey are important decision-making problems for policy makers in NHS. Also, value for money of implementing a nation-wide survey is criticized by practitioners due to the lack of new information from the survey results and they prefer patient responses to open-ended questions that provide more interesting insights. The proposed approach in this study allows policy makers to use social media sentiments on health care services as an indicator for deciding the frequency and timing of more expensive nation-wide questionnaire survey. As SERVQUAL is widely used to design questionnaire design for health care services, the social media sentiments, which are translated into SERVQUAL dimensions is a useful indicator to catch signals from the real patients. Also, the follow-up analysis to identify the frequency and keyness words that determine the sentiment scores of each dimension help policy makers to design more specific interventions and resource allocations can be more effective by tacking those areas.
The results of this study can be incorporated with event analyses since we have the respective dates of the tweets collected. By doing so, it will be possible to monitor social perceptions on certain policies or actions by the government. The results can be input data for policy decision-making if certain words are linked with outcomes of possible policies. We can also simulate the effects of policies by increasing or decreasing those words and monitor the changes of social perceptions due to the policy introduction. Government social media professionals can gain useful insights by interpreting social media data for decision-making (Mergel, 2013). By listening to online voices systematically, we can identify social perceptions and reflect these opinions to make significant improvement in designing public service (Criado et al., 2013).

5. CONCLUSION

In this study, we proposed a systematic way to analyze and track social perceptions on the quality of public services. Noisy social media data were filtered out by applying Doc2Vec and machine learning algorithms. The latent topics in social media data can be extracted and interpreted with the words belonging to the topics. This paper provides a method of acquiring more reliable sets of topics and using the topics for classifying tweets into one of SERVQUAL dimensions and other. Though it requires repetitive steps and statistical analyses, it can reduce manual interventions and subjective selections in the analysis. We validated the performances of classifications and sentiment measuring with the training data we obtained and the reviews from NHS. Nonetheless, this paper shows an example of getting robust topic sets and utilizing the topics for unravelling meaning of tweets.

This study is limited to one keyword used to collect tweets regarding to medical services. Even though all citizens have medical services through the NHS in the U.K., expanding keyword sets to collect a more accurate social media data is needed. However, this does not affect the effectiveness of the methodology of the study to collect and track social sentiments. This approach can be easily applied with the expanded keyword sets.

REFERENCE LIST

1. Al-Borie, H. M., & Sheikh Damanhouri, A. M. (2013). Patients' satisfaction of service quality in Saudi hospitals: a SERVQUAL analysis. *International Journal of Health Care Quality Assurance*, *26*(1), 20–30.
2. Altuntas, S., Dereli, T., & Yilmaz, M. K. (2012). Multi-criteria decision making methods based weighted SERVQUAL scales to measure perceived service quality in hospitals: A case study from Turkey. *Total Quality Management & Business Excellence*, *23*(11-12), 1379–1395.
3. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.
4. Chen, K., Kou, G., Shang, J., & Chen, Y. (2015). Visualizing market structure through online product reviews: Integrate topic modeling, TOPSIS, and multi-dimensional scaling approaches. *Electronic Commerce Research and Applications*, *14*(1), 58–74.
5. Criado, J. I., Sandoval-Almazan, R., & Gil-Garcia, J. R. (2013). Government innovation through social media. *Government Information Quarterly*, *30*(4), 319–326. http://doi.org/10.1016/j.giq.2013.10.003
6. Cronin, J. J., Jr, & Taylor, S. A. (1992). Measuring service quality: a reexamination and extension. *Journal of Marketing*, *56*(3), 55–68.
7. De Silva, D. (2013). *Measuring patient experience*. Health Foundation.
8. Goh, K. Y., Heng, C.-S., & Lin, Z. (2013). Social Media Brand Community and Consumer Behavior: Quantifying the Relative Impact of User- and Marketer-Generated Content. *Information Systems Research*, *24*(1), 88–107. http://doi.org/10.1287/isre.1120.0469
9. Greaves, F., Laverty, A. A., Cano, D. R., Moilanen, K., Pulman, S., Darzi, A., & Millett, C. (2014). Tweets about hospital quality: a mixed methods study. *BMJ Quality & Safety*, *23*(10), 838–846.
10. Greaves, F., Pape, U. J., King, D., Darzi, A., Majeed, A., Wachter, R. M., & Millett, C. (2012). Associations between Internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. *BMJ Quality & Safety*, *21*(7), 600–605.
11. Grönroos, C. (1984). A service quality model and its marketing implications. *European Journal of Marketing*, *18*(4), 36–44.
12. Grün, B., & Hornik, K. (2011). topicmodels: An RPackage for Fitting Topic Models. *Journal of Statistical Software*, *40*(13), 1–30. http://doi.org/10.18637/jss.v040.i13
13. Ham, C., Berwick, D., & Dixon, J. (2016). *Improving quality in the English NHS: A strategy for action*. The King's Fund.
14. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112). Springer.
15. Kim, K. (2015). *Developing Theory-based Text mining Framework to Evaluate Service quality in the Context of Hotel Customers' Online Reviews*.
16. King, D., Ramirez-Cano, D., Greaves, F., Vlaev, I., Beales, S., & Darzi, A. (2013). Twitter and the health reforms in the English National Health Service. *Health Policy*, *110*(2), 291–297.
17. Ladhari, R. (2009). A review of twenty years of SERVQUAL research. *International Journal of Quality and Service Sciences*, *1*(2), 172–198.
18. Lee, Habin, Tsohou, A., & Choi, Y. (2017). Embedding persuasive features into policy issues: Implications to designing public participation processes. *Government Information Quarterly*, *34*(4), 591–600.
19. Lee, Hanjoon, Delene, L. M., Bunda, M. A., & Kim, C. (2000). Methods of measuring health-care service quality. *Journal of Business Research*, *48*(3), 233–246.
20. Lehtinen, U., & Lehtinen, J. R. (1982). Service quality: a study of quality dimensions. Service Management Institute.
21. Lin, Z., Lu, R., Xiong, Y., & Zhu, Y. (2012). Learning ontology automatically using topic model. Presented at the International Conference on Biomedical Engineering and Biotechnology.
22. Liu, B. (2012). Sentiment Analysis and Opinion Mining: Introduction and Survey. Morgan & Claypool Publishers.
23. Mergel, I. (2013). A framework for interpreting social media interactions in the public sector. *Government Information Quarterly*, *30*(4), 327–334.

24. Mohammad, S., & Turney, P. (2013). Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence*, *29*(3), 436–465.
25. Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. Presented at the the ESWC Workshop on Making Sense of Microposts. Retrieved from https://arxiv.org/abs/1103.2903
26. Palese, B., & Piccoli, G. (2016). Online Reviews as a Measure of Service Quality. Presented at the Pre-ICIS SIGDSAIFIP WG. Symposium, Dublin.
27. Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, *49*(4), 41–50.
28. Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality. *Journal of Retailing*, *64*(1), 12–40.
29. Park, G.-W., Kim, Y., Park, K., & Agarwal, A. (2016). Patient-centric quality assessment framework for healthcare services. *Technological Forecasting and Social Change*, *113*, 468–474.
30. Purcărea, V. L., Gheorghe, I. R., & Petrescu, C. M. (2013). The assessment of perceived service quality of public health care services in Romania using the SERVQUAL scale. *Procedia Economics and Finance*, *6*, 573–585.
31. Sofaer, S., & Firminger, K. (2005). Patient perceptions of the quality of health services. *Annual Review of Public Health*, *26*, 513–559.
32. Swani, K., Brown, B. P., & Milne, G. R. (2014). Should tweets differ for B2B and B2C? An analysis of Fortune 500 companies' Twitter communications. *Industrial Marketing Management*, *43*(5), 873–881.
33. Teshnizi, S. H., Aghamolaei, T., Kahnouji, K., Teshnizi, S. M. H., & Ghani, J. (2018). Assessing quality of health services with the SERVQUAL model in Iran. A systematic review and meta-analysis. *International Journal for Quality in Health Care*.