

Text Mining Narrative Survey Responses to Develop Engagement Scale Items

John M. Ford
U.S Merit Systems
Protection Board
john.ford@mspb.gov

Doug Nierle
U.S Merit Systems
Protection Board
doug.nierle@mspb.gov

J. Peter Leeds
U.S Merit Systems
Protection Board
peter.leeds@mspb.gov

Thomas A. Stetz
Hawaii Pacific University
tstetz@hpu.edu

Abstract

A sixteen-item employee engagement scale was supplemented with items developed from literature review, from related scales, and from text mining narrative responses to an open-ended question about improving employee performance. The text mining procedure is described and may be useful to other scale developers. Some items derived from text mining performed as well as those developed using traditional methods. Possible modifications and extensions of the method are suggested.

1. Introduction

This article briefly reviews the early development of an employee engagement scale, and then focuses on a text mining procedure used to develop additional questions for that scale. The primary objective of this project was to improve this scale to better gauge the engagement level of civil servants in the Federal workforce. This was done by accessing a heretofore untapped source, the opinions offered by survey respondents in response to an open-ended question about the best way to improve their performance. This seemed a productive approach because previous research [1, 2] had established a link between employee engagement and both organizational and individual performance.

A secondary objective was to investigate the usefulness of text analysis of open-ended question responses to identify question topics useful to expanding a deployed measurement scale. Current professional practice in scale development focuses on formal methods, such as appropriating questions from existing questionnaires, deriving new questions from findings in the professional literature, and conducting focus groups with representative groups of potential survey respondents [3, 4]. A common characteristic of such professionally respectable development

methods is the existence of some link to the population of interest.

Less often discussed are the informal methods, such as individual inspiration, collaborative discussion between individual researchers, and unstructured group discussions among survey stakeholders. Items generated from these processes have, at best, indirect links to the population of interest. Survey researchers can be minimally concerned about the informal origins of such items because they are ultimately subjected to field testing with this population and statistical analysis of their contribution to the scale being developed [5].

Item developers sometimes review responses to open-ended questions related to a scale topic as one informal method for developing new scale items. Such review is often unsystematic and limited by the large quantity of narrative responses. A text analytic procedure could introduce systematicity and manageability to this process, elevating a mostly informal method into a more formal one linked to the target population. This text mining procedure, perhaps expanded and improved upon following this initial exploration, may be useful to other survey researchers developing and improving survey-based measurement scales.

2. Engagement scale development

Academicians, management theorists, and organizations themselves have similar definitions of “employee engagement.” Most definitions center on the idea that employees who are engaged have some type of heightened connection to their work, their organization, or the people they work for or with [6]. The importance of this connection lies in the significant relationship between increased levels of employee engagement in Federal agencies and improved agency performance outcomes. Specifically, the U.S. Merit Systems Protection Board (MSPB) has found that in agencies where

more employees were engaged better program results were produced, employees used less sick leave, fewer employees filed equal employment opportunity complaints, and there were fewer cases of work-related injury or illness [1, 2].

For the past two decades, the MSPB has conducted periodic, Merit Principles Surveys of Federal employees to elicit perceptions of their jobs, work environment, supervisors, and agencies. MSPB constructed a scale to measure the engagement level of Federal employees using data from 2005 Merit Principles Survey (MPS 2005). The MPS 2005 was administered to 70,000 full-time, permanent, nonseasonal Federal employees randomly sampled from a population of approximately two million. It was administered during the late summer and early fall of 2005. Twenty-four Federal agencies participated, and a total of 36,926 employees completed the survey for a response rate of just over 50 percent [7].

The engagement scale was initially developed post-hoc using responses to several dozen five-option Agree/Disagree items from the MPS 2005. These items had been used through several previous iterations of the MPS to track employee attitudes toward their work settings. A factor analysis of these items yielded a single dominant factor for the employee engagement construct. Thirty items correlated highly with this engagement factor. The initial engagement scale included 11 of these items. Five additional items were included that did not correlate highly with the engagement factor, but represented aspects of engagement identified as important in a review of previous engagement research [1]. A more detailed account of this development process is available in a presentation delivered to the International Personnel Assessment Council [8].

The resulting 16-item scale covered six workplace climate factors that foster engagement. They are: 1) Pride in One's Work; 2) Quality of Leadership; 3) Opportunity to Perform Well; 4) Recognition; 5) Prospects for Future Growth; and 6) Positive Work Environment. One factor remained for which no items existed in the MPS 2005 item bank: Effort Beyond Minimum Required. The items comprising this scale are the first 16 listed in the Appendix to this article. They are all one-sentence statements to which survey participants respond using a standard 5-option agreement scale (Strongly Agree, Agree, Neither, Disagree, or Strongly Disagree).

The Cronbach's alpha for the post hoc scale assembled from the MPS 2005 data was .926, indicating a high internal consistency among the sixteen items [9]. The engagement scale was used to

examine the relationship between engagement and several measures of Federal agency success. Agencies with more engaged (e.g., higher-scoring) employees had better program results, used less sick leave, received fewer EEO complaints, and had fewer cases of work-related illness [1].

Plans to use the engagement scale in subsequent administrations of the Merit Principles Survey afforded the opportunity to improve the scale. There were at least two reasons to do so. First, the Effort Beyond Minimum Required dimension identified in the literature review had not been covered. It was also suspected that there may be additional aspects of Federal employee engagement that are not present in the private sector, where most of the previous engagement measures had been developed.

An expanded engagement scale was assembled which included eight additional items. Two items were written to cover the missing Effort Beyond Minimum Required factor from the engagement literature review [8]. The text mining of narrative survey responses described in the next section initially contributed 4 items. Also included were two reference items from MSPB's 20011-2013 research agenda [10] that focus on supervisor feedback. These two items were not part of the engagement scale and were intended as discriminant validity benchmarks [11]. We should expect correlations between these items and the engagement scale to be lower than the correlations of intended engagement items.

Items for the literature review and research agenda were developed using standard item writing methods [BBB, CCC]. MSPB researchers referred to engagement factor definitions, then wrote, reviewed, and revised items collaboratively. The remaining four items were conceived based on the results of text mining. Researchers referred to relevant open-ended survey responses, then wrote, reviewed, and revised items based on them. The following section describes how text mining was used to identify relevant open-ended responses that were the basis for these four additional scale items.

3. Text Mining and Item Writing

The procedure followed for text mining is generally consistent with general content analysis methods familiar to many social science researchers [12-14]. Our goal was to automate a number of data reduction and text processing tasks to make useful content patterns more apparent to scale developers [15]. This process has been used with other text data, for example to mine information from banks of multiple choice test questions [16], and from a cross-

agency database of performance appraisals [17]. In both applications, the goal was to reach formal conclusions about the text data. The goal of this project was less formal and more exploratory: To find distinctions between less engaged and more engaged employees that could be the basis for new scale items not anticipated by previous research or experience.

3.1. Obtain text data

Among the questions included on the MPS 2005 was an open-ended question about improving employee performance: “Please describe one change to your work situation your agency or supervisor could make that would improve your personal job performance.” Responses to this question in the MPS 2005 data matrix were linked with responses to closed-ended questions, including those that comprise the engagement scale and classification into Less Engaged Somewhat Engaged, and Highly Engaged groups based on the aggregate engagement scale score. This allows matching of individual narrative responses to engagement scale scores—and the three level of engagement categories created for agency outcome analysis.

Such narrative responses are useful because they can capture unexpected employee attitudes [18, 19]. Because previous research had established a relationship between engagement level and individual performance [1, 2], there was reason to hypothesize differences in the ways more engaged and less engaged employees think about and describe barriers to high performance. We can reasonably expect responses to a question about improving performance to suggest aspects of engagement that might not occur to researchers writing new scale items. There would be some value to simply reviewing these responses for item-writing ideas. But the large sample size ($n=36,926$) makes this somewhat impractical. An automated approach could save time and identify themes that case-by-case review might not uncover.

3.2. Choose text mining software

The text mining tasks for this project were conducted using the WordStat content analysis software from Provalis Research [20]. There are alternative software choices which might have been used, including modules for the open-source R statistics program. WordStat combined sufficient power with ease of use. More important for our approach, it supports user-software collaboration

rather than requiring complete automation of the text mining process. The following steps were conducted using WordStat.

3.3. Count words and phrases

In this step, each occurrence of a word or phrase in each narrative response was tabulated. Responses to the open-ended question included 505,657 total words. There were many repetitions using 14,442 unique words. There were 1,531,692 unique two- to five-word phrases. Words and phrases used by fewer than five survey participants were dropped from the analysis to focus review on the most frequently occurring features. Although important in a more exhaustive review, lower frequency words do not help as much to identify a small number of broad themes in a large data set. This five-case cutoff was arbitrarily chosen, but necessary to impose practical constraints on the process [21]. The analysis then focused on 4,153 unique words and 20,478 unique phrases.

3.4. Apply exclusion list

A large percentage of words used in any sample of text do not carry meaning, but instead bring grammatical structure to sentences. An exclusion or “stop list” of words to be disregarded in subsequent steps of processing usually contains such high-frequency words (e.g., “a”, “and”, “the”). The software removes these words from consideration as key words and phrases.

This project used a widely-available exclusion list of 609 common English function words [22]. Applying the exclusion list reduced the data set to 3,668 unique words and 1,801 unique phrases.

3.5. Standardize word forms

In this step, sometimes called “stemming” or “lemmatization,” words that are grammatical variations of the same root are recoded as the same word. For example, the words “run,” “ran,” and “running” would be recoded as the same word [23]. Lemmatization was performed by the WordStat software. This process “reduced” our focus to 2,753 unique words and 2,000 unique phrases.

It seems counterintuitive that the number of phrases should increase because of this “data reduction” step. This results from an interaction between the five-case minimum and this step. Once lemmatization is applied, some formerly different words are considered the same. This causes an

increase in the frequency of some phrases as previously-different phrases are counted together. (Now “run away” and “running away” are counted as the same, for example.) This allows some phrases previously screened by the five-case requirement to be counted as they are now more frequent. (Three “run away” instances and three “running away” instances were formerly excluded. Now they are combined into six instances of “run away” that passes the threshold and is counted as a phrase.) This effect occurs for words as well as phrases, but in this analysis, this does not result in a net increase in included words. Although data has not, strictly-speaking, been reduced, the number of useful features is increased, along with the potential predictive power of models based on these features.

3.6. Identify distinguishing features

WordStat uses a Chi-Squared procedure to determine which words and phrases are used more frequently or more groups. This approach was used to compare responses from employees with higher or lower engagement scale scores. The software can use this information to create a specialized classifier which attempts to identify an employee’s level of engagement based on his or her narrative response about performance. WordStat used Chi-Squared feature selection with maximum Chi-Squared feature weighting to construct the classifier using 300 of the highest predicting words and phrases.

In theory, such a model could be used to assign engagement scale scores or engagement group membership based on narrative responses alone, functioning as “text scale” [24]. The model based on this data predicts engagement level with only 22.4% accuracy. It is clearly not an adequate substitute for our more traditional engagement scale. Fortunately, we do not need a powerful model for our purposes. Our focus was on the 300 features identified as part of the model-building process. They had the potential to reveal distinctions between more and less engaged employees which could be used to develop new engagement scale items.

The text mining procedure reduced 14,442 unique words and 1,531,692 phrases—that’s just over one and a half million (1,546,134) features—down to the 300 best predicting words and phrases. This is still a lot of information, and the terms identified refer to clusters of original responses. But this distilled set of features brings focus to review of the narrative responses. Reviewing 300 features is still a commitment, but is a doable task.

3.7. Review words and phrases

Researchers reviewed words and phrases which predicted either high or low levels of engagement. When words or sets of words on the list suggested a pattern, researchers used a WordStat Keyword-in-Context (KWIC) table to review entire responses containing the words or phrases.

During this review researchers removed from consideration any responses that seemed too closely linked to any of the original 16 engagement scale items, or that in their judgment could not be captured in an Agree/Disagree statement format. Most completely irrelevant responses are idiosyncratic to one or a small number of survey participants and were screened out of the analysis by the five-case-minimum requirement described in Section 3.3.

Some of the remaining frequent responses are also dropped because they are unsurprising differentiators. Less engaged employees, for example, are more likely to respond with “I don’t know.” They may be disengaged from the survey process as well as from their jobs. More engaged employees, on the other hand, are more likely to say “Nothing” or “Not applicable” when asked about improving their performance. Perhaps they do not look for help with this from outside sources. Some judgment is needed about what is truly off-topic at this stage. Any word or phrase that makes it this far plays some role in distinguishing between employee engagement levels. Not all can be translated into a scale item.

Researchers next reviewed responses containing highly predicting words and phrases for patterns. In this set of responses, we noticed two differentiating patterns. There were many references to supervisors, management, and other employees. Examining the responses containing these terms indicated that less engaged employees were more likely to hold their supervisors and upper management responsible for inhibiting their performance. More engaged employees were more likely to cite poorly performing employees at their own level as responsible for reducing their effectiveness.

There were also many references to workplace policies and lack of staff, funding, and other specific resources. The responses indicated that less engaged employees were more likely to feel constrained by organizational policies. More engaged employees were more likely to claim they could accomplish more if given more staff, more funding, more equipment, and more resources in general.

3.8. Write new scale items

Four new engagement scale items were written to capture these two distinctions. The same standard item writing, review, and revision process was followed with the ideas generated from text mining as with ideas generated from the literature and research agenda [4]. The engagement scale now included the original 16 items, four items written using text mining, two items to cover the missing Effort Beyond Minimum Required engagement factor, and two reference items. All 24 items are listed in the Appendix at the end of this article.

4. Results

The extended 24-item engagement scale was included in the 2010 Merit Principles Survey. Responses were obtained from 37,137 of 60,000 full-time Federal employees invited to participate. The same sampling and administration procedures were used as for the MPS 2005.

Table 1 contains item statistics that show the effects of removing different subsets of items from the engagement scale. The top half of the table focuses on subscales; the bottom half focuses on the four text mining-derived items. As a group, the four text mining items perform less well than the other scale components, including the two non-engagement comparison items. An examination of the individual items sheds more light on their individual effects.

If Item 21 were removed, this would improve the scale more than removing any other single item. Most survey participants agreed with this item, reducing item variance and lowering item statistics. This kind of ceiling effect—or a contrasting floor effect when almost no participants agree—is not an uncommon finding in an item’s first exposure. These effects can be usefully reduced by rewriting the item to soften or strengthen its tone. Such revision might make this item more useful.

Item 24 is the next-weakest item. In reviewing it, we noted that both Item 21 and Item 24 push the boundaries of good item writing by being long with embedded lists [25]. The MPS has a long history of using longitudinal tracking questions with this flaw. Even so, these items could have been written without the embedded examples, or with fewer examples. The narrative responses were the source of these embedded examples. This suggests that caution should be used in trying to make a small set of items “cover” a general trend in this way.

These two items will be removed from the scale, perhaps to be rewritten. The reduced set of text

mining items, labeled “Adjusted Text Mined (2)” in Table 1, has item statistics comparable to items in the original scale and those derived from literature review and from the research agenda. These items can be retained and increase the internal consistency of the engagement scale. The items arguably increase the validity of this scale as well by representing aspects of engagement derived from the responses of Federal employees.

Table 1. Subscale and item statistics.

Statistics are with Item(s) Removed	Scale Mean	Item(s) – Scale Correlation	Cronbach Alpha
Engagement Scale Subscales			
Original (16)	86.73	.679	.947
Literature (2)	86.67	.763	.946
Agenda (2)	86.92	.731	.946
Text Mined (4)	86.93	.371	.950
Adjusted Text Mined (2)	87.32	.498	.949
Items Derived from Text Mining			
Lack Resources (#21)	85.75	-.013	.952
Other Employees (#22)	87.12	.632	.948
Managers (#23)	87.43	.582	.948
Barriers (#24)	87.42	.281	.952

5. Discussion

There are any number of approaches, formal and informal, scale developers use to generate ideas for scale items. As mentioned in the introduction, a key quality features of any source of potential topics are links to both the construct being measured and the population of interest. The narrative response mining

method introduced here has both of those links, giving it better potential than informal methods. That it produced successfully performing items demonstrates that it belongs in the ranks of respectable methods. It can, of course, be improved. Several strategies are worth considering.

It is the nature of text mining to search opportunistically through text that was created for another purpose. In this project, the open-ended question about improving performance was included in the survey for other reasons and was not intended for use in scale development. A more targeted question designed for this purpose is one possible improvement. For example, following the engagement scale items with “What other factors contribute to how engaged you are with your work? What factors are barriers to engagement?” would have generated ideas more directly linked to the purpose of the scale. We might also have extended the group-comparison strategy made further use of other closed-ended questions on the survey to, investigating differences between supervisors and nonsupervisors, field and headquarters employees, and other demographics.

This project did not make full use of the text mining capabilities available in the software. The goal was simple detection of candidate topics related to employee engagement level which would then feed into the item authoring process. The procedure followed seemed sufficient to support this. But the process could have benefitted from additional exploration of the narrative text.

For example, we might have begun exploration of the overall themes presenting in the narrative responses using WordStat’s topic modelling capability [26, 27]. This would have revealed the most prevalent content themes in the responses. We might have then explored their relationship to employee engagement level. The prevalence across respondents of topics identified through topic modelling could have provided perspective on the prevalence of the topics we identified through engagement group comparison.

Some researchers have measured the presence of themes in text using pre-defined lists of words and phrases that signal the presence of each theme. The DICTION software, for example, uses this approach to profile political writing and speech [28]. Another program, LIWC, detects themes in journals and other personal documents [29]. It is the nature of text mining, sometimes, to search for the unknown. If we know in advance the relevant themes in our narrative responses, no further analysis is needed. It is possible that less specialized content dictionaries which search across a large set of possible themes

might prove useful for mining responses to open-ended questions. This area of research bears watching.

The open-source R statistical environment includes several user-contributed packages that support text analysis in general [26, 30]. While not specialized for scale development, nor as easy to use, they have potential for adaptation to our purpose. This is also a source to watch for emerging innovations useful to narrative response analysis.

We might also have used other text features in addition to words and phrases such as punctuation and more complex phrases than adjacent word sequences [31]. We were cautious that such improvements might decrease the interpretability of text mining results, expend too much time, or require access to expensive resources. As text mining develops as a discipline, advances may make improved tools and processes more accessible. At this point in time, the techniques used in this project are sufficient to mine narrative responses for dominant themes [32].

There are some data-level process decisions that might be made differently. Spellchecking might have reduced the “messiness” of our text data, making improvements similar to those observed from standardizing word forms. We had concluded that it might not be worth the necessary time. Our previous experience indicated it might not be necessary for this type of task using a large data set with strong themes [33]. Other techniques, such as grammatical parsing and identification of each word’s specific meaning [34], might also sharpen the text mining software’s focus.

Predictive models based on text data can sometimes be improved by focusing on trigrams (overlapping three-letter sequences within words) instead of the words themselves. While this can increase model precision, it may also reduce a human reviewer’s ability to discover themes among the trigrams. Perhaps interpretability could be improved with additional software feature development (or increased familiarity with existing features by the researchers).

Our ability to learn from integration of the text mining process into scale development could be aided by greater documentation of our overall procedures. This is not common in scale development, but would be advantageous to process improvement. For example, it would be useful to know the normal “attrition rate” of items produced from other methods, both formal and informal, to make a more granular comparison of the contribution of new approaches. There is general advice to draft two to three times as many items as will ultimately

appear on the scale [4], but it is not clear how expected attrition might vary under different authoring conditions and at different points in the scale testing and revision process.

There is one final issue that we might reflect upon. As social scientists, we are most comfortable with rigorous, reliable quantitative methods. Although we can be slow to admit it, we are also privately comfortable with any number of informal, loosely-documented, artful practices like those involved in the initial writing of scale items. Somehow, we are less comfortable with a mingling of the two—even when such combination is of practical benefit. Why is that? This case study suggests we might usefully challenge our comfort zones in this respect.

6. Conclusions

The text mining approach to identifying potential topics for new engagement scale items was effective because it generated usable items that covering of engagement not already addressed in the existing scale. It is a feasible approach to use when relevant source of respondent-generated text is available with an empirical (best) or logical (at least) link to the construct the scale is measuring. The authors believe this approach, with some improvements, can be useful to other survey researchers developing items to improve measurement scales.

7. Appendix – Engagement scale items

7.1. Original 16 engagement items

1. My agency is successful in accomplishing its mission.
2. The work I do is meaningful to me.
3. My work unit produces high quality products and services.
4. Overall, I am satisfied with my supervisor.
5. Overall, I am satisfied with managers above my immediate supervisor.
6. I know what is expected of me on the job.
7. My job makes good use of my skills and abilities.
8. I have the resources to do my job well.
9. I would recommend my agency as a place to work.
10. I have sufficient opportunities (such as challenging assignments or projects) to earn a high performance rating.
11. Recognition and rewards are based on performance in my work unit.

12. I am satisfied with the recognition and rewards I receive for my work.
13. I am given a real opportunity to improve my skills in my organization.
14. I am treated with respect at work.
15. My opinions count at work.
16. A spirit of cooperation and teamwork exists in my work unit.

7.2. Literature review items

17. At my job, I am inspired to do my best work.
18. I have the opportunity to perform well at challenging work.

7.3. Research agenda items

19. My supervisor provides constructive feedback on my job performance.
20. My supervisor provides timely feedback on my job performance.

7.4. Text mining items

21. Lack of resources, such as more staff, a larger budget, or more equipment and supplies, is a primary reason my job performance is not higher.
22. The performance and/or conduct of other employees are primary reasons my job performance is not higher.
23. The performance and/or conduct of my supervisors and managers are primary reasons my job performance is not higher.
24. Barriers to success, such as constraining rules or work processes, under-informed coworkers, or office politics, are the primary reasons my performance is not higher.

8. References

- [1] MSPB, The power of Federal employee engagement, U.S. Merit Systems Protection Board, Washington, DC, 2008.
- [2] MSPB, Managing for employee engagement, U.S. Merit Systems Protection Board, Washington, DC, 2009.
- [3] Groves, R., F. Fowler, M. Couper, J. Lepkowski, E. Singer, and R. Tourangeau, Survey methodology, Wiley-Blackwell, Hoboken, NJ, 2009.
- [4] Johnson, R. and G. Morgan, Survey scales: A guide to development, analysis, and reporting, Guilford Press, New York, NY, 2016.

- [5] Fowler, F., *Survey research methods*, Sage, Los Angeles, CA, 2009.
- [6] Macey, W., B. Schneider, K. Barbara, and S. Young, *Employee engagement: tools for analysis, practice, and competitive advantage*, Wiley-Blackwell, Hoboken, NJ, 2009.
- [7] MSPB, *Results of the Merit Principles Survey 2005*, U.S. Merit Systems Protection Board, Washington, DC, 2007.
- [8] Nierle, D., J. Ford, and J. Tsugawa, "Improving a Federal employee engagement scale," *International Personnel Assessment Council*, Santa Barbara, CA, 2010.
- [9] Cronbach, L., "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, no. 3, pp. 297-334, 1951.
- [10] MSPB, *U.S. Merit Systems Protection Board 2011—2013 research agenda*, U.S. Merit Systems Protection Board, Washington, DC, 2011.
- [11] Campbell, D. and D. Fiske, "Convergent and discriminant validation by the multitrait-multimethod matrix," *Psychological Bulletin*, vol. 56, no. 2, pp. 81-105, 1959.
- [12] Krippendorff, K. *Content analysis: An introduction to its methodology*. Sage, Thousand Oaks, CA, 2012.
- [13] Krippendorff, K. and M. Bock, *The content analysis reader*. Sage, Thousand Oaks, CA, 2008.
- [14] Neuendorf, K., *The content analysis guidebook*, Sage, Thousand Oaks, CA, 2016.
- [15] DeVellis, R., *Scale development: Theory and applications (4rd Ed.)*, Sage, Thousand Oaks, CA, 2017.
- [16] Ford, J., T. Stetz, M. Bott, and B. O'Leary, "Automated content analysis of multiple-choice test item banks," *Social Science Computer Review*, vol. 18, no. 3, pp. 258-271, 2000.
- [17] Stetz, T. and J. Ford, "Leadership and same-gender bias: A content analysis of promotion recommendations," *Journal of Psychological Issues in Organizational Culture*, vol. 1, no. 1, pp. 6-18, 2010.
- [18] Kulesa, P. and R. Bishop, "What did they really mean? New and emerging methods for analyzing themes in open-ended comments," in A. Kraut (Ed.), *Getting action from organizational surveys*, Jossey-Bass, San Francisco, CA, pp. 238-263, 2006.
- [19] Reja, U., K. Manfreda, V. Hlebec, and V. Vehovar, "Open-ended vs. close-ended questions in web questionnaires," In A. Ferligj and A. Mrvar (Eds.), *Developments in Applied Statistics*, FDV, Ljubljana, Slovenia, 2003.
- [20] Peladeau, N., *WordStat 7.1*. Provalis Research, Montreal, CA, 2017.
- [21] Deng, Q., M. Hine, S. Ji, and S. Sur, "Building an environmental sustainability dictionary for the IT industry," *HICSS-50*, Honolulu, HI, 2017.
- [22] Wilbur, W. and K. Sirotkin, "The automatic identification of stop words," *Journal of Information Science*, vol. 18, no. 1, pp. 46-66, 1992.
- [23] Frakes, W., *Stemming algorithms, Information retrieval: data structures and algorithms*. Prentice-Hall, Upper Saddle River, NJ, 1992.
- [24] Miller, T., *Data and text mining: A business applications approach*, Prentice-Hall, Upper Saddle River, NJ, 2004.
- [25] Tourangeau, R., L. Rips, and K. Rasinski, *The psychology of survey response*. Cambridge University Press, Cambridge, UK, 2000.
- [26] Jockers, M., *Text analysis with R for students of literature*, Springer, New York, NY, 2014.
- [27] Roberts, M., B. Stewart, D. Tingley, C. Lucas, J. Leder-Luis, S. Gadarian, B. Albertson, and D. Rand, "Structural topic models for open-ended survey responses," *American Journal of Political Science*, vol. 58, no. 4, pp. 1064-1082, 2014.
- [28] Hart, R. and A. Curry, "The third voice of American politics," *Presidential Studies Quarterly*, vol. 46, no. 1, pp. 73-97, 2016.
- [29] Pennebaker, J., *The secret life of pronouns: What our words say about us*, Bloomsbury Press, New York, NY, 2011.
- [30] Kwartler, T., *Text mining in practice with R*, Wiley, Hoboken, NJ, 2017.
- [31] Zeng, A., *Mastering feature engineering principles and techniques for data scientists*, O'Reilly Media, Sebastopol, CA, 2017.
- [32] Bondi, M. and M. Scott, *Keyness in texts*. John Benjamins Publishing, Philadelphia, PA, 2010.
- [33] MSPB, *Making the right connections: Targeting the best competencies for training*, U.S. Merit Systems Protection Board, Washington, DC, 2011.
- [34] Miner, G., J. Elder, A. Fast, T. Hill, R. Nisbet, and D. Delen, *Practical text mining and statistical analysis for non-structured text data applications*, Academic Press, New York, NY, 2012.