Methods to Integrate Natural Language Processing Into Qualitative Research

Marissa D. Abram1, Karen T. Mancini1, and R. David Parker2

Abstract
Background: Qualitative methods analyze contextualized, unstructured data. These methods are time and cost intensive, often resulting in small sample sizes and yielding findings that are complicated to replicate. Integrating natural language processing (NLP) into a qualitative project can increase efficiency through time and cost savings; increase sample sizes; and allow for validation through replication. This study compared the findings, costs, and time spent between a traditional qualitative method (Investigator only) to a method pairing a qualitative investigator with an NLP function (Investigator + NLP).

Methods: Using secondary data from a previously published study, the investigators designed an NLP process in Python to yield a corpus, keywords, keyword influence, and the primary topics. A qualitative researcher reviewed and interpreted the output. These findings were compared to the previous study results.

Results: Using comparative review, our results closely matched the original findings. The NLP + Investigator method reduced the project time by a minimum of 120 hours and costs by $1,500.

Discussion: Qualitative research can evolve by incorporating NLP methods. These methods can increase sample size, reduce project time, and significantly reduce costs. The results of an integrated NLP process create a corpus and code which can be reviewed and verified, thus allowing a replicable, qualitative study. New data can be added over time and analyzed using the same interpretation and identification. Off the shelf qualitative software may be easier to use, but it can be expensive and may not offer a tailored approach or easily interpretable outcomes which further benefits researchers.

Keywords
natural language processing, NLP, unstructured data, mixed methods

Introduction
The global increase of data has resulted from the expanded incorporation of electronic devices and software, with increasingly accelerated use over the last 30 years. Quantitative data are produced from everyday items: cars, televisions, electronic medical devices, smart watches, and a myriad of other devices and processes. In 2020, the data volume in bytes exceeds 40 times the number of stars in the observable universe at 44 zettabytes \((1,000^7 \text{ bytes})\) and with increased production, new and re-emerging analytic methods are deployed to increase our understanding of the world (Desjardins, 2019). Because of a drive to understand the world through data, new statistical techniques are emerging, and older known methods are being given new use. As the amount of qualitative and unstructured data is increasing, developing new methods or integrating methods from other disciplines could expand the capacity of qualitative researchers to analyze larger amounts and more diverse data types. Linking qualitative and quantitative researchers and methods has the potential to expand our understanding of the emerging world of data.

Machine Learning (ML) is one of these re-emerging methods which was developed in the 1950s (Samuel, 1959). Depending on the definition of ML used, there range anywhere from four general concepts to thousands of ML approaches. ML is now used as an umbrella concept for statistical methods that process large amounts of data using algorithms developed and coded in a software application by which a computer iterates and refines analyses (or a statistical model) over time, thereby “learning” the best approach to produce more relevant results. Data mining is an example of a common ML approach where an algorithm is developed and employed to look for relationships between the data which are previously unknown.

1 Adelphi University College of Nursing & Public Health, Long Island, NY, USA
2 College of Health, University of Alaska Anchorage, AK, USA

Corresponding Author:
R. David Parker, College of Health, University of Alaska Anchorage, 1901 Bragaw Street, Suite 220, Anchorage, AK 99508, USA. Email: david.parker@alaska.edu

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
ML has been successful recently in drug discovery, evidence based medicine, and multiple medical healthcare treatments (Alsawas et al., 2016; Lee et al., 2019).

Natural Language Processing (NLP) is an example of an ML method. Like traditional analytics, NLP is a constellation of methods categorized under a concept and not a specific application. NLP is used for qualitative data in the traditional sense or additionally for qualitative data which is collected in other formats, including open ended or free form feedback on a customer satisfaction survey, medical provider notes in an electronic medical record (EMR), or a transcript of research participant interviews (Koleck et al., 2019).

NLP can be a cumulative methodological process for researchers beginning with the development of baseline functions and knowledge, which can then be extended to include additional data, programming, and analytic techniques (Miller & Brown, 2018). Two basic and essential NLP concepts are corpus and dictionary. While specific definitions vary, a corpus is generally defined as the entire body of text which contains the contextual information on words in the text, while a dictionary is a list of words in the language of analysis. NLP methods provide exceptional opportunities to analyze large amounts of unstructured, qualitative data, however, as the corpus and dictionary are constructed by people, they are not bias free nor completely objective. Additionally, to study a new content area, each would need to be updated (Guetterman et al., 2018; Koleck et al., 2019). The potential for investigator introduced bias holds true for any analysis in which a person defines the approach and interprets the findings. NLP offers a distinct advantage, the same algorithm will be applied to all the data, this changes most subjective bias to a systematic bias. When the data are electronic, it is then far less cumbersome for review and analysis by additional persons. As with any replicable approach, over time, there should be a decreased bias and the data should approach convergence between observed outcomes and expected outcomes or observed outcomes and the “truth.”

Qualitative research by nature is criticized as an overly subjective research method partially due to the use of unstructured or semi structured data. As many analyses are conducted by a single investigator, this can be time and resource consuming and increasingly expensive for analyzing even small data sets (Renz et al., 2018). Compared to quantitative research, qualitative research provides deep insights into a phenomenon. By design, qualitative research is intended to provide thick, rich descriptions, which can then be used to develop additional insights leading to further research. As with quantitative methods, qualitative may have replication challenges, as it can be interpretative and based on a co-creation of researcher experience with the phenomena. At a time when scientific research is suffering from a “great replication crisis” and multiple projects have been identified as not replicable, indicating this is a time to explore approaches to address these concerns (“Challenges in irreproducible research,” 2018). Whether through poor documentation of methodological procedures, misapplication of analytics, or intentional misrepresentation, this replication problem calls for an increased effort to ensure a consistent and replicable approach to research. The increase of free form text, or unstructured data, presents the need to build on traditional, qualitative approaches. More easily accessible, standardized qualitative research methods to manage bigger data sets need to be developed.

NLP methods create opportunities for qualitative researchers to push beyond the single investigator method, where the person recruits, interviews, transcribes, and analyzes a small study sample, which inhibits generalizability (Carminati, 2018). Building on a strong qualitative methods study by augmenting this approach with emerging technology, investigators will be able to harness the power of more data, providing deeper insights, and increasing the impact of an investigation (Guetterman et al., 2018). While larger data sets, whether in qualitative or quantitative research, do not always translate into more precise knowledge, if the methods are not appropriate, if unidentified systematic bias remains, or if there is an error in the interpretation, the use of sophisticated analytical techniques will not mitigate the problem.

This project replicated a qualitative interview study by integrating an NLP component into the same qualitative research analysis plan, and then comparing the original and new findings. NLP identified the important terms and arrange them into topics. These topics were then analyzed by an experienced qualitative researcher to identify the important themes. These themes from the NLP identified terms were then compared to the original analysis of the same interview data conducted by a qualitative researcher alone.

Methods

The deidentified data for this project were derived from nine confidential interviews of nurses who work in the substance use field which had been previously analyzed, from which the results were published (Abram, 2018). The data were provided without any identifiers and as this project used anonymous data, it meets the Federal definition of non-human subjects research and exempt from institutional review (45 CFR Part 46). The project methods were conducted in two stages: data preprocessing, where the data were prepared for analysis through standardization and formatting; and data analyses.

An extensive literature review was conducted to determine which statistical software program and packages would be used to conduct this project, including both peer reviewed journals and grey literature, such as technical reports, blogs on natural language processing, and information technology related websites (Alsawas et al., 2016; Guetterman et al., 2018; Hong et al., 2018; Koleck et al., 2019; Lee et al., 2019; Matsutani et al., 2019; Miller & Brown, 2018; Sievert & Shirley, 2014). Our preference for software was an open source freeware platform as we hoped to develop a methodology which could be provided for other researchers to use and would not require significant financial expenditures. After this review, there were two primary contenders, R Programming Language and Python. There was more information which was also more
recent available for Python and NLP, therefore, this met our
criteria for selection. Python is a free programming language
available for download from the Python Foundation, a
non-profit group which has oversees official releases. Available
since 1990, Python has grown to become the most used
software application for data analyses primarily because it is
free and is an easier language to learn.

Python is greatly enhanced using an additional software
component called an integrated development environment
(IDE). This add on software greatly enhances to the human
interaction with the software as it moves the coding more into
a recognizable set of human language commands. The same
metrics were applied in our search for an IDE, free and easy to
use. Our findings supported the software called Spyder given
its widespread use, relatively easy use, and its limited impact
on computing resources (“Spyder,” 2019). All analyses were
completed using Python 3.7.4 64-bit with the Spyder 3.3.6
(IDE) (“Python,” 2020; “Spyder,” 2019).

The initial step required the data to be transformed into a
usable format. The data were provided in Word (.docx) and
Adobe Acrobat (pdf) formats. Adobe Acrobat DC converted
the .pdf files to an electronically readable format using its
internal optical character recognition (OCR). Manual removal
of handwritten notes and formatting errors introduced by OCR
was completed. All interview data were then converted into .txt
files to prepare for data preprocessing and analyses.

To further reduce time required and to increase replicability
in future studies, a programming function was written to man-
age the data preprocessing. A function is highly beneficial as it
can be saved and used in additional analyses where the same
processes need to be used. The primary packages and modules
used to create this function included: genism, nltk, nump, and
pyLDAvis (“Python,” 2020; “Spyder,” 2019). The standard
data preprocessing methods for NLP were employed and
included: a single file merger; text case standardized to lower;
removal of stop words (ex. “and,” “or” “but,” “the”); punctua-
tion removal; word stemming (removal of word endings allow-
ing a focus on the word root); and text conversion of numbers.
Excluding words which do not contribute to the identification
of relevant themes is a standard process in many approaches to
qualitative research as it improves a researcher’s ability to
identify significant statements, formulate meanings, and cluster
themes (Konovalov et al., 2010; Morrow et al., 2015). While
there are many standard corpuses available, given the special-
ized wording expected to be present in the interview inter-
ceted with nurses about addiction, we created a new
dictionary and corpus for analyses (Alsuhaibani et al., 2018;
Neuraz et al., 2019). Corpus creation for a specific project is a
process which can range from simple, with less precision, to
complex with a higher level of specificity. For this project, we
used an approach where each word was included in the corpus
with the importance of the word created using a weighted fre-
cency of occurrence. For instance, if a word was used 30
times, it would have a higher level of importance due to its
occurrence than a word used five times.

Latent Dirichlet Allocation (LDA) was used as the initial
modeling technique in these analyses. LDA analysis is a com-
mon tool used in natural language processing as a generative
probabilistic model of a corpus (Blei et al., 2003). LDA
assumes documents are a random mixture of words over latent
topics, whereas each topic may be characterized by the distri-
bution over the contained words. Two standard measures of
performance in NLP modeling are perplexity and coherence
scores. As perplexity is an indicator of how well a model fits
the data and given the interpretive, non-predictive nature of our
model, we did not use perplexity scores. Perplexity provides an
indicator of the next expected word based on the previously
used words. As we are replicating a previously conducted study
by a qualitative researcher, we used coherence scores as indi-
cators of the word relativism within the topics. Coherence
scores measure the degree of semantic similarity between high
scoring words in each topic, so we used coherence as a measure
to arrange the words within the topics without obtaining the
scores.

To conduct the qualitative assessment of the NLP generated
key words, the LDA output including the words and their influ-
ence values were provided to a qualitative researcher who had
no affiliation with the previous study. The researcher used the
key words and influence scores to develop topic themes and an
overall theme for the data. The themes developed using NLP +
qualitative researcher were then compared to the themes from
the original manuscript which were developed solely by a qua-
Lititive researcher and no machine learning process.

Results

The combined interview data contained 169,666 words. Given
the intent to extract themes, the standard process of creating
word pairs (bigram) and three word groupings (trigram) was
used only to further refine, append, and update the stop words
list. These groupings provide additional context for interpreta-
tion and thematic grouping than individual words alone as they
consider two- or three-word sets. repository was conducted
Removal of additional words, such as “uh,” and “umm,” and
similar phraseology not included in the stop words to increase
thematic clarity. This iterative method resulted in a unique
word dictionary (n = 1,846) after all processing was
completed.

To determine the number of topics contributing added value
of the statistical model, we used the calculated coherence
scores. Coherence scores serve as an indicator of the impor-
tance of word by showing its distance from the location other
words in the model. Words which are closely linked, as indi-
cated by a higher score, are then aggregated into a topic. The
sum of all coherence scores in a topic explains the variability
and predictive ability of that topic in capturing the topics of
importance. To determine how many topics were represented,
we allowed the modeling process to run until no topics of
significance were identified.

From these data, we found that a model with two topics had
a coherence value of 0.2188; an eight-topic model had a
Table 1. Model Topics and the Contributing Words With Probability Influence.

| Topic | Influence Value × WORD BASE | Ending(s) |
|-------|-----------------------------|-----------|
| 1     | 0.048 patient(s)            |           |
|       | 0.038 nurse(e, ed, ing)     |           |
|       | 0.024 time(s)               |           |
|       | 0.023 people                |           |
|       | 0.016 something             |           |
|       | 0.015 care(d, s)            |           |
|       | 0.013 talk(s)               |           |
|       | 0.013 substance(s)          |           |
|       | 0.011 kind                  |           |
|       | 0.010 feel                  |           |
| 2     | 0.025 day(s)                |           |
|       | 0.020 work(ed, s)           |           |
|       | 0.020 sometimes             |           |
|       | 0.019 medic (action, ine, ines) |   |
|       | 0.015 little                |           |
|       | 0.015 give(n, s)            |           |
|       | 0.014 person                |           |
|       | 0.013 call(ed, s)           |           |
|       | 0.013 alway(s)              |           |
|       | 0.013 differ(ent, s)        |           |

Table 2. Comparison of Investigator Only Findings with Natural Language Processing and Investigator Findings.

| Investigator Only | NLP + Investigator |
|-------------------|--------------------|
| Defining the role for self | Providing patient-centered care to clients with substance abuse is layered and sensitive |
| Learning the role | The overall interpretation was that the RN role in addiction care needs further definition and support |
| Navigating with ease in an unchanging culture | The complex work of a registered nurse (RN) caring for clients with substance addiction not always recognized. |

coherence value of 0.2054; and a 14-topic model’s coherence value was 0.2884. A parsimonious approach in statistical modeling guides us to choose the simplest model that has the best value, in this case, there is no statistically significant benefit for a model with more than two topics using this LDA approach. This was determined by the small difference between the scores (Ding et al., 2018). Within each topic, the 10 words with the highest influence as identified by coherence scores were selected for further analyses. The influence score shows the probability of inclusion in the topic by word. The top ranked words for each topic, along with the word’s influence score are shown in Table 1.

The percentages of influential words from the corpus included each topic represents the topic influence in this two topic model. Topic 1 had the highest influence at 52.6% influence and Topic 2 words contained 47.4% attributable influence toward the model. This means that Topic 1 explained more of the predictive approach than did Topic 2. As this NLP approach used only identifies topics based on the estimated probability of the words in each and does not provide an interpretation, we provided the 10 most relevant terms from each topic to an experienced qualitative researcher who had no knowledge of the original project. This researcher used these words to construct a thematic interpretation based on the NLP results as shown in Table 2. The findings were: Topic 1: providing patient-centered care to clients with substance abuse is layered and sensitive; Topic 2: the complex work of a registered nurse (RN) caring for clients with substance addiction not always recognized. The overall interpretation was that the registered nurse (RN) role in addiction care needs further definition and support.

Using the identified NLP terms, the researcher identified overall theme sets. Table 2 compares the new findings with those of the original study which has been previously reported (Abram, 2018). After the analyses and interpretations were completed by the qualitative researcher on this project, the findings were provided to the investigator on the original project for comparison. After review, the original investigator agreed with interpretation of findings and found them to be in alignment with the original results.

Discussion

This study evaluated the ability to automate components of qualitative analyses and then compare the outcomes of these themes to the original manual process. NLP captured the overall thematic descriptions however, as used, did not provide rich descriptions that described the nuances of the participant experience, for example the what made the phenomena layered and sensitive. This basic, easy to implement approach of NLP would work best with methods such as content analysis. This method is however, also useful to the researcher who is trying to define or describe a phenomenon using a large qualitative or unstructured data source that seeks to generate the findings from the experience of the participant.

There are many positive implications of integrating NLP and other ML approaches into qualitative research. Cost and time savings are among the primary benefits. For investigators who choose this method, the use of a thematic or project based dictionary allows for specialized terms to increase the accuracy of identification of influence. Another benefit to our approach was that the automated, machine learning based transcription represents a significant cost savings over human transcription. The use of automated transcription services with investigator verification lowered the costs from $700 in the original study to approximately $38.00 in our NLP integrated approach (“Amazon Transcribe Pricing,” 2020). Using NLP for this project, the primary concepts were identified in less than 2 minutes of analysis. The entire process, including data cleaning, preprocessing, and coding required approximately 40 hours yielding code that we can use for other, similar projects. Had we replicated this study through standard qualitative approaches, listening to the interviews would have taken 14 hours, in addition to the time to transcribe and analyze the data. The original
investigator estimates the time consumed for this project was at a minimum 120 hours.

Additionally, the software applications used for this project were free, including Python and Google Sheets. Qualitative software, like quantitative software, can be expensive and may limit an investigator’s ability to complete research. Pricing is negotiated on a per institution or per system contract and can vary widely in cost with larger universities able to offer lower costs than smaller institutions. Investigators without an academic affiliation pay the highest costs (“Atlas.ti Qualitative Data Analysis Shop,” 2020).

There are limitations to our approach and project. For instance, while two of the researchers were qualitative nurse researchers, one is a quantitative epidemiologist. We believe this interdisciplinary approach adds strength but communicating concepts and terms between the different research approaches increased the time for project development and completion. The original research format was an interview, therefore, the original researcher may have captured additional information in the form of undocumented responses, including intonation. These could have been lost from a transcription analysis without the human interaction. Given that secondary data analysis is limited due to the inability to collect additional data, we could not expand our data given we did not have access to the interviewees. The weaknesses related to secondary data are mitigated as the specific focus of this project is to use a new method (NLP and a qualitative researcher) and compare the findings to the findings derived by the original, qualitative researcher alone. As we do not seek to add to the literature of nurses working in substance use recovery settings but to the knowledge of methods to enhance qualitative methods, we believe we were able to successfully complete our project despite the traditional research limitations.

**Conclusion**

Modified and new methods are important to analyze the increasing amount of data in the world. Unstructured data, which is often derived from free form fields and a type of qualitative data, is often left under analyzed given the complexity of analysis (Hong et al., 2019). Unstructured health and medical data offer many possibilities to increase our understanding of disease and symptomology, especially in the area of emerging infections or conditions which require a clinical diagnosis as these data are contained in many areas within health records. The ability to efficiently and cost effectively harness these data do not only represent a great opportunity in the area of qualitative research but also has implications for more traditional qualitative analysis from the novel methods employed for big data.

For this project, we assembled an interdisciplinary team including an experienced qualitative researchers and a quantitative researcher with advanced statistical skills and intermediate Python programming abilities to demonstrate the ability to identify relevant packages and modules through a literature review and integrate their use by constructing a programming function. Any type of researcher with an interest in programming and statistical knowledge of the analyses employed could complete such a project. Future research will extend the methods used here to more complicated and data intensive projects.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### ORCID iD

R. David Parker https://orcid.org/0000-0002-1955-7445

### References

Abram, M. D. (2018). The role of the registered nurse working in substance use disorder treatment: A hermeneutic study. *Issues in Mental Health Nursing*, 39(6), 490–498. https://doi.org/10.1080/01612840.2017.1413462

Alsawas, M., Alahdab, F., Asi, N., Li, D. C., Wang, Z., & Murad, M. H. (2016). Natural language processing: Use in EBM and a guide for appraisal. *Evidence Based Medicine*, 21(4), 136–138. https://doi.org/10.11136/ebmed-2016-110437

Alsuhaibani, M., Bollegala, D., Maehara, T., & Kawarabayashi, K. I. (2018). Jointly learning word embeddings using a corpus and a knowledge base. *PLoS One*, 13(3), e0193094. https://doi.org/10.1371/journal.pone.0193094

Amazon Transcribe Pricing. (2020). *Amazon web services*. https://aws.amazon.com/transcribe/pricing/

Atlas.ti Qualitative Data Analysis Shop. (2020). https://atlasti.cleverbridge.com/74/purl-order

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(January), 993–1022.

Carminati, L. (2018). Generalizability in qualitative research: A tale of two traditions. *Qualitative Health Research*, 28(13), 2094–2101. https://doi.org/10.1177/1049732318788379

Challenges in Irreproducible Research. (2018). *Nature*(Special). https://www.nature.com/collections/prbtkwmwv

Desjardins, J. (2019). How much data is generated each day? https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bbdf29f/

Ding, J., Tarokh, V., & Yang, Y. (2018). Model selection techniques: An overview. *IEEE Signal Processing Magazine*, 35(6), 16–34.

Guetterman, T. C., Chang, T., DeHonckheere, M., Basu, T., Scruggs, E., & Vydiaswaran, V. G. V. (2018). Augmenting qualitative text analysis with natural language processing: Methodological study. *Journal of Medical Internet Research*, 20(6), e231. https://doi.org/10.2196/jmir.9702

Hong, N., Wen, A., Mojarad, M. R., Sohn, S., Liu, H., & Jiang, G. (2018). Standardizing heterogeneous annotation corpora using HL7 FHIR for facilitating their reuse and integration in clinical NLP. *AMIA Annual Symposium Proceedings*, 2018, 574–583.
Hong, N., Wen, A., Shen, F., Sohn, S., Wang, C., Liu, H., & Jiang, G. (2019). Developing a scalable FHIR-based clinical data normalization pipeline for standardizing and integrating unstructured and structured electronic health record data. *JAMIA Open, 2*(4), 570–579. https://doi.org/10.1093/jamiaopen/ooz056

Koleck, T. A., Dreisbach, C., Bourne, P. E., & Bakken, S. (2019). Natural language processing of symptoms documented in free-text narratives of electronic health records: A systematic review. *Journal of the American Medical Informatics Association, 26*(4), 364–379. https://doi.org/10.1093/jamia/ocy173

Konovalov, S., Scotch, M., Post, L., & Brandt, C. (2010). Biomedical informatics techniques for processing and analyzing web blogs of military service members. *Journal of Medical Internet Research, 12*(4), e45. https://doi.org/10.2196/jmir.1538

Lee, K. C., Udelsman, B. V., Streid, J., Chang, D. C., Salim, A., Livingston, D. H., Lindvall, C., & Cooper, Z. (2019, September 25). Natural language processing accurately measures adherence to best practice guidelines for palliative care in trauma. *Journal of Pain and Symptom Management, 59*(2), 225–232. e2.

Matsutani, T., Ueno, Y., Fukunaga, T., & Hamada, M. (2019). Discovering novel mutation signatures by latent Dirichlet allocation with variational Bayes inference. *Bioinformatics, 35*(22), 4543–4552. https://doi.org/10.1093/bioinformatics/btz266

Miller, D. D., & Brown, E. W. (2018). Artificial intelligence in medical practice: The question to the answer? *The American Journal of Medicine, 131*(2), 129–133. https://doi.org/10.1016/j.amjmed.2017.10.035

Morrow, R., Rodriguez, A., & King, N. (2015). Caliazi’s descriptive phenomenological method. *The Psychologist, 28*(8), 643–644.

Neuraz, A., Looten, V., Rance, B., Daniel, N., Garcelon, N., Llanos, L. C., Burgun, A., Burgun, A., & Rosset, S. (2019). Do you need embeddings trained on a massive specialized corpus for your clinical natural language processing task? *Studies in Health Technology and Informatics, 264*, 1558–1559. https://doi.org/10.3233/shti190533

Python (Version 3.7.4). (2020). [64 bit]: Python foundation. https://www.python.org/

Renz, S. M., Carrington, J. M., & Badger, T. A. (2018). Two strategies for qualitative content analysis: An intramethod approach to triangulation. *Qualitative Health Research, 28*(5), 824–831. https://doi.org/10.1177/1049732317753586

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development, 3*(3), 210–229. https://doi.org/10.1147/rd.33.0210

Sievert, C., & Shirley, K. (2014, jun). *LDAvis: A method for visualizing and interpreting topics* [Paper Presentation]. Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, Baltimore, MD, United States.

Spyder (Version 3.3.6). (2019). The scientific python development environment. Retrieved from https://www.spyder-ide.org/