Natural Language Processing and Machine Learning Approach to Teaching and Learning Research Philosophies and Paradigms

Marcia Mkansi and Ntombi Mkalipi
University of South Africa, South Africa

Marcia.mkansi@gmail.com (corresponding author)
36898570@mylife.unisa.ac.za

Abstract: The paper cogitates on the critical advent of 4th IR focusing on the concept of machine learning (ML) underpinned by natural language processing (NLP) to demonstrate how research philosophies and paradigms can be better taught and learned for students' benefit. A systematic literature review was earmarked for its depth and textual inquiry from scholarly arguments. The main purpose of this paper is to aver a progressive technological approach towards better comprehensive research paradigms and philosophies, which are complex domains with diverse variety in higher education and a cause of discomfort for students at post-graduate levels. Using quantitative algorithm, the natural language processing and machine learning-inspired digital model poses questions that place students in a reflexive mode and draws their articulated responses as inputs that model their worldviews against a host of philosophies in the database. The paper revealed that, discordant with previous scholars who advocated for a single philosophical assumption for a field, subject or researcher, such as the existence of a pure positivist and/ or pure interpretivist, purist philosophical assumptions should be challenged to benefit students and academics. It means that the digital discovery of research philosophies and paradigms extends the work of previous theorists to the technologically inspired discovery of episteme, ontology and axiology. By its nature, the use of NLP becomes an advanced channel on how we know what we know and the nature of the reality and values being displayed. The paper contributes to the evocation of deep learning arising from new philosophies and methods. The inquiry-based teaching approach transforms learning from the generic push-teaching method that assumes universality to the fostering of a reflexive approach that helps resolve the deep ideological approaches that caused the polarisation. The manner in which NLP and ML are able to extract information relevant to knowledge or philosophical discovery paves the way for approaches that can lead to the depolarisation and decolonisation of research philosophies, which can ultimately boost the development of research students.

Keywords: Research philosophies and paradigms, Natural language processing, Teaching and learning, Research methodology, Machine learning

1. Introduction

Research paradigm advocates (Alvesson and Skoldberg, 2009; Guba and Lincoln, 1994; Killam, 2013; Polit and Beck, 2008; Saunders et al., 2012; Steen and Roberts, 2011) can attest that to understand research, one must examine the philosophies that underlie the researcher's paradigm. This chain of evidence encourages academics to observe the same phenomena in different ways, and subsequently, to derive different kinds of knowledge from the various different philosophical perspectives (Blaikie, 1993, 2000; Guarino, 1998). However, the knowledge-creation process is replete with vast array of data that is hard to capture in a single mind, book, journal or encyclopaedia. Furthermore, the dichotomies that exist because of the preferential philosophical approaches of supervisors can also impact students' research development (Acheampong et al., 2015; Silverman, 2010). The dichotomies and data of the research approaches are amplified by the parallels and dilemmas of the concepts, which make it hard for students to comprehend the research philosophies, paradigms and their role in knowledge generation (Mkansi and Acheampong, 2012; Scotland, 2012; Sefotho, 2015). The dichotomies and dilemmas are especially applicable to research at the post-graduate level, and research has revealed the undesirable picture of a near-total absence of understanding among the students in relation to the philosophies and paradigms (Hesse-Biber, 2015; Scotland, 2012).

Although the above-mentioned scholars have revealed the breadth of the issues being experienced as challenges in the teaching, learning, and understanding of research philosophies at post-graduate level, little, if any, efforts have been made to minimise the challenges by exploring some of the latest technologies. In light of the observed gap, how can technology help lecturers and students at post-graduate level to teach and learn a complex subject with different types of philosophies? Also, how can technology contribute to Te'eni et al.'s (2015) genre of knowledge discovery-based research methodologies, and Davison and Martisons' (2016) call for an inclusive approach to the development of new theories?

Reference this paper: Mkansi, M. and Mkalipi, N., 2023. Natural Language Processing and Machine Learning Approach to Teaching and Learning Research Philosophies and Paradigms *The Electronic Journal of Business Research Methods*, 21(1), pp. 14-30, available online at www.ejbrm.com

In an effort to address the above two questions, and also to improve the knowledge and understanding of research philosophies and paradigms at post-graduate level, this study use natural language processing (NLP) and machine learning (ML) (which are technological approaches) to introduce a way of simplifying the teaching and learning of research philosophies. NLP present opportunities for the comprehension and processing of human language and its translation into machine text whilst ML provides opportunities for comprehension and processing, in great detail, of research information and data (big data) that is beyond the command of a lecturer, book, journal, and/or student. It is the latter capabilities that have foregrounded the quantitative algorithm underlying the digital-based model in this study. This digital-based model processes a range of qualitative questions based on epistemological, ontological and axiological assumptions derived from consolidated literature. The systematic literature review became a compelling method for analysis, and the NLP and ML demonstrated the possibilities of achieving the objectives of the research. The manner in which NLP and ML are able to extract information relevant to knowledge or philosophical discovery paves the way for approaches that can lead to the depolarisation and decolonisation of research philosophies, which can ultimately boost the development of research students. This study underpins research methodology in business and management through the capability of technology in augmenting human knowledge using algorithms that have the capabilities to extract and process information from a large database, transforming it into deep learning that can guide decision making and knowledge production for scholars and students engaging in research for knowledge generation output. The advancement of the business and management research methodology is aligned with modern technology advancement and reduces the stereotyping and over-compensated positivist and interpretivist philosophies towards broader reflexive modes and plenitudes of philosophies and paradigms - including newly developed philosophies and paradigms most suitable for business and management.

The following section presents a background to research philosophies and paradigm challenges vis-a-vis and contemporary views on teaching and learning to dovetail with NLP and ML-based philosophical and paradigm positioning. Section 3 presents the methods of dataset collection and the analysis and annotation process that underpins the digital-based model architecture's performance, input and output data. The final section provides insights into the implication of the model and future research.

2. Philosophy and Paradigm Challenges

The theoretical background to research philosophies and paradigms gained momentum in the 13th century (Jensen, 2000; Lincoln, 1990). Eminent scholars of the time, such as John Locke, David Hume, René Descarte, Immanuel Kant and Hans Reichenback were some of the pioneers of philosophical assumptions; they also acted as advocates for knowledge generation and played a role in the development of research methods (Denzin and Lincoln, 2003; Killam, 2013; Miller and Grimwood, 2015; Saunders et al., 2019). However, since the origination of the concepts, there has been little progress towards technological advancement. This applies especially to those concepts that are reportedly difficult, and that consist of large amounts of data with many variations, such as research philosophies and paradigms, which causes, in effect, a living gulf between theory and practice.

In an effort to expose the gap, several scholars (Hesse-Biber, 2015; Scotland, 2012; Sefotho, 2015) revealed a series of issues related to understanding, and they aimed to raise awareness about the concepts and the challenges being experienced in connecting the relationship between research philosophical assumptions, methods and knowledge. For example, Acheampong et al. (2015) and Silverman (2010) revealed dichotomies in research methods that are due to the preferential philosophical approaches of supervisors, and which highlight the potential impact on students' ability to do sound research. A study by Hesse-Biber (2015) on the problems and prospects in the teaching of research revealed the pedagogical challenges students faced in understanding how research methods connect to philosophical assumptions, and at worst, some of the students did not know that they have philosophical assumptions.

Makombe's (2017) investigation into the relationship between research philosophies, methods and design reported that a good percentage of researchers avoid the discussion of their guiding paradigm due to a lack of knowledge related to the topic. Mkansi and Acheampong (2012), and Mackenzie and Knipe (2006) reported on the students' dilemmas in understanding the research philosophies and paradigms; problems that are caused and amplified by the incomprehensible classifications and contradictory terminology of philosophies. In emphasising the importance of research paradigms and philosophies, Sefotho (2015) suggested that the philosophy must be the guiding force that drives the researcher in developing a thesis, and if absent, the researcher's investigation is directionless. Mkansi (2018), similarly revealed a lack of application or acknowledgement of research philosophies in high impact knowledge hubs.

In a review of the gaps in technology aimed at improving the teaching and learning of research methods, this study found several technological advancements that complement qualitative and quantitative data analysis techniques such as SPSS (Field, 2009), Analysis of Moment Structures (Amos) (Arbuckle (2008), Maxqda (Saillard, 2011; Faherty, 2010) and Nvivo (Faherty, 2010). There are few exceptions that explored and studied the NLP and ML application to research methods. For example, Chen et al. (2021) demonstrates how and why machine learning can advance social science from analysing causality and correlations to prediction. Whereas Chang et al. (2021) use NLP to illustrate its strength in rapid analysis of big mixed methods data in times of catastrophic change. In addition, Javed et al. (2021) evaluates the objectivity of ML as a field of study. Although these studies make significant headways in their application of NLP and ML to research methodologies, their focus is mostly limited to data analysis of primary qualitative, quantitative and mixed methods, rather than a diagnostic inquiry-based learning of research philosophies and paradigms.

Beyond application to research methods, Kanchan and Yadav's (2022) systematic review of NLP research conducted from 1997 to 2021, revealed the application of sentiment analysis in many business spheres that includes product reputations to customers reviews etc., but none on the aspect of research philosophies and paradigms. Similarly, Zhao et al.'s (2021) report of 404 studies that used NLP for requirements engineering excludes matters of research philosophies and paradigms. Viewed differently, although contemporary applications of NLP to teaching and learning of research methods exist, they serve as forms of advanced alternatives or compliments of other technology-based analysis tools. Thus, research philosophies remain complex domains in higher education (Muhaise et al., 2020; Groeesler, 2017) and a cause of discomfort for students at post-graduate levels. While research competencies are becoming increasingly invaluable for employability, there is insufficient research on innovative pedagogical approaches to support students in this aspect (Daniel, 2018). Hence, this study's consideration of NLP and ML techniques. A review of the defined NLP and ML is presented in subsequent sections and its application in the study.

2.1 Natural Language Processing

Research is a field that is characterised by an unprecedented amount of data, ranging from 'episteme' or knowledge, and 'ontologia' as an inner cognitive state that is connected to, among others, observations, experiments, abstracts, narratives, interpretations of experiences, and conceptualised practice (case studies). Some of the research concepts, such as philosophies and paradigms, consist of a large, diverse amount of data and variations that have blurred the similarities and differences (see example screenshot of some of the philosophes in Figure 2). There is such a vast number of variations and large amount of data related to research philosophies and paradigms, and to complicate matters it is not available in a single book, journal or review, which leads to the polarisation and limited development of research students.

Natural language processing presents opportunities for the comprehension and processing, in great detail, of research information and great amounts of data that is beyond the comprehension of a lecturer, book, journal, and/or student. The relevance of NLP lies in its ability to analyse and help machines understand the nuances in human languages (Marr, 2016). NLP is concerned with the processing of written and spoken human language through computational techniques, and is equally dependent on machine learning (ML) in its pursuit of helping machines comprehend the human language (Marr, 2016). In this regard, it is necessary to capture the user's expression of their worldview through a series of questions that serve to present a specific customised representation of the user's research philosophical assumptions beyond the cluster-level results from ML.

There are various types or concepts of NLP, for example, there is natural language understanding (NLU), or natural language translation, which are both concerned with the comprehension or understanding of human language and its translation into machine text (Bonaccorso, 2017; Kaminski, 2017). Furthermore, natural language generation (NLG) generates human language text from machine text, or numbers, by making choices based on the grammatical content, correctness and readability of the generated language (Bonaccorso, 2017; Kaminski, 2017). Information extraction (IE) has also become popular with the rise of social media, and is mainly used for sentiment analysis (Derczynski et al., 2014; Jiang, 2012). This study adopted the NLU concept for its strength in capturing an unstructured understanding of users' inputs or texts of their worldview, and its translation into a supervised ML algorithm.

A search of the literature did not reveal any previous work done in relation to the use of NLP for the prediction or classification of user ideas and concepts into research philosophies and paradigms. However, it is noted that NLP has been used in marketing research (Leeson et al., 2019; Yu and Kwok, 2011), aviation (Kumar and Zymbler, 2019, and education (Waters et al., 2017;). The evident lack of NLP application in research philosophies provided the impetus for addressing Davison and Martinson's call for new ways of developing theories, including those emanating from indigenous backgrounds.

The NLP procedure involves the following procedures: tokenising, part of speech tagging (PoS Tagging), word embeddings, stemming and lemmatisation, and named entity recognition (NER) (see screenshot of Figure 3 in the methodology section).

- Tokenising uses a lexer (lexical analysis) to identify and separate instances of a sequence of characters or words, referred to as tokens, in a given sentence (Zhao and Kit, 2011). These tokens are used as input for the process of parsing or text mining. Parsing defines the grammatical rules for the tokens and the relationship between them in an abstract form by producing a dependency tree (Zeroual and Lakhouaja, 2018).
- Part of speech tagging (PoS) assigns morpho-syntactical features to words based on their context, thus enabling simple syntactic searches.
- Lemmatisation reduces words to their dictionary form, taking into account the meaning of words in sentences or nearby sentences, whereas stemming establishes relationships between words by reducing them to their basic or root form (Pedrycz and Chen, 2016).
- Named entity recognition involves the extraction of specific words or entities from within text. These entities are identified and linked to a category instance in a knowledge base to resolve their contextual meaning (Chang et al., 2016).
- Word embeddings, which is a distributed representation of text, is often used to overcome the weakness computers have in processing natural language by mapping words to vectors of numerical values, and also to map the relationships between words by creating similar representations for words with similar meaning (Li and Sha, 2017).

2.2 Machine Learning

Machine learning presents opportunities for the comprehension and processing, in great detail, of research information and data (big data) that is beyond the comprehension of a lecturer, book, journal, and/or student. The strength of ML is in augmenting human knowledge, using algorithms that have the capabilities to extract and process information from a large database (Carpenter, 2019), transforming it into deep learning that can guide decision-making and knowledge production. Yet, such ML opportunities remain unexplored in the research methods' space.

Machine learning can be classified into unsupervised, supervised, and semi-supervised learning (Brunton et al., 2019). Supervised learning algorithms (for example, naïve Bayes; LR; SVM; regressions such as linear, and the gaussian process; classifications, such as decision trees, and random forests; optimisation and control techniques, such as genetic algorithms, and deep model predictive controls) build models by learning relationships between descriptive features (input) and target features (output) based on historic datasets (Kelleher et al., 2015). The algorithm is trained by supplying it with known inputs and their matching responses, and from the learned relationship it can predict responses for unknown inputs (Shouval et al., 2013). Unsupervised learning machines require no supervision as they are capable of discovering information independently, through the use of, for example, clustering techniques such as k-means, spectral clustering, and dimensionality reductions such as autoencoders and diffusion maps (Brownlee, 2018). Semi-supervised techniques, such as reinforcement learning (Q-learning, Markov decision process, etc.), and generative models, such as generative adversarial networks, learn with partially labelled data (Brunton et al., 2019). This study adopted the supervised ML concept to complement NLP.

3. Methodology

This study used NLP and ML to discover research philosophies and paradigms. The class of techniques in use belongs to computer science and is highly quantitative in nature. Quantitative research deals with numeric data (Creswell, 2013, 2014; Saunders et al., 2016), which for the purposes of this study were mainly NLP algorithms. The study combined systematic literature and algorithm methodology to investigate how NLP can help lecturers and students at post-graduate level to teach and learn a complex subject (namely, the subject of research philosophies and paradigms) which has more than 180 different types of philosophies collected from various sources of literature.

The methodology embodies the research philosophies and paradigms artefacts developed for this study, which contributes to Peffers et al.'s (2018) quest for the acknowledgement of design science research in information systems. In addition, it demonstrates the approaches that can lead to a depolarisation of research philosophies that can boost the development of research students and knowledge production. The systematic literature review provided contextual data (for the database), and the variables crucial for gathering input data from users based on the multidimensional philosophical assumptions offered by Saunders et al. (2016) and Guba and Lincoln (1994).

As such, the next section presents the methods for data collection, followed by a discussion of the data annotation process. The systematic literature review represents the foundation of the model architecture developed. Put differently, the research philosophies and paradigms (RPP) model's contextual database is grounded on scientific knowledge that stretches from the early centuries to date.

3.1 Data set and Collection

In gathering and preparing the data set to serve as contextual data (database) and which would also be used for deriving the input questions, the present study followed Denyer and Transfield's (2000) principles of a systematic literature review (see Figure 1). The principle of a systematic literature review involves scoping the subject matter by using keywords (for example, research, philosophies, paradigms, interpretivism, positivism, etc.) to search for relevant subject matter across a host of information hubs (journals, reviews, encyclopaedia, conferences, etc.). Information relevant to the literature review was published in the International Scientific Indexing (ISI), Scopus list, Science Electronic Library Online (SciELO) and the International Bibliography of Social Sciences (IBSS), dating back from the early centuries up to the year 2020. In addition, the researchers consulted research methods textbooks available from various publishing houses and google scholars, including those by popular scholars, such as Creswell (2014), and Saunders et al. (2016). The in-depth search of the information hubs produced a total of 180 research philosophies. However, most of the philosophies did not have enough reviews and data, which proved difficult for the NLP algorithm. As such, the research philosophies with limited information were excluded, leaving a total of 180 usable research philosophies for input to the system.

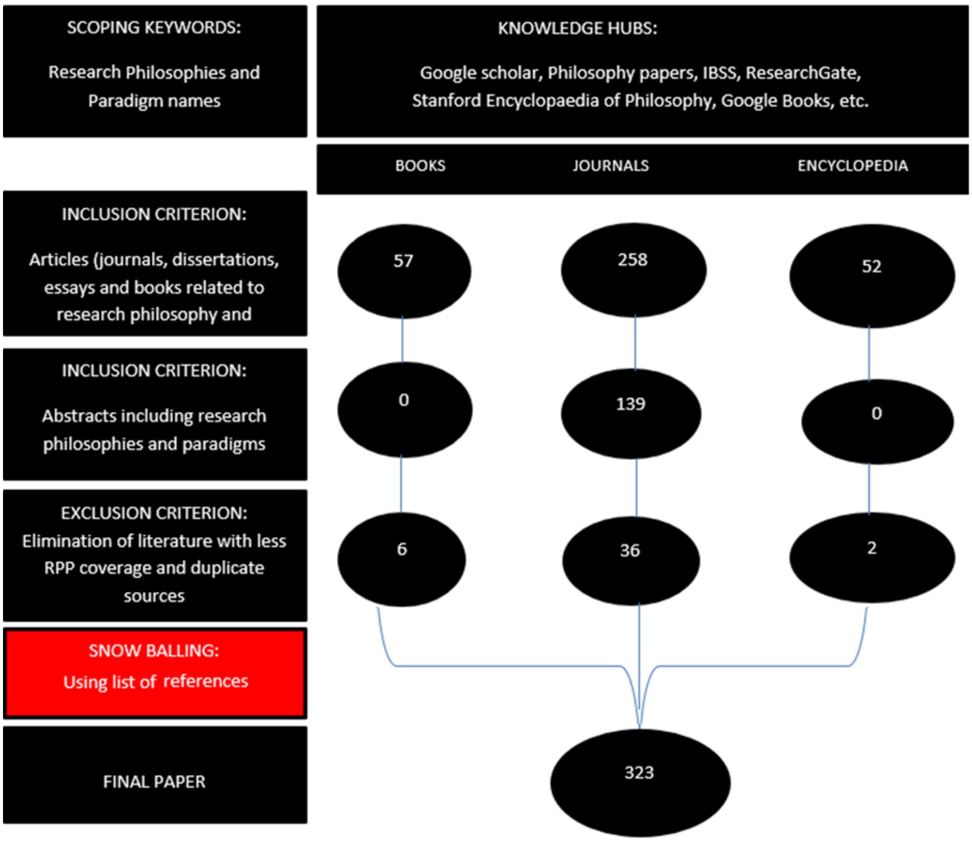

SCOPING KEYWORDS:
Research Philosophies and Paradigm names
KNOWLEDGE HUBS:
Google scholar, Philosophy papers, IBSS, ResearchGate, Stanford Encyclopaedia of Philosophy, Google Books, etc.
BOOKS
JOURNALS
ENCYCLOPEDIA
INCLUSION CRITERION:
Articles (journals, dissertations, essays and books related to research philosophy and
57
258
52
INCLUSION CRITERION:
Abstracts including research philosophies and paradigms
0
139
0
EXCLUSION CRITERION:
Elimination of literature with less RPP coverage and duplicate sources
6
36
2
SNOW BALLING:
Using list of references
FINAL PAPER
323

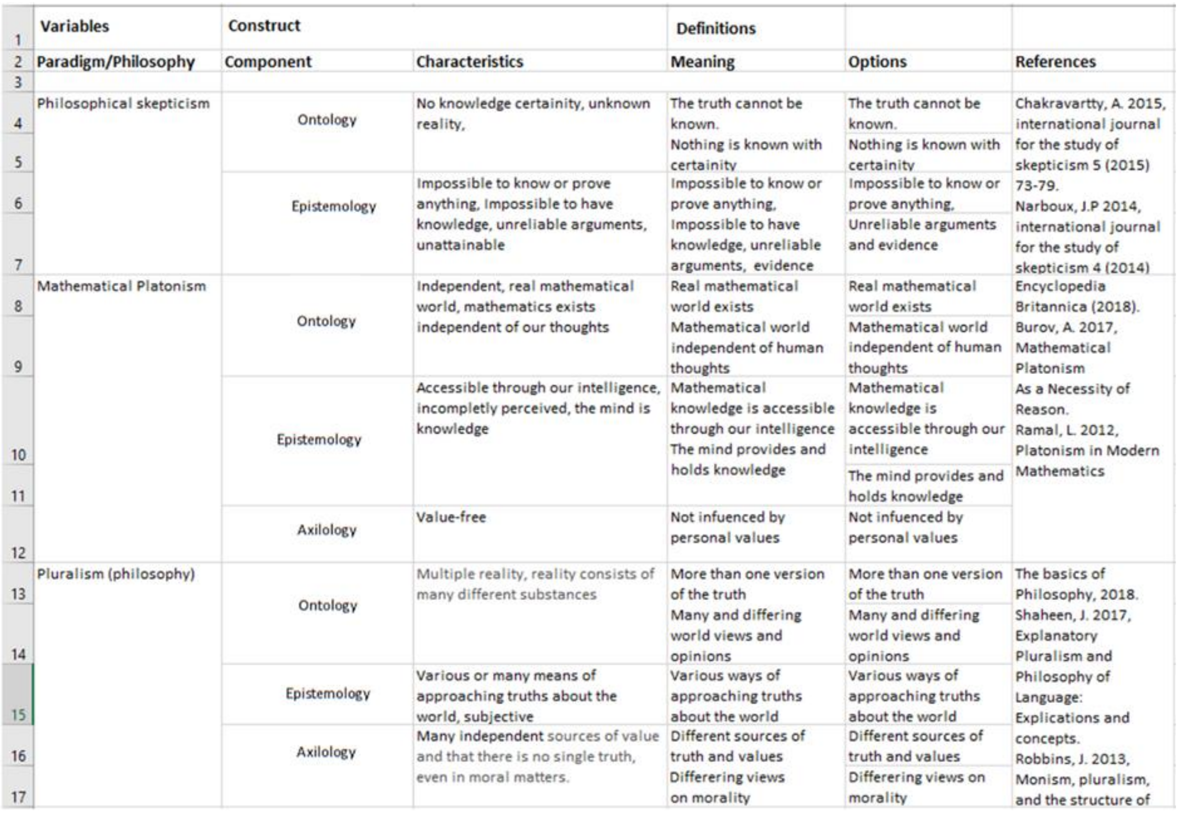
Figure 1: Process of Systematic Data Collection of Research Philosophies and Paradigms

(Adapted from Denyer and Transfield, 2000)

The literature findings supported the development of the model in terms of the contextual data (database) and the design of questions that reflect the multidimensional structure of a research philosophy based on the various ontological, epistemological, and axiological stances. In particular, data derived from the literature sources was crucial in grounding both the RPP model and the study's objectives on credible scientific knowledge, and for identifying relevant factors from the different sources of data. The relevance of data sources in achieving conformability and credibility, whilst limiting subjectivity is highly endorsed by Saunders et al. (2019) and Yin et al. (2018) (see Appendix 1 for a sample corroboration of research philosophies' data against different sources of literature).

3.2 Data Analysis and Annotation

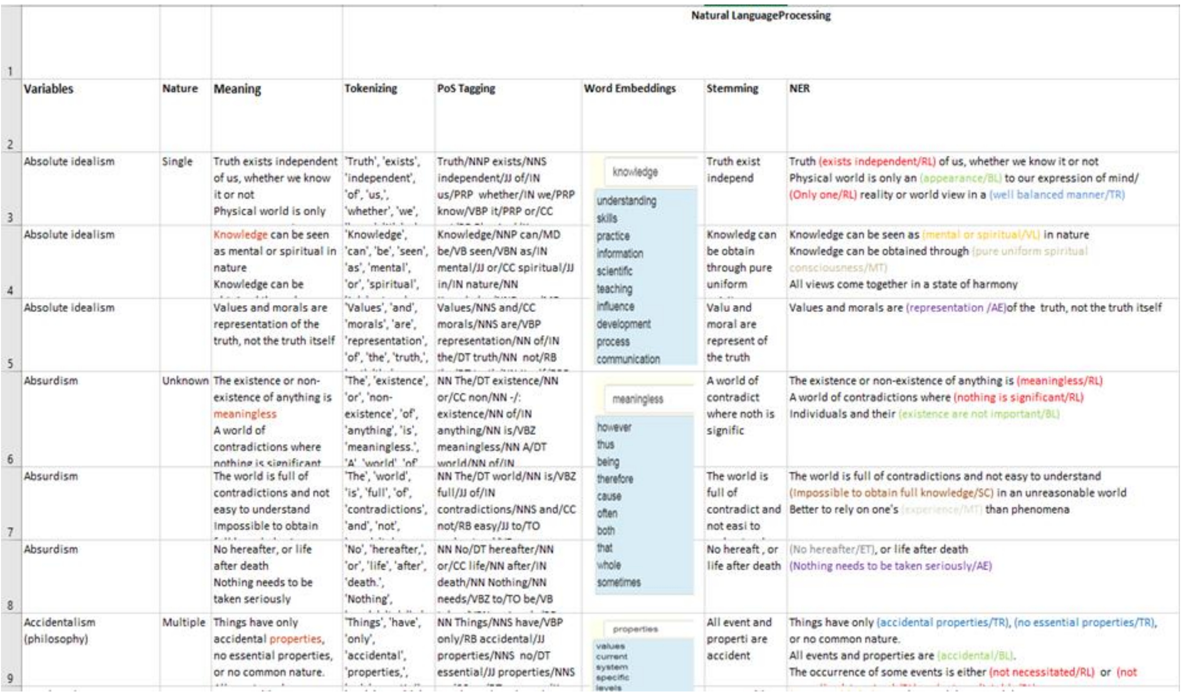
The data was first analysed using a summative content analysis which involved counting the number of scholarly articles that included research philosophies' keywords, including the interpretation of their respective underlying context, as endorsed by Hsieh and Shannon (2005). The analysis focused on the meaning, characteristics, and the options emanating from the meaning, which was useful for the NLP analysis. A Microsoft Excel spreadsheet was used to organise each philosophy within the epistemological, ontological, and axiological lenses or framework, as offered by Saunders et al. (2016), Seddon and Scheepers (2012), and Guba and Lincoln (1994) (see example screenshot of Figure 2). This firstly, involved the Microsoft Excel spreadsheet with the content analysis of the streamlined 180 research philosophies to create a corpus of RPPs (see Figure 2).



1	Variables	Construct		Definitions		
2	Paradigm/Philosophy	Component	Characteristics	Meaning	Options	References
3						
4	Philosophical skepticism	Ontology	No knowledge certainity, unknown reality,	The truth cannot be known. Nothing is known with certainity	The truth cannot be known.	Chakravartty, A. 2015, international journal for the study of skepticism 5 (2015) 73-79. Narboux, J.P 2014, international journal for the study of skepticism 4 (2014)
5					Nothing is known with certainity	
6		Epistemology	Impossible to know or prove anything, Impossible to have knowledge, unreliable arguments, unattainable	Impossible to know or prove anything, Impossible to have knowledge, unreliable arguments, evidence	Impossible to know or prove anything,	
7					Unreliable arguments and evidence	
8	Mathematical Platonism	Ontology	Independent, real mathematical world, mathematics exists independent of our thoughts	Real mathematical world exists Mathematical world independent of human thoughts	Real mathematical world exists	Encyclopedia Britannica (2018). Burov, A. 2017, Mathematical Platonism As a Necessity of Reason. Ramal, L. 2012, Platonism in Modern Mathematics
9					Mathematical world independent of human thoughts	
10		Epistemology	Accessible through our intelligence, incompletly perceived, the mind is knowledge	Mathematical knowledge is accessible through our intelligence The mind provides and holds knowledge	Mathematical knowledge is accessible through our intelligence	
11					The mind provides and holds knowledge	
12		Axilology	Value-free	Not infuenced by personal values	Not infuenced by personal values	
13	Pluralism (philosophy)	Ontology	Multiple reality, reality consists of many different substances	More than one version of the truth Many and differing world views and opinions	More than one version of the truth	The basics of Philosophy, 2018. Shaheen, J. 2017, Explanatory Pluralism and Philosophy of Language: Explications and concepts. Robbins, J. 2013, Monism, pluralism, and the structure of
14					Many and differing world views and opinions	
15		Epistemology	Various or many means of approaching truths about the world, subjective	Various ways of approaching truths about the world	Various ways of approaching truths about the world	
16		Axilology	Many independent sources of value and that there is no single truth, even in moral matters.	Different sources of truth and values Differering views on morality	Different sources of truth and values	
17					Differering views on morality	

Figure 2*: Sample Multidimensional Philosophical Assumptions Summative Content Analysis*

Following the presentation of research philosophies and paradigm and their associated meaning, the study carried out data annotation. Data annotation commonly involves including parts of speech (POS), or tags, which label each word in terms of its grammatical category (Reppen, 2010). Data tags are crucial for addressing data inputs from users of RPP model and helps to eliminate issues associated with searching for polysemous words (i.e. can be used as a modal verb or a noun) to disambiguate and focus the search results (see Figure 3) below for a sample screenshot of data annotation). The labels or annotated data are crucial for training data using the BoW (bag-of-words) and machine learning algorithm.



					Natural LanguageProcessing			
1								
2	Variables	Nature	Meaning	Tokenizing	PoS Tagging	Word Embeddings	Stemming	NER
3	Absolute idealism	Single	Truth exists independent of us, whether we know it or not Physical world is only	'Truth', 'exists', 'independent', 'of', 'us,', 'whether', 'we',	Truth/NNP exists/NNS independent/JJ of/IN us/PRP whether/IN we/PRP know/VBP it/PRP or/CC	knowledge understanding skills practice information scientific teaching influence development process communication	Truth exist independ	Truth (exists independent/RL) of us, whether we know it or not Physical world is only an (appearance/BL) to our expression of mind/ (Only one/RL) reality or world view in a (well balanced manner/TR)
4	Absolute idealism		Knowledge can be seen as mental or spiritual in nature Knowledge can be	'Knowledge', 'can', 'be', 'seen', 'as', 'mental', 'or', 'spiritual',	Knowledge/NNP can/MD be/VB seen/VBN as/IN mental/JJ or/CC spiritual/JJ in/IN nature/NN		Knowledg can be obtain through pure uniform	Knowledge can be seen as (mental or spiritual/VL) in nature Knowledge can be obtained through (pure uniform spiritual consciousness/MT) All views come together in a state of harmony
5	Absolute idealism		Values and morals are representation of the truth, not the truth itself	'Values', 'and', 'morals', 'are', 'representation', 'of', 'the', 'truth,',	Values/NNS and/CC morals/NNS are/VBP representation/NN of/IN the/DT truth/NN not/RB		Valu and moral are represent of the truth	Values and morals are (representation /AE)of the truth, not the truth itself
6	Absurdism	Unknown	The existence or non-existence of anything is meaningless A world of contradictions where nothing is significant	'The', 'existence', 'or', 'non-existence', 'of', 'anything', 'is', 'meaningless.', 'A', 'world', 'of'	NN The/DT existence/NN or/CC non/NN -/: existence/NN of/IN anything/NN is/VBZ meaningless/NN A/DT world/NN of/IN	meaningless however thus being therefore cause often both that whole sometimes	A world of contradict where noth is signific	The existence or non-existence of anything is (meaningless/RL) A world of contradictions where (nothing is significant/RL) Individuals and their (existence are not important/BL)
7	Absurdism		The world is full of contradictions and not easy to understand Impossible to obtain	'The', 'world', 'is', 'full', 'of', 'contradictions', 'and', 'not',	NN The/DT world/NN is/VBZ full/JJ of/IN contradictions/NNS and/CC not/RB easy/JJ to/TO		The world is full of contradict and not easi to	The world is full of contradictions and not easy to understand (Impossible to obtain full knowledge/SC) in an unreasonable world Better to rely on one's (experience/MT) than phenomena
8	Absurdism		No hereafter, or life after death Nothing needs to be taken seriously	'No', 'hereafter,', 'or', 'life', 'after', 'death.', 'Nothing',	NN No/DT hereafter/NN or/CC life/NN after/IN death/NN Nothing/NN needs/VBZ to/TO be/VB		No hereaft , or life after death	(No hereafter/ET), or life after death (Nothing needs to be taken seriously/AE)
9	Accidentalism (philosophy)	Multiple	Things have only accidental properties, no essential properties, or no common nature.	'Things', 'have', 'only', 'accidental', 'properties,',	NN Things/NNS have/VBP only/RB accidental/JJ properties/NNS no/DT essential/JJ properties/NNS	properties values current system specific levels	All event and properti are accident	Things have only (accidental properties/TR), (no essential properties/TR), or no common nature. All events and properties are (accidental/BL). The occurrence of some events is either (not necessitated/RL) or (not

Figure 3: Example of Data Annotation Using NLP Methodology Application to RPP

3.3 RPP Model Architecture

The architecture of the RPP model consists of four major components, namely: input data (client side), RPP data from literature or context data (database on the server), bag of words (BoW) and ML algorithm (server), and output data detailed in the corresponding sections (accessed from client side, retrieved from the database). Figure 4 provides an understanding of the behavior of the system and how it performs the classification of a researcher's sentiment into research philosophy and paradigm categories.

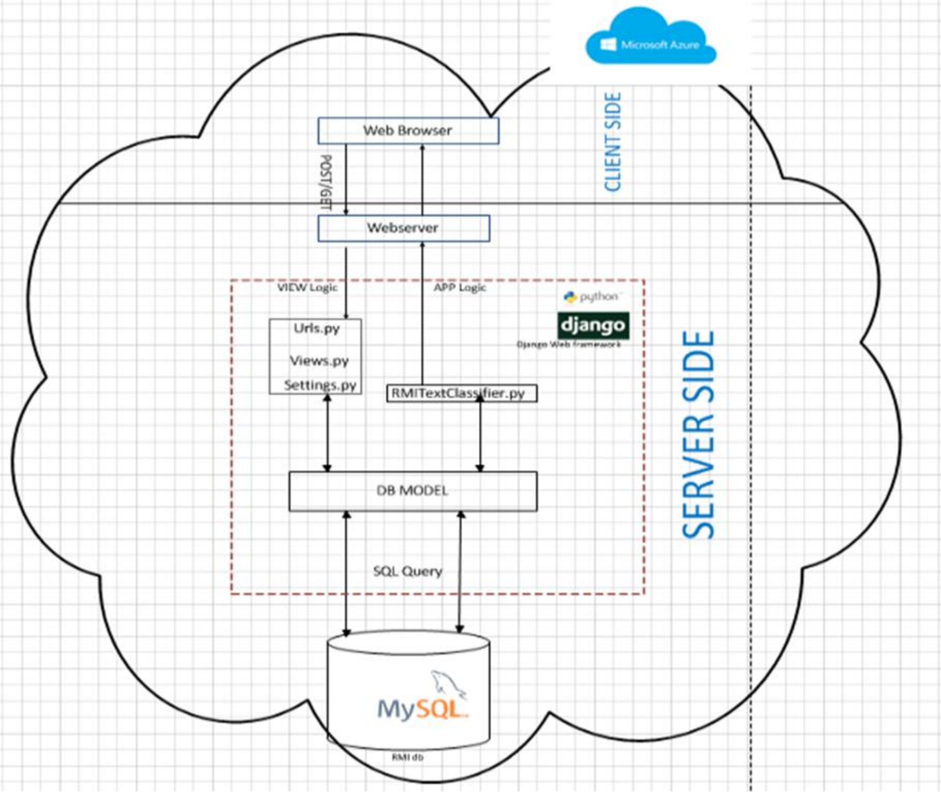


Figure 4: Research Methods Index NLP Architecture Diagram

The NLP application is hosted on the Microsoft Azure public cloud computing platform, as shown in Figure 4, for ease of access from multiple locations. The study adhered to the architectural style referred to as Representational State Transfer (RESTful) client-server design in deploying the NLP application. Python was used to write NLP code that is used for the classification of input. The Django Web framework was used for constructing the NLP web API, and was also used for the passing of data between the user interface and the back-end. The Django Web framework was also used for the development of the user interface. The Django Web framework supports the Model View Template (MVT) pattern, where the developer provides the model and Django uses the template and views to map the model to the Uniform Resource Locator (URL). The Django Web framework then renders the URL to the user interface.

The MySQL database is used to store information on research philosophies and paradigms, user account information, user input, system roles and reports. A user will be provided with a link to the system which will allow them to register an account, answer the questionnaire and then view their report.

3.4 Input Data (Client or User Side)

Input data for the NLP RPP model is derived from users' response to questions that represents the different variables of the multidimensional philosophical assumptions within the epistemological, ontological, and axiological lenses or framework, as offered by Saunders et al. (2016) and Guba and Lincoln (1994). The users' responses describe the specific worldview needed to run the RPP model. For example, the following questions are representative of the philosophical structure of the multidimensional assumptions by former scholars and are asked to gather inputs:

1. Questions based on the premise of ontology. The premise of these questions is to explore the user's ontological stance.

- How many versions of the truth can there be in a given situation?
- How can truth be influenced?

2. Questions based on the premise of epistemology. The premise of these questions is to explore the user's epistemological stance.

- How is knowledge acquired, that is, how do we know what we know?
- What influences what we know?
- How many sources of knowledge are there?
- *How can knowledge be advanced?*

3. Questions based on the premise of axiology. The premise of these questions is to explore the user's axiological stance.

- What is the importance of values and ethics?
- How can personal values influence the truth?
- What determines our values and ethics?

The user enters input data through the web browser that resides on their specific personal computer that is connected to the internet through an ISP or other means. The inputs are crucial for tokenisation and for performing the BoW and machine learning algorithm against the list of common words in the database, and to possibly identify the words or emerging worldviews that fall outside the purview of the database. A researcher uses the web browser to connect to the RMI application. The web browser is linked to the webserver through a POST/GET method.

3.5 RPP Context Data From Literature (Database on the Server)

The context data was derived from the 180 usable items of information on the different types of research philosophies collected from the various sources of literature, as previously discussed. The annotated context data is presented in the RPP model following the NLP methodology or procedure (see Figure 3). Appendix 1 provides an example of how data pertaining to each philosophy was collected and corroborated from different sources for conformability and credibility. The different sources helped not only with the study's credibility and conformability, but also provided coverage of the different emphases related to the phenomenon that was necessary to produce a rich and adequate volume of text for algorithm performance purposes. Further, corroboration of the different emphases of all the philosophies was essential for drawing more reliable and meaningful conclusions about each philosophy.

The strength of a small targeted database or (corpora) for the investigation of special uses, such as the RPP model, has been greatly discussed by Reppen (2010). The MySQL database stores all the RPP's data, participant data, information and responses which are then used to generate a report recommending research philosophy and paradigm categories to participants. The communication between the client and the webserver is through the Hypertext Transfer Protocol (HTTP). The function of the server is mainly to store, process ML algorithm discussed in the next section, and deliver web content (html pages) as requested by the client.

3.6 Classification Model: Bag of Words and Machine Learning Algorithm (Server)

The data that was sourced from journals, encyclopaedias, books, and so forth, and stored in the RPP's corpus, was split into train-test sets (at 90-10 ratio) for the purpose of training and testing the classification algorithms. According to Sebastiani (2002), ML algorithms cannot process text data directly, therefore the corpus data needed to be represented in a numerical form. The Bag-of-Words (BoW) model was used to pre-process and represent the RPP's corpus data in numeric vectors. This model is based on a vocabulary and measure of the frequency of all known words in a corpus (Aggarwal and Zhai, 2012). The Term Frequency – Inverse Document Frequency, or TF-IDF, was then used to rescale word frequencies by computing how many times the words appear in each of the RPPs, with the most frequently appearing words, such as 'there and that', being ignored.

This then produced a final score which was used as a weight for each of the RPPs. These weights or frequencies of known words in a category are considered as features to be used as equivalent fixed-length numerical vectors (as shown in Table 1). The numerical vectors or features are used to train classifiers such as the naïve Bayes, SVM and Logistic Regression to determine which one produces the highest level of confidence in classifying text into different RPP categories. The example of BoW model presented in Table 1 is also used to represent input text in numeric vectors to be used to determine the RPP category it belongs to.

Table 1: Vector Representation of RPPs in BoW

	World	Compose	Fundamental	substance	External	nature	kind	Independent	term	Language	Object	exist	function	relate
RPP 1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
RPP 2	0	0	0	0	1	1	1	1	1	1	0	0	0	0
RPP 3	1	0	0	0	0	0	0	0	0	0	1	1	1	1

RPP1: The world is composed of two fundamental substances

RPP2: External natural kind terms independent of language

RPP3: Objects exist and function only as relational in the world

Once the user has completed the questionnaire, the input is concatenated into a string that undergoes pre-processing before being vectorised. A comparison of the features derived from user input against the corpus category features yields the three topmost RPPs that are closely linked to the input, which in turn, represents a researcher's worldview. Furthermore, the NLP is linked to a lexical database, such as Wordnet, to find synonyms for other complex philosophical terms, and the emerging constructs necessary for deep learning. The output results are presented in a chart showing the degree to which a researcher is aligned with a particular RPP.

The server side refers to the NLP application that processes input, classifies it into RPP categories and then issues a report for the participant. This report is based on the information request from the MySQL database. The study ensures the validation of annotation or labels using supervised learning which helps with label data, or semi-structured supervised learning (taking the labels we have to leverage on data without labels). The latter labels are used to infer labels to emerging data that is outside the purview of the database.

On running the Naïve Bayes, SVM and LR from the same dataset represented by the BoW model it can be observed that the Naïve Bayes outperforms all the other algorithms (see Table 2). This is due to the Naïve Bayes' ability to be linearly scalable with the data points and the number of predictor features (Keogh, 2015; Sebastiani, 2002) and its ability to provide a probabilistic explanation for classifications based on the Bayes theorem.

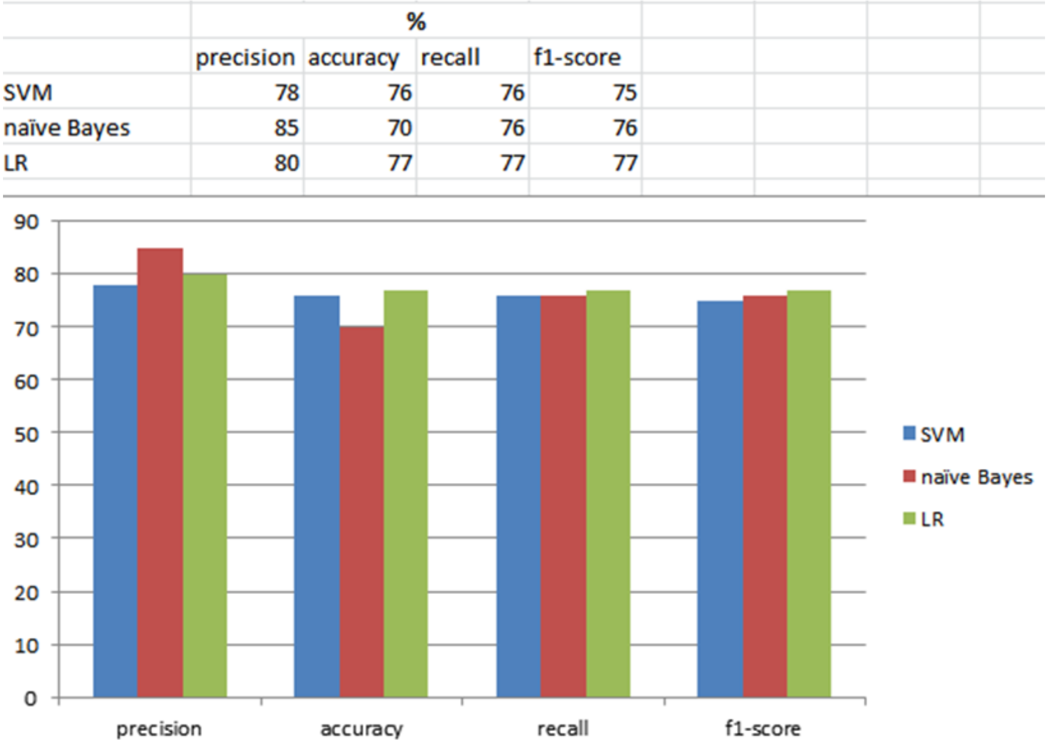
Table 2: Accuracy and Precision of NLP Algorithms

Algorithm	Accuracy	Precision rate
Naïve Bayes	70%	85%
SVM	76%	78%
LR	77%	80%

Further, the Naïve Bayes algorithm is suitable for this study due to (1) its ability to learn, even if there are small amounts of training data, thereby significantly reducing the time it takes to train, (2) it yields parameter estimates that are good, and (3) with independent predictors it performs better than the discriminative LR and SVM algorithms (Keogh, 2015; Sebasitani, 2002; Uddin et al., 2012).

As for accuracy levels, Gaizauskas and Wilks (1998) attest that the precision of the information extraction (IE) system ranges around 50%. The natural language component of the RMI reported in this study produces a 70% accuracy and 85% precision level, which lends and builds support for the previous study by Gaizauskas and Wilks (1998).

In terms of confidence level, Gandrabur et al. (2006) argue that the confidence level related to NLP is subject to data, purpose of the study, and the variations in language. According to Gaizauskas and Wilks (1998), IE is as a subset of knowledge discovery and data mining research that extracts useful aspects of textual information from natural language texts, such as the one provided by users in the study. Confidence levels are further reflected based on precision, recall, and F1-score, as shown in the comparative view of different algorithms in Figure 5.



	%			
	precision	accuracy	recall	f1-score
SVM	78	76	76	75
naïve Bayes	85	70	76	76
LR	80	77	77	77

Figure 5: Comparative View of the Algorithms

3.7 Model Assumptions

The RPP model is not without assumptions. The acknowledgement of assumptions is a practice greatly endorsed by Seddon and Scheepers (2012) who argue for boundary conditions beyond which the findings of the study might not apply. In this study, there is an assumption in the analysis that the respondents had the same level of cognition, use of English and expertise in research philosophies. There is also an inherent risk in text analysis because the same phrase can mean different things in different cultures. Put differently, text analysis is subject to imbalance and linguistic variations issues.

In addition, the considered research philosophies are (at this point of the model) limited relatively to the contextual data in the RPP database. Put differently, the RPP model at the time of writing this paper has not identified emerging or indigenous worldviews outside the database, but rather calculates the users' Bow against the stored contextual data. This is mainly because although the system is built to recognise emerging concepts beyond the database, the system only considers words outside the purview of the RPP model as an emerging philosophy, if there is a theme or high pattern from more people. In its current form, the model only highlights words that might be new to the RPP database.

The RPP model relies mainly on text, and excludes pictures, audio, video, social media and any other form of data. The selection of feature vectors based on text reliance or keywords limits the classification to text extracted from the users' inputs, thus limiting an understanding of the relationship between words. The latter is mitigated, to a certain extent, by the link between the RMI model and WordNet to try and explore the relationships between words. The domain focus of NLP to research philosophies limits its scalability.

3.8 Output Data (Accessed From Client Side, Retrieved From the Database)

This section presents the output of the realistic context of the RPP model application using NLP, the technological techniques used to present a way of making the teaching and learning of research philosophies and paradigms easier. The system was tested firstly, for performance against its ability to capture a user's exact inputs regarding his/her worldviews to align, correlate with, identify and reveal specific philosophies espoused by the user, against those collected from literature in the RPP database. Secondly, to demonstrate percentages and areas where a user's worldview correlates with more than one research philosophy. Lastly, for precision and accuracy.

Natural language processing is the most important indicator, as it gives the user freedom of expression through open-ended questions. Using the exact user's expression, the NLP processes the unstructured data using an algorithm based on the methodology previously discussed in the section on the NLP process (of tokenising, lemmetisation, and named entity recognition). Using ML algorithm, the NLP model dives deep to correlate the user's worldview beyond the preferential philosophical approaches of supervisors, subject, and field of study. By so doing, NLP provides an interesting approach to resolving the deep ideological approaches which result in polarisation.

The output data presents the following four categories: details of the user's customised worldview or research philosophies, a presentation of how the user's inputs were tokenised, lemmatised, and the name recognition entity process.

In the first instance, the output data reveals the type of variations in the philosophical assumptions that the user espouses within a cluster or clusters. Obviously, in the first instance, the observation is linked to the cluster output(s) of the NLP. For example, the pie chart in Figure 6 shows that the user's philosophical stance is skewed more to the realism cluster (i.e. cluster with different philosophies with assumptions closer to realism's assumptions).

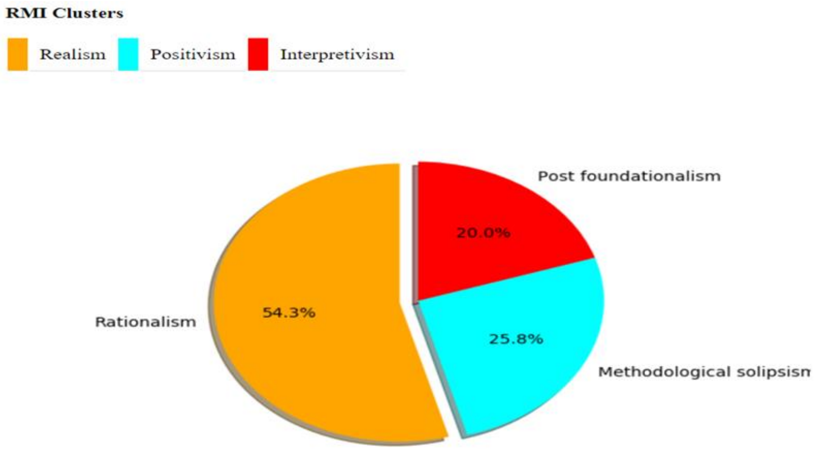
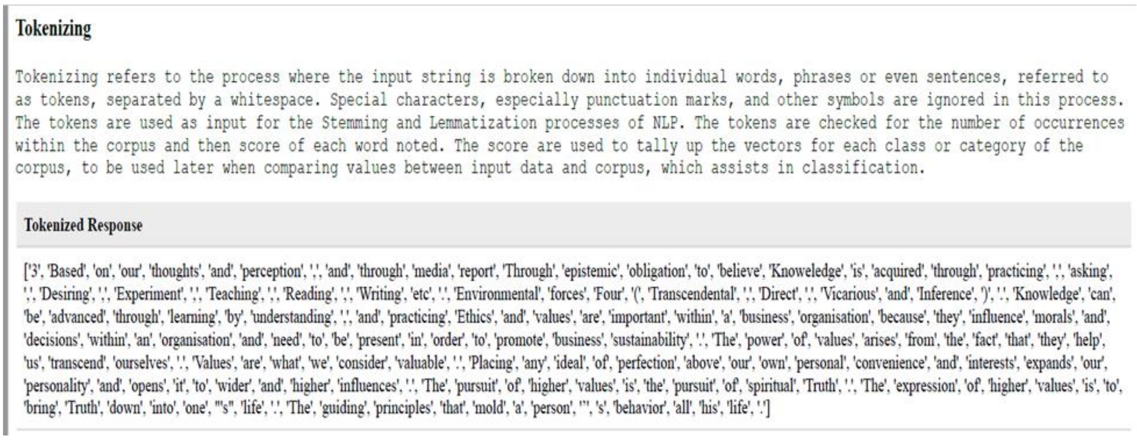


Figure 6: NLP Output Results

Most importantly, the pie chart reveals the specific philosophical assumption within the realist cluster with clear percentages of the most aligned philosophy. Notably, the user's expression reveals an alignment to philosophies in multiple clusters (i.e. realism, interpretivism, and positivists clusters), which indicates that a user possesses multiple philosophical assumptions.

Beyond the revelation of the cluster and actual philosophy(ies), the NLP results also present a transparent process of how the user's unstructured data was tokenised and lemmetised, across different name entities. For example, Figures 7, 8 and 9 below show how the user's unstructured data was broken down into different tokens, lemmetised, and the name entity recognition (NER) patterns. Of particular importance about NER, is that it reveals details of the actual philosophies with hyperlinks that crawl to the database of research philosophies and paradigm literature that provides the user with more information about each correlated philosophy.



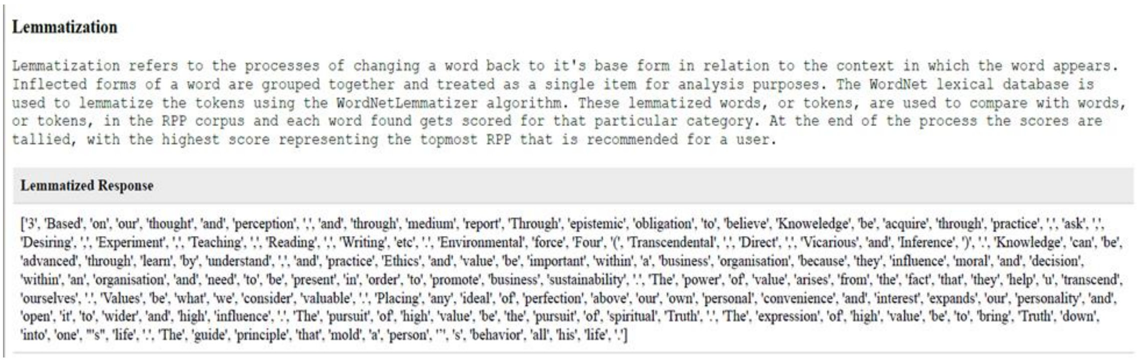
Tokenizing

Tokenizing refers to the process where the input string is broken down into individual words, phrases or even sentences, referred to as tokens, separated by a whitespace. Special characters, especially punctuation marks, and other symbols are ignored in this process. The tokens are used as input for the Stemming and Lemmatization processes of NLP. The tokens are checked for the number of occurrences within the corpus and then score of each word noted. The score are used to tally up the vectors for each class or category of the corpus, to be used later when comparing values between input data and corpus, which assists in classification.

Tokenized Response

['3', 'Based', 'on', 'our', 'thoughts', 'and', 'perception', ',', 'and', 'through', 'media', 'report', 'Through', 'epistemic', 'obligation', 'to', 'believe', 'Knoweledge', 'is', 'acquired', 'through', 'practicing', ',', 'asking', ',', 'Desiring', ',', 'Experiment', ',', 'Teaching', ',', 'Reading', ',', 'Writing', 'etc', '.', 'Environmental', 'forces', 'Four', '(', 'Transcendental', ',', 'Direct', ',', 'Vicarious', 'and', 'Inference', ')', '.', 'Knowledge', 'can', 'be', 'advanced', 'through', 'learning', 'by', 'understanding', ',', 'and', 'practicing', 'Ethics', 'and', 'values', 'are', 'important', 'within', 'a', 'business', 'organisation', 'because', 'they', 'influence', 'morals', 'and', 'decisions', 'within', 'an', 'organisation', 'and', 'need', 'to', 'be', 'present', 'in', 'order', 'to', 'promote', 'business', 'sustainability', '.', 'The', 'power', 'of', 'values', 'arises', 'from', 'the', 'fact', 'that', 'they', 'help', 'us', 'transcend', 'ourselves', '.', 'Values', 'are', 'what', 'we', 'consider', 'valuable', '.', 'Placing', 'any', 'ideal', 'of', 'perfection', 'above', 'our', 'own', 'personal', 'convenience', 'and', 'interests', 'expands', 'our', 'personality', 'and', 'opens', 'it', 'to', 'wider', 'and', 'higher', 'influences', '.', 'The', 'pursuit', 'of', 'higher', 'values', 'is', 'the', 'pursuit', 'of', 'spiritual', 'Truth', '.', 'The', 'expression', 'of', 'higher', 'values', 'is', 'to', 'bring', 'Truth', 'down', 'into', 'one', "'s", 'life', '.', 'The', 'guiding', 'principles', 'that', 'mold', 'a', 'person', "'", 's', 'behavior', 'all', 'his', 'life', '.']

Figure 7: Tokenisation of the User's Unstructured Data



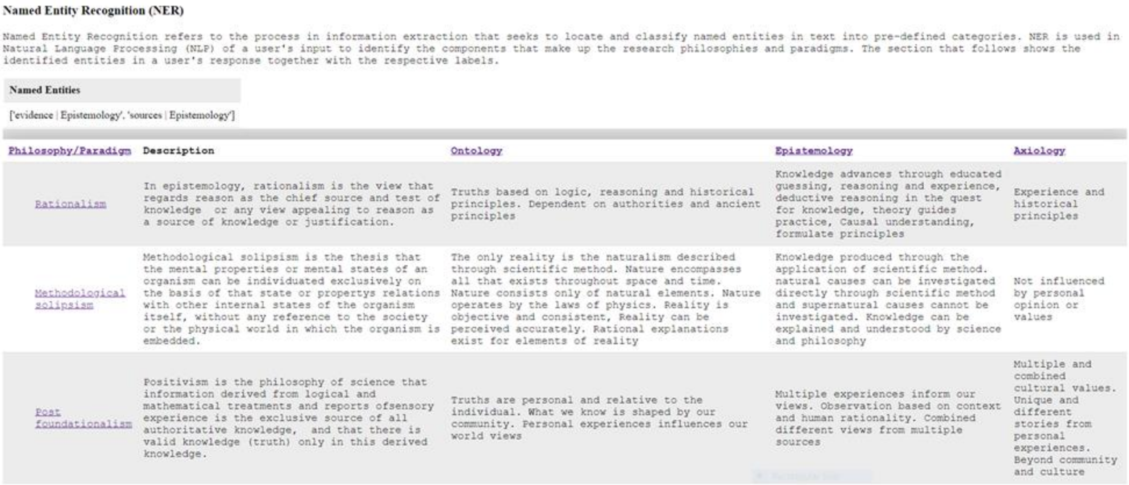
Lemmatization

Lemmatization refers to the processes of changing a word back to it's base form in relation to the context in which the word appears. Inflected forms of a word are grouped together and treated as a single item for analysis purposes. The WordNet lexical database is used to lemmatize the tokens using the WordNetLemmatizer algorithm. These lemmatized words, or tokens, are used to compare with words, or tokens, in the RPP corpus and each word found gets scored for that particular category. At the end of the process the scores are tallied, with the highest score representing the topmost RPP that is recommended for a user.

Lemmatized Response

['3', 'Based', 'on', 'our', 'thought', 'and', 'perception', ',', 'and', 'through', 'medium', 'report', 'Through', 'epistemic', 'obligation', 'to', 'believe', 'Knoweledge', 'be', 'acquire', 'through', 'practice', ',', 'ask', ',', 'Desiring', ',', 'Experiment', ',', 'Teaching', ',', 'Reading', ',', 'Writing', 'etc', '.', 'Environmental', 'force', 'Four', '(', 'Transcendental', ',', 'Direct', ',', 'Vicarious', 'and', 'Inference', ')', '.', 'Knowledge', 'can', 'be', 'advanced', 'through', 'learn', 'by', 'understand', ',', 'and', 'practice', 'Ethics', 'and', 'value', 'be', 'important', 'within', 'a', 'business', 'organisation', 'because', 'they', 'influence', 'moral', 'and', 'decision', 'within', 'an', 'organisation', 'and', 'need', 'to', 'be', 'present', 'in', 'order', 'to', 'promote', 'business', 'sustainability', '.', 'The', 'power', 'of', 'value', 'arises', 'from', 'the', 'fact', 'that', 'they', 'help', 'u', 'transcend', 'ourselves', '.', 'Values', 'be', 'what', 'we', 'consider', 'valuable', '.', 'Placing', 'any', 'ideal', 'of', 'perfection', 'above', 'our', 'own', 'personal', 'convenience', 'and', 'interest', 'expands', 'our', 'personality', 'and', 'open', 'it', 'to', 'wider', 'and', 'high', 'influence', '.', 'The', 'pursuit', 'of', 'high', 'value', 'be', 'the', 'pursuit', 'of', 'spiritual', 'Truth', '.', 'The', 'expression', 'of', 'high', 'value', 'be', 'to', 'bring', 'Truth', 'down', 'into', 'one', "'s", 'life', '.', 'The', 'guide', 'principle', 'that', 'mold', 'a', 'person', "'", 's', 'behavior', 'all', 'his', 'life', '.']

Figure 8: User's Unstructured Data Using



Named Entity Recognition (NER)

Named Entity Recognition refers to the process in information extraction that seeks to locate and classify named entities in text into pre-defined categories. NER is used in Natural Language Processing (NLP) of a user's input to identify the components that make up the research philosophies and paradigms. The section that follows shows the identified entities in a user's response together with the respective labels.

Named Entities

['evidence | Epistemology', 'sources | Epistemology']

Philosophy/Paradigm	Description	Ontology	Epistemology	Axiology
Rationalism	In epistemology, rationalism is the view that regards reason as the chief source and test of knowledge or any view appealing to reason as a source of knowledge or justification.	Truths based on logic, reasoning and historical principles. Dependent on authorities and ancient principles	Knowledge advances through educated guessing, reasoning and experience, deductive reasoning in the quest for knowledge, theory guides practice, Causal understanding, formulate principles	Experience and historical principles
Methodological solipsism	Methodological solipsism is the thesis that the mental properties or mental states of an organism can be individuated exclusively on the basis of that state or propertys relations with other internal states of the organism itself, without any reference to the society or the physical world in which the organism is embedded.	The only reality is the naturalism described through scientific method. Nature encompasses all that exists throughout space and time. Nature consists only of natural elements. Nature operates by the laws of physics. Reality is objective and consistent, Reality can be perceived accurately. Rational explanations exist for elements of reality	Knowledge produced through the application of scientific method. natural causes can be investigated directly through scientific method and supernatural causes cannot be investigated. Knowledge can be explained and understood by science and philosophy	Not influenced by personal opinion or values
Post foundationalism	Positivism is the philosophy of science that information derived from logical and mathematical treatments and reports ofsensory experience is the exclusive source of all authoritative knowledge, and that there is valid knowledge (truth) only in this derived knowledge.	Truths are personal and relative to the individual. What we know is shaped by our community. Personal experiences influences our world views	Multiple experiences inform our views. Observation based on context and human rationality. Combined different views from multiple sources	Multiple and combined cultural values. Unique and different stories from personal experiences. Beyond community and culture

Figure 9: NER of User's Philosophical Assumption

4. Conclusion and Recommendations

This study offers a progressive technological approach towards understanding research paradigms and philosophies. In particular, the study uses NLP and machine learning to introduce users to a wealth of research paradigms and philosophies that are aligned to their worldview, rather than those that are most familiar to their teacher or supervisor. The natural language processing and machine learning inspired digital model poses questions that place students in a reflexive mode and draws their articulated responses as inputs that model their worldviews against a host of philosophies in the database. The inquiry-based teaching approach transforms learning from the generic push-teaching method that assumes universality, to the fostering of a reflexive approach that helps resolve the deep ideological approaches that caused the polarisation. The polarisation and dichotomies that Silverman (2010) and Acheampong et al. (2015) identified in research methods may be largely eradicated, leading to a new discovery of deep learning, and an awareness and understanding of the concepts that can boost the development of research students and knowledge production.

Beyond providing insight into how technology can complement the teaching and learning of a complex subject that is characterised by high variety, the findings of this study contribute to the evoking of deep learning arising from new philosophies and methods. The application of NLP to the subject of research philosophies extends its relevance beyond the subjects of marketing, aviation, and customer relationship management that were quite evident in literature.

On consideration of the rate of automation in the fourth industrial revolution, and how technology permeates every industry, it is worthwhile applying the novelty of NLP to investigate research philosophies for teaching and learning research philosophies. At best, the NLP's insight into knowledge production can add to the discussion of how knowledge production can potentially be transformed. The latter elevates scientific knowledge beyond what might normally be expected of library of research methods, and provides an alternative direct response to Davison and Martisons' (2016) call for new ways of developing theories. It also reveals how NLP can help to facilitate a responsive interactive teaching and learning process that has the potential to improve understanding of the concepts.

Furthermore, this study revealed developments that are discordant with previous scholars who advocated for a single philosophical assumption for a field, subject, or researcher, such as the existence of pure positivist by Mach, and/ or pure interpretivist (Banks, 2013). Whether the discovery of users' multiple philosophical assumptions will help to curb the parallels, debates, and dilemmas documented by previous scholars, remains an interesting question (Mkansi and Acheampong, 2012; Scotland, 2012; Sefotho, 2015). Future research can engage scholars with extreme right or left philosophical assumptions to use the research method index (RMI) system to assess whether they are who they say they are. And if not, what can change moving forward?

The digital discovery of research philosophies and paradigms extend the work of previous theorists to the technologically-inspired discovery of episteme, ontology, and axiology. By its nature, the use of NLP becomes an advanced channel to how we know what we know, and what is the nature of the reality and values being displayed? Further studies could formulate several hypotheses and theorems that arise from technological foundations. It further reinforces the assertions of Habermas (1972) and James and Vinnicombe (2002) that commonly emphasise the unneutrality of knowledge, and how every researcher has deeply embedded worldviews that differently shape knowledge production and research methods.

Moreover, the findings present an attempt towards reducing the gulf between theory and practice, especially the lack of knowledge and awareness of the concepts. Future studies can explore whether the technological discovery inspires the application and acknowledgement of the concepts in major knowledge outlets, such as journals with high impact factors and by students in different fields.

Lastly, the study provides an alternative method of analysing and cross-validating data through the demonstration of NLP to shine some light on Aram and Salipanter's (2003) call for new communities of relevance and rigorous knowledge production. The NLP approach reveals how scholars can apply various methods to data while employing different research paradigms that yield lean insights for mass customisation in teaching and learning.

The study is not without limitation, as its first demonstration of NLP does not measure the impact of the software on improving users' understanding of the concepts and how it can shape research methods. Hence, the recommendation for interested parties (academics and post-graduate students) to explore the second and third phase of the RMI project which deals with determining how the system can improve understanding of the concepts, how research philosophies shape research methods, and help generate knowledge. Future studies can use the first phase of the software project (www.rmi.unisa.ac.za) to conduct social experiments and report on the effects and impact of the software on their teaching and learning of research philosophies and paradigms at third-year (undergraduate) or post-graduate levels.

Reference

Acheampong, E, Mkansi, M, Kondadi, K, and Qi, B (2015) Polarization in research methods. In J. Mendy, & S. D. Geringer (Eds.), *Leading Issue in Business Research Methods Volume 2*. Academic Conferences and Publishing International Limited.

Acheampong, E, Mkansi, M, Kondadi, KR, and Qi, B (2012) Polarization in research methods application: Examining the examiner. *11th European conference on Research Methodology for Business and Management Studies*. Bolton, UK.

Aggarwal, CC, and Zhai, C (2012). *Mining Text Data*. Springer.

Alvesson, M, and Skoldberg, K (2009) *Reflexive methodology: new vistas for qualitative research*. Sage.

Aram, JD, and Salipanter, PF Jr (2003) Bridging scholarship in management: Epistemological reflections. *British Journal of Management, 14*(3), 189-205.

Arbuckle, JL (2008) *AmosTM 17.0 User's Guide*. Available at http://www.jou.ufl.edu/research/lab/pdf/Amos-17.0-User's-Guide.pdf (accessed 20 June 2011)

Banks, EC (2013) Metaphysics for positivists: Mach versus the Vienna Circle. *Discipline Filosophiche, 23*, 57-77.

Bertoni, M, Rondini, A, and Pezzotta, G (2017) A systematic review of value metrics for PSS design. *Procedia CIRP*, *64,* 289-294.

Blaikie, N (1993) *Approaches to social enquiry*. Polity Press.

Blaikie, N (2000) *Designing social research*. Polity Press.

Bonaccorso, G (2017) *Machine learning algorithms: Reference guide for popular algorithms for data science and machine learning.* Packt.

Brownlee, J (2018) *Supervised and Unsupervised Machine Learning Algorithms*. Available at https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/ (accessed 30 January 2020)

Brunton, SL, Noack, B R, and Koumoutsakos, P (2019) Machine learning for fluid mechanics. *Annual Reviews of Fluid Mechanics*, *52,* 477-508.

Carpenter, S (2019) Contemporary digital systems. In M. Mkansi, N. McLennan, and G. de Villiers (Eds.), *Contemporary Issues in Operations and Supply Chain Management* (pp. 146-175). Van Schaik.

Chang, T, DeJonckheere, M, Vydiswaran, VGV, Li, J, Buis, LR, Guetterman, TC (2021) Accelerating Mixed Methods Research with Natural Language Processing of Big Text Data. *Journal of Mixed Methods Research*, 15(3), 398-412.

Chen, Y, Wu, X, Hu, A, He, G, Ju, G (2021) Social Prediction: a new research paradigm based on machine learning. *The Journal of Chinese Sociology*, 8(1), 1-21.

Creswell, JW (2013) *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Sage.

Creswell, JW (2014) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (4th ed.). Sage.

Chang, AX, Spitkovsky, VI, Manning, CD, and Agirre, E (2016) *A comparison of named-entity disambiguation and word sense disambiguation.* Available at https://nlp.stanford.edu/pubs/chang2016entity.pdf (accessed 30 January 2020)

Daniel, B K (2018) Empirical verification of the "TACT" framework for teaching rigour in qualitative research methodology. *Qualitative Research*. Advance online publication.https://doi.org/10.1108/QRJ-D-17-00012.

Davison, R M, and Martinsons, M. G (2016) Context is king! Considering particularism in research design and reporting. *Journal of Information Technology*, *31*(3), 241-249.

Denyer, D, and Tranfield, D (2009) Producing a systematic review. In D. Buchanan, & A. Bryman (Eds.), *The Sage Handbook of Organizational Research Methods* (pp. 671–689). Sage Publications.

Denzin, N, and Lincoln, Y (2003) The discipline and practice of qualitative research. In N. Denzin & Y. Lincoln (Eds.), *Collecting and interpreting qualitative materials* (2nd ed.). Sage.

Derczynski, L, Maynard, D, Rizzo, G, van Erp, M, Gorrell, G, Troncy, R, Petrak. J. & Bontcheva, K (2015). Analysis of named entity recognition and linking for tweets. *Information Processing and Management, 51*(2), 32-49.

Faherty, VE (2010) *Wordcraft: Applied Qualitative Data Analysis (QDA) Tools for Public and Voluntary Services*. Sage Publications.

Field, AP (2009) *Discovering Statistics Using SPSS*. Sage Publications.

Groeesler, A (2017) Teaching Research Methods: An Occasional Paper. *Institute of Teaching and Learning Innovation (ITaLi)*. The University of Queensland.

Gaizauskas, R, and Wilks, Y (1998) Information Extraction: beyond document retrieval, *Journal of Documentation*, 54(1), 70-105.

Gandrabur, S, Foster, G, and Lapalme, G (2006) Confidence estimation for NLP applications. *ACM Transactions on Speech and Language Processing, 3*(3), 1-29.

Guarino, N (1998) Formal Ontology in Information Systems. Paper published in *Proceedings of FOIS'98*, Trento, Italy, 6–8 June 1998 (pp. 3-15). Amsterdam: IOS Press.

Guba, EG, and Lincoln, YS (1994) Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of Qualitative Research* (pp. 105–117). Sage.

Habermas, J (1972) *Knowledge and human interest*. Beacon Press.

Hesse-Biber, S (2015) The problems and prospects in the teaching of mixed methods research. *International Journal of Social Research Methodology*, *15*(5), 463-477.

Hsieh, H F, and Shannon, SE (2005) Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277-1288.

James, K, and Vinnicombe, S (2002) Acknowledging the individual in the researcher. In D. Partington (Ed.), *Essential skills for management research* (pp. 84-98). Sage.

Javed, S, Adewumi, TP, Liwicki, FS, Liwicki, M (2021) Understanding the Role of Objectivity in Machine Learning and Research Evaluation. *Philosophies*, 6(22), 1-8.

Jensen, HS (2000) A history of the concept of knowledge. *Zagreb International Review of Economics and Business*, *3*(2), 1-16.

Jiang, J (2012) Mining text data: Information extraction from text. In C. C. Aggarwal & C. Zhai (Eds.), *Mining Text Data* (pp. 11-36). Springer.

Kaminski, J, Jiang, Y, Piller, F, and Hopp, C (2017) Do user entrepreneurs speak different? Applying natural language processing to crowdfunding videos. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2683-89). Denver.

Kanchan, N, and Yadav, PR (2022) Realisation of natural language processing and machine learning approaches for text-based sentiment analysis. *Expert Systesms*, DOI: 10.1111/exsy.13114.

Kelleher, J D, Namee, B M, and D'Arcy, A (2015) *Fundamentals of Machine Learning for Predictive Data Analytics.* Massachusetts Institute of Technology.

Keogh, E (2015) *Naïve Bayes classifier.* Available at: http://www.cs.ucr.edu/~eamonn/CE/Bayesian%20Classification%20withInsect_examples.pdf (accessed 01 October 2020)

Killam, L (2013) *Research terminology simplified: Paradigms, axiology, ontology, epistemology and methodology.* World Bank Publishers.

Kumar, S, and Zymbler, MA (2019) A machine learning approach to analyse customer satisfaction from airline tweets. *Journal of Big Data,* 6(62), 1-16.

Leeson, W, Resnick, A, Alexander, D, and Rovers, J (2019) Natural Language Processing (NLP) in qualitative public health research: A proof of concept study. *International Journal of Qualitative Methods,* 18, 1-9.

Li, Q, and Sha, S (2017) Learning stock market sentiment lexicon and sentiment-oriented word vector from StockTwits. *Proceedings of the 21st Conference on Computation Natural Language Learning* (pp. 301-310). Vancouver, Canada. Association for Computational Linguistics.

Lincoln, Y (1990) The making of a constructivist: A remembrance of transformations past. In E. Guba (Ed.), *The Paradigm dialog* (pp. 67-87). Sage Publications.

Loh, WY (2011) Classification and Regression Trees. *Wiley Interdisciplinary Review, Data Mining and Knowledge* Disc, 1, (pp. 14-23).

Mackenzie, N, and Knipe, S (2006) Research dilemmas: Paradigms, methods and methodology. *Issues in Educational Research*, *16*(2), 193-205.

Makombe, G (2017) An expose of the relationship between paradigm, method and design in research. *The Qualitative Report*, *22*(12), 3363-3382.

Marr, B (2016) *What is the difference between artificial intelligence and machine learning?* Available at https://www.google.co.za/amp/s/www.forbes.com/sites/bernardmarr/2016/12/06/what-is-the-difference-between-artificial-intelligence-and-machine-learning/amp/ (accessed 30 January 2020)

Miller, PK, and Grimwood, T (2015) Mountains, cones, and dilemmas of context: The case of "ordinary language" in philosophy and scientific method. *Philosophy of the Social Sciences*, *45*(3), 331-355.

Mkansi, M (2020) Research methods index. In D. Remenyi (Ed.), *Innovation in Teaching of Research Methodology Excellence Awards: An Anthology of Case Histories 2020*. Academic Conferences and Publishing International Limited.

Mkansi, M, and Acheampong, EA (2012) Research philosophical debates and classifications: Students' dilemma. *Electronic Journal of Business Research Methods*, *10*(2), 132-140.

Mkansi, M (2018) Research paradigm and philosophies swept under the carpet: A summative content analysis. *17th European Conference on Research Methods in Business and Management*, 12-13 July, Universita Roma Tre, Rome, Italy. Academic Conferences and publishing limited.

Muhaise, H, EJiri, AH, Muwanga-Zake, JWF, and Kareyo, M (2020) The Research Philosophy Dilemma for Postgraduate Student Researchers. *International Journal of Research and Scientific Innovation* Vol 7, No 4, pp. 2321–2705.

Pedrycz, W, and Chen, S (2016) *Sentiment Analysis and Ontology Engineering: An Environment of Computational Intelligence*. Springer International Publishing.

Peffers, K, Tuunaneh, T, and Niehaves, B (2018) Design science research genres: introduction to the special issue on exemplars and criteria for applicable design science research. *European Journal of Information Systems*, *27*(2), 129-139.

Polit, D, and Beck, C T (2008) *Nursing research: Generating and assessing evidence for nursing practice*. Lippincott Williams and Wilkins.

Reppen, R (2010) Building a corpus: What are the key considerations? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics.* Routledge Handbooks.

Saillard, KE (2011) Systems Versus Interpretive with Two CAQDAS Packages: Nvivo and MAXQDA. *Forum Qualitative Social Research*, *12*(1), Art. 34.

Saunders, M, Lewis, P, and Thornhill, A (2016) *Research methods for business students* (5th ed.). Pearson Education.

Saunders, M, Lewis, P, and Thornhill, A (2019) *Research methods for business students.* Pearson Education.

Sebastiani, F (2002) Machine learning in automated text categorization. *ACM Computing Surveys*, 34, 1–47.

Scotland, J (2012) Exploring the philosophical underpinnings of research: Relating ontology and epistemology to the methodology and methods of the scientific, interpretive, and critical research paradigms. *English Language Teaching*, *5*(9), 9–16.

Seddon, P, and Scheepers, R (2012) Towards the improved treatment of generalisation of knowledge claims in IS research: Drawing general conclusions from samples. *European Journal of Information Systems*, *21*(1), 6-21.

Sefotho, MM (2015) A researcher’s dilemma: Philosophy in crafting dissertations and theses. *Journal of Social Science*, *42*(12), 23–36.

Shouval, R, Bondi, O, Mishan, H, Shimoni, A, Unger, R, and Nagler, A (2014) Application of machine learning algorithms for clinical predictive modeling: A data-mining approach in SCT. *Bone Marrow Transplantation*, *49*(3), 332-337.

Silverman, D (2010) *Doing qualitative research* (3rd ed.). Sage.

Steen, M, and Roberts, T (2011) *Handbook of midwifery research*. Wiley-Blackwell.

Te’eni, D, Rowe, F, Agerfalk, PJ, and Lee, JS (2015) Publishing and getting published in *EJIS*: marshalling contributions for a diversity of genres. *European Journal of Information Systems*, *24*(6), 559-568.

Waters, A, Grimaldi, P, Lan, A, Baraniuk, R (2017) Short-Answers Responses to STEM exercises: Measuring Response Validity and its Impact on Learning. *Proceedings of the 10th International Conference on Data Mining*, 374-375.

Yin, RK (2018) *Case study research and applications: Design and methods* (6th ed.). Sage.

Zeroual, I, and Lakhouaja, A (2018) Data science in light of Natural Language Processing: An overview. *Procedia Computer Science*, *127*(2018), 82-91.

Zhao, H, and Kit, C (2011) Integrating unsupervised and supervised word segmentation: The role of goodness measures. *Information Sciences*, *181*(1), 163-183.

Zhao, L, Alhoshan, W, Ferrari, A, Letsholo, KJ, Ajagbe, MA, Chioasca, E, Batista-Navarro, RT (2021) Natural Language Processing for Requirements Engineering: A systematic Mapping Study. ACM Comput. Surv, 54(3), 1-41.

Appendix 1: Corroboration of Research Philosophies Data

Corroboration of research philosophies data				**Data Sources**			
Paradigm/ philosophy	**Year**	**References**	**Title of data sources**	**Journals**	**Encyclopaedia**	**Books**	**Dissertations**
Infinitism	2018	Aikin, Scott.	Epistemic infinitism, 2018, doi:10.4324/0123456789-P077-1. Routledge *Encyclopedia of Philosophy*.		●		
	2013	Turri, J., & Klein, P.	Infinitism in epistemology	●			
Innatism	2010	Hill, J	The Synthesis of Empiricism and Innatism in Berkeley’s Doctrine of Notions. *Berkeley Studies*. 21. 3–15.	●			
	2015	Winch C	Innatism, Concept Formation, Concept Mastery and Formal Education. *J Philos Educ* [Internet]	●			
Internalism	2017	Pappas, George	"Internalist vs. Externalist Conceptions of Epistemic Justification", *The Stanford Encyclopedia of Philosophy*		●		
	2017	Wilson, R.A.	Externalism and Internalism in the Philosophy of Mind			●	
Interpretivism	2015	Aliyu et al.	Positivist and Non-Positivist Paradigm in Social Science Research: Conflicting Paradigms or Perfect Partners?	●			

Corroboration of research philosophies data				**Data Sources**			
Paradigm/ philosophy	**Year**	**References**	**Title of data sources**	**Journals**	**Encyclopaedia**	**Books**	**Dissertations**
	2014	Vosloo, J.J	A Sport management programme for educator training in accordance with the diverse needs of South African Schools				●
	1967	Meyers	Peirce on Cartesian Doubt. Transactions of the Charles S. Peirce Society				●
	2012	Wahyuni	The Research Design Maze: Understanding Paradigms, Cases, Methods and Methodologies				●
	2012	Saunders et al.	*Research Methods for Business Students*			●	
Irrealism (philosophy)	2016	Cohnitz, Daniel and Rossberg, Marcus	"Nelson Goodman", *The Stanford Encyclopedia of Philosophy*		●		
	2014	Cohnitz	Nelson Goodman		●		
	2013	Bhaskar	A_Realist_Theory_of_Science			●	
Liberal naturalism	2018	Macarthur	Liberal naturalism and the scientific image of the world, Inquiry	●			
	2017	Crespo	Economics and Other Disciplines: Assessing New Economic Currents Routledge Advances in Social Economics			●	