Journal of Research in Business : e33-47 Özel Sayı • ISSN: 2630-6255
DOI: 10.29228/JRB.1026523

RESEARCH ARTICLE • FORSCHUNGSARTIKEL • ARAŞTIRMA MAKALESİ

INFLUENCE OF DIFFERENT THEORIES OF ETHICS ON ORGANIZATIONAL CODES OF CONDUCT OR ETHICS: A COMPARATIVE SEMANTIC ANALYSIS

FARKLI ETİK KURAMLARININ KURUMSAL ETİK TUTUM VE DAVRANIŞ KURALLARINA ETKİSİ: BİR KARŞILAŞTIRMALI SEMANTİK ANALİZ

M. Murat ALBAYRAKOĞLU[*]
Mehmet N. AYDIN[**]

Abstract

The aim of this study was to investigate the influence of various theories of ethics on codes of conduct or codes of ethics of computing and data organizations. To quantify and evaluate the differences in influence, four *Python* libraries, namely *difflib*, *gensim*, *nltk*, and *spaCy*, and, in addition, a web-based proprietary semantic similarity tool, *Compare Text*, were used. The codes of seven computing and data organizations for Information Technology (IT) professionals and scholars were compared to the descriptions of five different schools of ethical thought through four different tools. The findings were tabularized, summarized in radar charts, and their implications were discussed: It was found that there are some differences of influence on the codes by different theories. However, the percentages of similarities calculated by each tool were observed to differ, on some occasions, considerably. Finally, contributions and limitations of the current work and recommendations for further studies were presented.

Keywords: Codes of conduct or ethics, computing and data organizations, semantic similarity, theories of ethics.

JEL Classification: M31, O39

* **Corresponding Author:** PhD. Student, Kadir Has University, School of Graduate Studies, MIS Program, murat.albayrakoglu@khas.edu.tr, ORCID: 0000-0002-5057-5641.

** Assoc. Prof., Kadir Has University, Faculty of Economics, Administrative and Social Sciences, Department of MIS, mehmet.aydin@khas.edu.tr, ORCID: 0000-0002-3995-6566.

To cite this article: Albayrakoğlu, M. M. & Aydın, M. N. (2022). Influence of different theories of ethics on organizational codes of conduct or ethics: A comparative semantic analysis. *Journal of Research in Business*, 7(1), e33-47.

"There is no requirement of Ethics Committee Approval for this study."

Submitted: 20.11.2021 **Revised:** 07.02.2022
Accepted: 01.03.2022 **Online Yayın:** 30.03.3022

Öz

Bu çalışmanın amacı çeşitli etik kuramlarının, bilgisayar ve veri derneklerine ait etik tutum ve davranış kuralları üzerindeki etkilerini incelemektir. Etkilerin arasındaki farkları ölçmek ve değerlendirmek için, dört *Python* kütüphanesi, *difflib*, *gensim*, *nltk* ve *spaCy*, ve bunlara ek olarak web-tabanlı ve lisanslı bir semantik benzerlik aracı olan *Compare Text* kullanılmıştır. Bilişim Teknolojisi (BT) profesyonelleri ve akademisyenlerine ait yedi bilişim ve veri derneğinin kuralları, beş farklı etik düşünce ekolünün betimlemeleri ile karşılaştırılmıştır. Bulgular tablolanmış, radar çizimleri biçiminde özetlenmiş ve sonuçları tartışılmıştır. Farklı kuramların, kurallar üzerinde bir miktar farklı etkilerinin olduğu saptanmıştır. Bunun yanısıra, bazı araçlar ile hesaplanan benzerlik yüzdelerinin diğerlerininkinden ciddi ölçüde farklı olduğu gözlemlenmiştir. Son olarak, mevcut çalışmanın katkıları ve sınırları ile gelecekteki çalışmalar için öneriler sunulmuştur.

Anahtar Kelimeler: Bilişim ve veri dernekleri, etik tutum ve davranış kuralları, semantik analiz, etik kuramları.

JEL Sınıflandırılması: M31, O39

1. Introduction

When there is a large corpus of documents to analyze, either because the number of documents is high or the average length of documents is long, it becomes a tedious task to compare them manually among themselves or with one or more other documents. For some corpora, the task is even impossible for human beings. This is where Natural Language Processing (NLP) tools become handy.

Furthermore, one can use semantic analysis to discover subtle differences, however slight, among the documents of a corpus since NLP tools have the ability to quantify similarity. This study involves using NLP over a small corpus to discover the similarity of each document with each one in another corpus. Documents in the second corpus were thought to have some influence on those in the first.

The aim is to discover which theories of ethics, sometimes called schools of ethical thought, have the most significant influence or impact on computer and data organizations' codes of conduct or ethics. Without computers and artificial intelligence (AI), such assessments were made through verbal arguments, but the findings cannot be quantified. In that case, it is hardly possible, if not totally impossible, to distinguish subtleties among influences. In contrast, this study seeks methods and tools to help a researcher quantify the differences rather than a mere qualitative assessment.

The language of the codes used in this study is English. The study covers organizations within the org domain. Seven codes of conduct or ethics of associations or institutes of IT professionals and scholars are included in the study.

Texts about five theories of ethics were used as bases for comparison. Textual analyses were carried out by comparing each of the seven organizational codes with each of the five schools of thought. *Python* programming language has several libraries that can be used to compare sentences, short texts, or entire documents.

Four analyses were carried out using *Python*'s *difflib* (Python, 2021), *gensim* (Řehůřek, 2021a), and *nltk* (NLTK, 2021), and *spaCy* (spaCy, 2021a) libraries, and *Compare Text* by cortical.io (2021). *Compare Text* is a cloud-based, free, but proprietary tool available on the Web.

2. Literature Review

This literature review consists of discussions about ethics, theories of ethics, computer and data ethics along with the codes of conduct or ethics for computer and data professionals, and methods and tools for semantic similarity analytics.

2.1. Ethics

Ethics is a concept that is hard to define. In general, it refers to moral judgment applied to attitude and behavior of individuals toward others over some continua between two extremes: positive and negative, as shown below:

Positive	Negative
Right	Wrong
Good	Bad
Just	Unjust
Virtuous	Vicious
Proper	Improper

As individual attitude or behavior gets closer to the positive, it is considered ethical and, therefore, more acceptable. Each individual is a conscientious moral agent who uses moral reasoning to act ethically (Donlevy & Walker, 2011).

Ethics can be viewed from two points of view. The first involves the theoretical aspects of acceptable attitude and behavior: it is called theoretical ethics, or philosophy of ethics. The second is called applied ethics or practical ethics. Applied ethics deals with the application of theoretical ethics to solve ethical problems of real life. Its scope varies from the individual to various forms of social formations, including for-profit and not-for-profit organizations (Shafer-Landau, 2019).

2.2. Theories of Ethics

There are different ways of classifying and describing schools of ethical thought (Deigh, 2010; Shafer-Landau, 2012; Rachels & Rachels, 2015). A classification offered by Quinn (2015) and his definition and discussion of each school of thought will be used as a basis for comparing different codes of conduct or ethics in historical order.

Virtue ethics Ethical decisions and choices of a person about doing what is right, good, just, virtuous, or proper are based on the character (who the person is) and deeds (what the person does) of that person. The four virtues of an ethical person are wisdom, fairness, courage, and self-control.

Kantian ethics Kantian or deontological ethics is based on a set of rules to do good. It is founded on the following three principles:

1. Do what you would want to be done to you, by others, to others.
2. Always apply the same rules to everybody, including yourself.
3. A person is never a means but an end for herself.

Act-utilitarian ethics If what a person chooses to do or not to do causes more good than harm, she should do it. Otherwise, she should refrain from doing it respectively. The utility is the difference between the positive and negative possible consequences of an act. Therefore, a person must act in the best interest of the greatest number of people to end up with the greatest net utility. In act-utilitarian ethics, negative consequences upon others, or the minorities, can be, perhaps sadly, acknowledged and yet tolerated, or even outright ignored.

Rule-utilitarian ethics While based on the concept of utility, rule-utilitarian ethics considers a specific action to be morally justified only if the act conforms with a moral rule. A moral rule should still create more utility as compared to other possible rules. However, it should also guarantee that negative consequences of or potential harm due to a specific action have been eliminated as much as possible.

Social contract A social contract is some kind of an explicit or implicit collective understanding that it is in everybody's interest to have rules that provide safety and security for everyone. These rules are essential for the weakest, the most vulnerable part of society so that even fragile people can survive. For the most part, the social contract is unwritten. It is inherited at birth, and it requires individuals not to break moral codes or laws.

There are other schools of thought, such as theological ethics, relativistic ethics, etc. Fundamentally, the schools that should be used in devising codes of conduct or ethics for big data should be principled and universal. Any school of ethical thought that depends on a particular ideology effective during a specific period, in a certain society, or within a specific geographical area should be avoided.

Among the schools that must be avoided are theological and relativistic ethics. The former is driven by dogma and, therefore, is based on worldview or ideology. Furthermore, depending on religion or sect followed by each society, moral rules may and, actually, do vary. The latter tolerates variations among the moral approaches depending on the culture. According to relativists, culture defines a framework for moral rules and how they are used for moral justification. Since both theological and relativistic ethics allow for some sort of discrimination in applying moral principles to solve moral problems, they were not included in this study.

2.3. Computer and Data Ethics

According to O'Leary (2016), there is a distinction between computer ethics and data ethics: In his article titled "Ethics for Big Data and Analytics," the author purported that big-data ethics should be somewhat different than computer ethics. While the latter was about the uses of information technology (IT) and its consequences, the former involved the uses of what is processed by IT and its consequences. He criticized the limitations of professional organizations' codes of ethics or conduct by regarding those as static and conservative. Instead, he wrote, the codes should constantly evolve and prescribe not only what not to do but also what to do when new technologies emerge. Table 1 shows the codes of conduct or ethics of the computer and data organizations. Six of the links have already been provided by O'Leary. A Web search yielded the seventh: Oxford-Munich Code of Conduct for Professional Data Scientists (2020).

Table 1: Codes of conduct or ethics (O'Leary, 2016; Oxford-Munich Code of Conduct, 2020)

Organization and Abbreviation	Code
Association for Computing Machinery (ACM)	ACM Code of Ethics and Professional Conduct (www.acm.org/about-acm/acm-code-of-ethics-and-professional-conduct)
American Statistical Association (ASA)	Ethical Guidelines for Statistical Practice (www.amstat.org/asa/files/pdfs/EthicalGuidelines.pdf)
British Computer Society (BSC)	BCS Code of Conduct (www.bcs.org/membership/become-a-member/bcs-code-of-conduct/)
INFORMS – Certified Analytics Professional Program (INFORMS)	Code of Ethics/Conduct (https://www.certifiedanalytics.org/ethics.php)
Data Science Association (DSA)	Data Science Code of Professional Conduct (www.datascienceassn.org/code-of-conduct.html)
Institute of Electrical and Electronics Engineers (IEEE)	Code of Ethics (www.ieee.org/about/corporate/governance/p7-8.html)
Oxford-Munich Code of Conduct for Professional Data Scientists (O-M)	Code of Conduct (www.code-of-ethics.org/code-of-conduct/)

2.4. Approaches to Text Analysis

Textual analysis is the generic name given to a collection of methods to make sense of text documents so that their content can be interpreted correctly (McKee, 2003). It combines formal statistical methods and less formal interpretive techniques to identify the patterns of word usage (Ignatow & Mihalcea, 2018). Text mining is used to analyze texts from two different yet complementary perspectives: syntactic and semantic. While syntactic analysis covers the conformity of a given text to the vocabulary and grammatical structures of a language, semantic analysis involves the meaning from different points of view, such as mood, attitude, emotion, etc.

Gomaa and Fahmy (2013) and Vijaymeena and Kavitha (2016) presented two surveys of text similarity approaches in data mining. The former authors classified the similarity approaches into three groups:

- String-based,
- Corpus-based, and
- Knowledge-based.

The latter authors mentioned four similarity measures within the framework of a generic text mining architecture:

- Character-based,
- Term-based,
- Corpus-based, and
- Knowledge-based

Strings can take the form of a character, a word, or a term. The latter can either be a single word or a compound expression consisting of a set of ordered words. In both cases, terms usually belong to a specific domain, such as biology, chemistry, psychology, etc. Corpora cover a range from sentences to short texts (Shrestha, 2011).

3. Methodology

The fundamental assumption of this study is stated as follows: The degree of semantic similarity between the description of an ethical theory and a particular code of conduct or ethics is a measure of the influence of the former on the latter. To evaluate the influence or impact of each theory of ethics on the codes of conduct or ethics of computer and data organizations, five theories of ethics were used as bases for comparison. Since the text of each theory was taken from a textbook (Quinn, 2015) written for pedagogical purposes, it required some purging. Consequently, the examples and references to other theories were removed from the text. Seven codes of conduct or ethics of associations or institutes of IT professionals and scholars were examined without any significant modification.

Textual analyses were carried out by comparing each of the five schools of thought with the seven organizational codes one-by-one. Overall, four separate analyses were performed with *Python*'s *difflib* (Python, 2021), *gensim* (Řehůřek, 2021a), and *nltk* (NLTK, 2021), and *spaCy* (spaCy, 2021a) libraries, and *Compare Text* by cortical.io (2021).

3.1. Semantic Similarity Methods

Semantic similarity is a metric based on an algorithm to determine how close or how distant longer textual elements, such as sentences, paragraphs, or documents, are from each other within a given context. To evaluate the semantic similarity of longer documents, such as letters, news, reports, or stories, sophisticated analyses based on text-mining techniques are needed (Rozeva & Zerkova, 2017).

In text mining, each text first goes through a preprocessing phase that consists of two tasks: linguistic processing and contextual processing. During preprocessing, specific words or terms that convey the meaning of the document are extracted. In the second phase, the document is represented as a set of vectors made up of features (entities) from the first phase. A collection of vectors from different documents amount to a matrix that requires some reduction without a loss of meaning to make it manageable. In the third phase, one of the following methods is used most of the time:

- Distribution analysis,
- Clustering,
- Trend analysis, and
- Association rules,

utilizing the domain knowledge encoded in lexicons, ontologies, taxonomies, and thesauruses. The final phase is a presentation phase that involves some facilities, such as graphics, to summarize the findings.

3.2. Research Design

As far as open-source tools for semantic comparisons are considered, a good number of *Python* packages, of which four were used in the current study, are available. *R* can also be used to carry out semantic similarity analysis. (See, for example, Kumar & Kumar, 2016.) A host of R packages can be used to conduct semantic similarity analyses (CRAN, 2020), among other NLP tools. For this particular work, however, four *Python* libraries and a web-based proprietary semantic similarity tool, *Compare Text*, were selected and used to keep the scope of research manageable.

The first method chosen was the *Gestalt Pattern Matching* (GPM) implemented in *Python*'s built-in *difflib* library. It simply divides the number of matching characters of two strings by the total number of characters to estimate the similarity of two texts (Ratcliff & Metzener, 1988). Therefore, the similarity of the two strings is found as follows:

$$S = \frac{2K_m}{s_1 + s_2} \tag{1}$$

where

S = Similarity of two strings ($0 \le S \le 1$)

K_m = Number of characters of the substrings of a string that match those of the other.

s_1 = Length of the first string ($s_1 > 0$)

s_2 = Length of the second string ($s_2 > 0$)

The algorithm eliminates any irrelevant element such as whitespace or a blank line and then seeks the longest contiguous matching subsequence. Once the sought subsequence is found, similar sequences are investigated before and after the initial subsequence. (Python, 2021). The algorithm goes on to work until all similar subsequences are exhausted. In our work, the **get_close_matches** method of *difflib* was used with its defaults to obtain "good enough" matches.

The second method utilizes AI, based on the Natural Language Toolkit for document tokenization and Generate Similar tool to quantify the similarity of two texts. The corresponding *Python* libraries are *nltk* and *gensim,* respectively. The **sent_tokenize** and **word_tokenize** methods of *nltk* were used in tandem to generate the necessary Bag of Words (BoWs) for each theory-code pair. Afterward, dictionaries for each file were generated, and finally, similarities based on the *Term Frequency-Inverse Document Frequency* (TF-IDF) model that comes with *gensim* (Řehůřek, 2021b) were calculated. The TF-IDF statistically measures how relevant a token is to a document in a collection of documents (Leskovec, Rajaraman & Ullman, 2020). In its simplest implementation, the Term Frequency (TF) is calculated as the ratio of the number of occurrences of a token in a document to the length of the document. For a given set of *N* documents:

$$TF_{ij} = \frac{\varphi_{ij}}{\max_k \varphi_{kj}} \tag{2}$$

where

TF = Term frequency ($0 < TF \leq 1$)

i= Word index ($i = 1, 2, 3\ldots$)

j = Document index ($j = 1, 2, 3\ldots, N$ where N is the number of documents)

φ_{ij} = Frequency of term i in document j

$\max_k \varphi_{kj}$ = The number of the word that occurs most (word k) in document j

The Inverse Document Frequency (IDF) is the ratio that measures the occurrence of a token across a collection of documents: the higher the IDF, the less common the token included in the document set.

$$IDF_i = \log\left(\frac{N}{n_i}\right) \tag{3}$$

where

IDF_i = Inverse document frequency of document i

n_i = Total number of occurrences of word i in N documents

The overall *TF-IDF* score of document i is simply the result of the multiplication of *TF* and *IDF*:

$$TF\text{-}IDF_i = TF_{ij} \times IDF_i \tag{4}$$

The third method chosen was *cosine similarity*. Cosine similarity is implemented as default in the *spaCy* library for *Python* programming language (spaCy, 2021a). *spaCy* was selected because it comes with a pre-trained pipeline (spaCy, 2021b). The existence of the pipeline reduces the requirement for the number of lines of code to type for a program that would have the same functionality.

Cosine similarity is widely used for information retrieval and text mining. Each document is represented by a vector, called a term-frequency vector, consisting of term frequencies. If two non-zero vectors belong to an inner product space, cosine similarity yields a measure of how similar these two vectors, and hence the two documents, are. It is just the cosine of the vectors of length one (Han, Kamber & Pei, 2012). The formula for cosine similarity, therefore, is:

$$S(\boldsymbol{x},\boldsymbol{y}) = \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|} \tag{5}$$

where

$S(\boldsymbol{x}, \boldsymbol{y})$ = Similarity of two term-frequency vectors $\boldsymbol{x}$ and $\boldsymbol{y}$

$\boldsymbol{x}$ = term-frequency vector of document x

$\boldsymbol{y}$ = term-frequency vector of document y

$\|\boldsymbol{x}\|$= Euclidean norm of vector $\boldsymbol{x}$

$\|\boldsymbol{y}\|$= Euclidean norm of vector $\boldsymbol{y}$

Since it is readily available on the Cloud, cortical.io's free *Compare Text* tool was chosen to conduct this study's fourth and last part (Cortical.io, 2021). The tool is based on a theory called Semantic Folding, inspired by how the brain processes information (Webber, 2016). It is a proprietary and patented method implemented on the Cloud using the Software as a Service (SaaS) model for different areas of application.

The system measures the extent of overlap between two pieces of text based on their fingerprints which are numerical representations of meaning. It understands the relatedness of two items by measuring the overlap of their fingerprints. It uses unsupervised learning using a single text or a corpus in a particular language to create a semantic space called Retina Database. A program called Retina Engine converts the corpus into semantic fingerprints. Semantic fingerprints are actually sparsely distributed word vectors. For each word of the corpus, thousands of semantic characteristics are captured.

Semantic Folding is based on the Hierarchical Temporal Memory (HTM), which in turn is based on Sparse Distributed Representations (SDRs) inside the brain, to encode data and input it into HTM networks. Semantically related word vectors are placed close to each other over a two-dimensional semantic map. The Retina Engine converts the symbolic representations into an SDR consisting of 0s and 1s. Thus, a language is decomposed into words, and then, the Retina Engine converts words into SDRs. Two words are similar if they are conceptually related, and their similarity can be easily measured using the Euclidean distance.

4. Research Findings

Tables 2-5 show semantic similarities measured using each tool described in the subsection Methods, Models, Algorithms, and Tools above. The tables show semantic similarities between the computer and data organizations' codes of conduct or ethics and descriptions of ethical schools of thought by Quinn (2015). Similarity percentages were rounded to the nearest integer to preserve uniformity of data since *Compare Text* reports similarities only as integer percentages. Findings for each code are summarized on the right and for each theory at the bottom of the tables.

When the averages of theories of ethics on the tables are compared, the numbers are not uniform, and they vary from one tool to another between orders of 5 to 10. In Tables 2 and 4, both *difflib* and *gensim-nltk* models reported similarities of the same order of magnitudes. However, while the *difflib* model did not point out a dominant theory, the *gensim-nltk* model yielded the highest similarities for the social contract. The DSA's and the O-M's codes on Table 4 scored the same similarities as the social contract for three (95%) schools of thought and one (97%) school of thought, respectively.

The smallest similarities were reported by the *gensim-nltk* model shown in Table 3. Although the average score of 9% similarity is the highest for rule-utilitarianism, only four out of seven organizations scored highest for that theory. In fact, the model's highest scores are the most heterogeneous among the four models.

Finally, similarities reported by *Compare Text* tool, shown in Table 5, were between 45-57%: neither as low as those of the *difflib* model nor as high as those of the *gensim-nltk* and the *spaCy* models. In this last model, the social contract was the sole dominating theory of ethics for all organizations.

Figure 1 further summarizes the findings shown in Tables 2-5 on four radar charts, each for one model. The charts summarize both the overall influence of the schools of ethical thought and the effect of each school on a particular organization in detail. Whereas the results of the *difflib* model (Figure 1.a) are not so discernible, the *gensim-nltk* model (Figure 1.b) reflects the heterogeneity of its results. The dominance of the social contract among the five schools of ethical thought is obvious for the *spaCy* (Figure 1.c) and the *Compare Text* (Figure 1.d) models.

Table 2: Similarities found by using difflib

	Theories of Ethics								
Organizations	*Virtue Ethics*	*Kantian Ethics*	*Act-Utilitarianism*	*Rule-Utilitarianism*	*Social Contract*	*AVERAGE*	*MAX*	*MIN*	*RANGE*
ACM	97%	97%	**99%**	98%	**99%**	98%	99%	97%	2%
ASA	94%	96%	96%	96%	**97%**	96%	97%	94%	3%
BCS	85%	88%	88%	87%	**89%**	87%	89%	85%	4%
INFORMS	86%	**91%**	**91%**	89%	90%	89%	91%	86%	5%
DSA	93%	93%	**94%**	**94%**	92%	93%	94%	92%	2%
IEEE	76%	75%	76%	**81%**	77%	77%	81%	75%	6%

O-M	94%	**97%**	95%	95%	96%	95%	97%	94%	3%
AVERAGE	89%	**91%**	**91%**	**91%**	**91%**				
MAX	97%	97%	99%	98%	99%				
MIN	76%	75%	76%	81%	77%				
RANGE	21%	22%	23%	17%	22%				

Table 3: Similarities found by using *gensim* and *nltk*

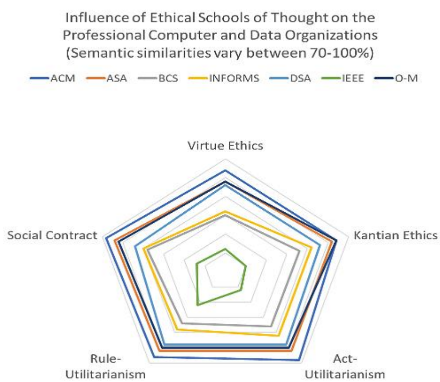
Organizations	***Theories of Ethics***					*AVERAGE*	*MAX*	*MIN*	*RANGE*
	Virtue Ethics	*Kantian Ethics*	*Act-Utilitarianism*	*Rule-Utilitarianism*	*Social Contract*				
ACM	6%	10%	9%	**14%**	8%	9%	14%	6%	8%
ASA	4%	5%	4%	**6%**	5%	5%	6%	4%	2%
BCS	7%	**12%**	11%	10%	7%	9%	12%	7%	5%
INFORMS	4%	4%	**7%**	6%	5%	5%	7%	4%	3%
DSA	5%	5%	7%	**8%**	5%	6%	8%	5%	3%
IEEE	7%	9%	**11%**	9%	8%	9%	11%	7%	4%
O-M	5%	**10%**	8%	**10%**	5%	8%	10%	5%	5%
AVERAGE	5%	8%	8%	**9%**	6%				
MAX	7%	12%	11%	14%	8%				
MIN	4%	4%	4%	6%	5%				
RANGE	3%	8%	7%	8%	3%				

Table 4: Similarities found by using *spaCy*

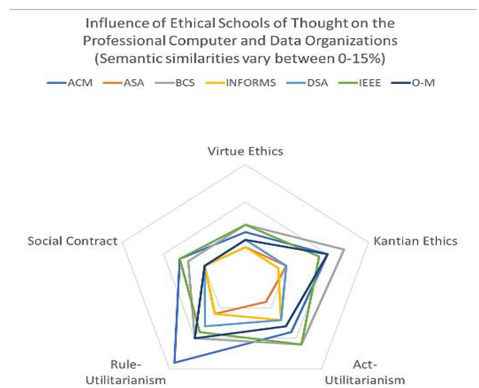
	Theories of Ethics								
Organizations	*Virtue Ethics*	*Kantian Ethics*	*Act-Utilitarianism*	*Rule-Utilitarianism*	*Social Contract*	*AVERAGE*	*MAX*	*MIN*	*RANGE*
ACM	96%	96%	96%	95%	**98%**	96%	98%	95%	3%
ASA	96%	95%	96%	95%	**97%**	96%	97%	95%	2%
BCS	97%	97%	97%	96%	**98%**	97%	98%	96%	2%
INFORMS	95%	96%	95%	95%	**97%**	96%	97%	95%	2%
DSA	94%	**95%**	**95%**	**95%**	**95%**	95%	95%	94%	1%
IEEE	96%	96%	96%	95%	**98%**	96%	98%	95%	3%
O-M	95%	96%	**97%**	96%	**97%**	96%	97%	95%	2%
AVERAGE	96%	96%	96%	95%	**97%**				
MAX	97%	97%	97%	96%	98%				
MIN	94%	95%	95%	95%	95%				
RANGE	3%	2%	2%	1%	3%				

Table 5: Similarities found by using Compare Text

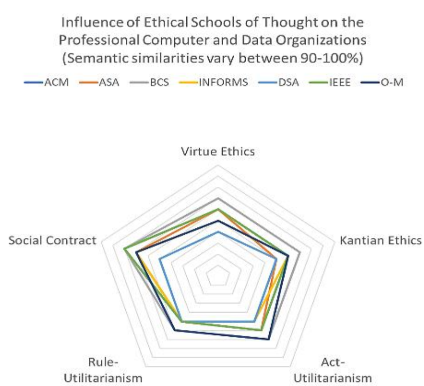
	Theories of Ethics								
Organizations	*Virtue Ethics*	*Kantian Ethics*	*Act-Utilitarianism*	*Rule-Utilitarianism*	*Social Contract*	*AVERAGE*	*MAX*	*MIN*	*RANGE*
ACM	46%	46%	53%	50%	**54%**	50%	54%	46%	8%
ASA	49%	48%	52%	49%	**54%**	50%	54%	48%	6%
BCS	49%	49%	48%	46%	**57%**	50%	57%	46%	11%
INFORMS	48%	47%	48%	46%	**54%**	49%	54%	46%	8%
DSA	53%	48%	55%	55%	**56%**	53%	56%	48%	8%
IEEE	53%	49%	50%	49%	**57%**	52%	57%	49%	8%
O-M	46%	45%	49%	47%	**50%**	47%	50%	45%	5%
AVERAGE	49%	47%	51%	49%	**55%**				
MAX	53%	49%	55%	55%	57%				
MIN	46%	45%	48%	46%	50%				
RANGE	7%	4%	7%	9%	7%				



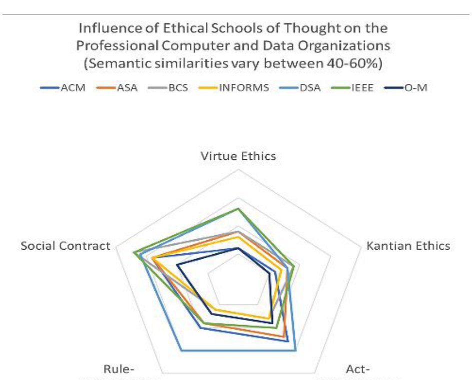
(a) The *difflib* model



(b) The *gensim-nltk* model



(c) The *spaCy* model



(d) The *Compare Text* model

Figure 1: Radar diagrams for the results of the four models

5. Discussion and Conclusion

When the results of Tables 2-5 and Figures 1.a-d are considered, there is significant overlap and slight differences in influences of schools on the codes. The overlap can be attributed to the shared context: ethics. In the texts of theories, there are references to the rules, morals, and sometimes law. References to ethically acceptable and unacceptable behavior are common. So are the adjectives used to qualify behavior.

Overall, this study shows that different schools of ethical thought affect the codes of conduct or ethics of professional computer and data organizations at slightly different levels. In their current state, the research and findings are sufficient as a pilot study. However, each model used to determine the dominance of the theories of ethics should be further investigated in detail if any room for improvement, especially in data preparation, exists.

Unfortunately, that is not the case for Compare Text since the proprietary nature of the tool used to conduct the study does not allow researchers to verify the algorithm to process data and obscures its transparency. The need for using open-source tools to make the analyses and compare their results with this one to validate further the current study results is obvious. Also, multiple data sources for the descriptions of the theories could be sought.

Only one pre-trained generic pipeline was used in the study. However, it could be more helpful to work with a corpus focused on ethics literature. On the other hand, the generic pipeline might help the AI interpret the text similar to that can be interpreted by a layperson. Other pre-trained pipelines, such as WordNet could be tried.

Overall, the current study provided some sense of the dominance of a specific theory of ethics on the computing and data organizations' codes of conduct or ethics. If necessary, refinements can be made; the approach can be extended to the codes of conduct or ethics of corporations, industry sectors, government agencies, and non-governmental organizations.

Author Contribution

CONTRIBUTION RATE	EXPLANATION	CONTRIBUTORS
Idea or Notion	Form the research idea or hypothesis	M. Murat Albayrakoğlu Mehmet N. Aydın
Literature Review	Review the literature required for the study	M. Murat Albayrakoğlu
Research Design	Designing method, scale, and pattern for the study	M. Murat Albayrakoğlu Mehmet N.Aydın
Data Collecting and Processing	Collecting, organizing, and reporting data	M. Murat Albayrakoğlu
Discussion and Interpretation	Taking responsibility in evaluating and finalizing the findings	M. Murat Albayrakoğlu Mehmet N. Aydın

Conflict of Interest

No conflict of interest was reported by the authors.

Financial Support

The author(s) has not received any financial support for this study.

References

Cortical.io. (2021). *Compare Text*. Retrieved May 20, 2021, from https://www.cortical.io/freetools/compare-text/.

CRAN. (2021). *CRAN task view: Natural language processing*. Retrieved June 05, 2021, from: https://cran.r-project.org/web/views/NaturalLanguageProcessing.html.

Deigh, J. (2010). *An Introduction to ethics*. Cambridge.

Donlevy, J. K., & Walker, K. D. (2011). *Working through ethics in education and leadership: Theory, analysis, plays, cases, poems, prose, and speeches*. Sense Publications.

Gomaa, W. H., & Fahmy, A. A. (2013). "A survey of text similarity approaches.", *International Journal of Computer Applications, 68*(*13*), 13-18.

Han, J., Kamber, M. & Pei, J. (2012). *Data mining: Concepts and techniques*, 3rd ed. Morgan Kaufmann.

Ignatow, G. & Mihalcea, R. (2018). *An introduction to text mining: Research design, data collection, and analysis*. SAGE Publications.

Leskovec, J., Rajaraman, A. & Ullman, J. D. (2020). *Mining of massive datasets*, 3rd ed. Cambridge.

McKee, A. (2003). *Textual analysis: A beginner's guide*. SAGE Publications.

Kumar, A., & Kumar, P. (2016). *Mastering text mining with R*. Packt Publishing.

O'Leary, D. (2016). "Ethics for Big Data and Analytics." *IEEE Intelligent Systems, 31*, 81-84.

NLTK. (2021). *NLTK 3.6.2 documentation: Natural Language Toolkit*. Retrieved June 21, 2021, from https://www.nltk.org/.

Oxford-Munich Code of Conduct. (n.d.). *Code of conduct.* Oxford-Munich Code of Conduct for Professional Data Scientist. Retrieved May 19, 2021, from http://www.code-of-ethics.org/code-of-conduct/.

Python. (2021). *Python documentation—The Python standard library—Text processing services: difflib—Helpers for computing deltas*. Retrieved June 06, 2021, from https://docs.python.org/3/library/difflib.html.

Quinn, M. J. (2015). *Ethics for the information age*, 7th ed. Pearson.

Rachels, J., & Rachels, S. (2015). *The elements of moral philosophy, 9th ed*. McGraw-Hill.

Ratcliff, J. W. & Metzener, D. E. (1988). *Pattern matching: The gestalt approach*. Retrieved Nov. 11, 2021, from https://www.drdobbs.com/database/pattern-matching-the-gestalt-approach/184407970?pgno=5.

Řehůřek, R. (2021a). *GENSIM topic modeling for humans: What is Gensim?* Retrieved on June 21, 2021, from https://radimrehurek.com/gensim/intro.html.

Řehůřek, R. (2021b). *models.tfidfmodel – TF-IDF model.* Retrieved on June 25, 2021, from https://radimrehurek.com/gensim/models/tfidfmodel.html.

Rozeva, A., & Zerkova, S. (2017). "Assessing semantic similarity of texts: Methods and Algorithms," in the *Proceedings of the 43rd International Conference of Applications of Mathematics in Engineering and Economics, AIP Conf. Proc. 1910*, 060012-1-060012-8.

Shafer-Landau, R. (2012). *The fundamentals of ethics, 2nd ed*. Oxford.

Shafer-Landau, R. (2019). *Living ethics: An introduction with readings*. Oxford.

Shrestha, P. (2011). "Corpus-based methods for short text similarity." *Rencontre des Etudiants Chercheurs en Informatique pour le Traitement Automatique des Langues, June 2011, Montpellier, France. 2* (*1*), pp.297-302.

spaCy. (2021a). *spaCy: Industrial-strength natural language processing in Python*. Retrieved July 01, 2021, from https://spacy.io/.

spaCy. (2021b). *spaCy: Trained Models & Pipelines*. Retrieved July 01, 2021, from https://spacy.io/models.

Vijaymeena, M. K., & Kavitha, K. (2016). "A survey on similarity measures in text mining." *Machine Learning and Applications: An International Journal, 13*(*1*), 19-28.

Webber, F. E. S. (2016). *Semantic folding: Theory and its application in semantic fingerprinting [White paper]*. Retrieved May 20, 2021, from https://www.cortical.io/static/downloads/semantic-folding-theory-white-paper.pdf.

Resume

Mehmet Murat ALBAYRAKOĞLU (Ph.D. Student) works as an adjunct instructor at Kadir Has University, Faculty of Economics, Administrative and Social Sciences. He is also a Ph.D. student in the Management Information Systems Program of Kadir Has University. He holds a bachelor's degree in metallurgical engineering and graduate degrees in mechanical engineering, business administration, and IT and management. He had several publications in optimization and multi-criteria decision making (MCDM) with analytic hierarchy process (AHP) in manufacturing, technology management, and supply chain management (SCM).

Mehmet N. AYDIN (Assoc. Prof.) is an Associate Professor at Kadir Has University, Faculty of Economics, Administrative, and Social Sciences, Department of Management Information Systems. He completed his Ph.D. at the University of Twente in Management Information Systems Department. His area of research includes cloud computing, innovation network, e-business, and information system (IS) design. He has published over 50 articles in international journals, including Information Systems Frontiers, IEEE Access, Applied Sciences, Computers & Education, Computers and Electronics in Agriculture, and Journal of Enterprise Information Management.