Journal of System Management (JSM)
Online ISSN: 2538-1571
Print ISSN: 2322-2301
6(2), Summer 2020, pp. 155-168
Research Paper

Designing an Intelligence Model for Auditing Professional Ethics in Knowledge Contents Production

Ebrahim Nazari Farokhi [1] **Alireza Poorebrahimi** [2]
Mohammad Nazari Farokhi [3]

Abstract

One of the main concerns of scholars is the observance of intellectual property rights in the production of knowledge content. Data and knowledge derived from the results of research need to be produced based on ethical issues in order to provide context for solving problems through innovation, scientific accuracy and relevance on the basis of ethical principles. In this research, the intelligent knowledge audit model was designed. The present research is applied and in the first stage due to the recognition of the dimensions of the model, is exploratory, in the second phase, the comparative study of the selected models, and in the third stage is analytical. The data gathering tool were a researcher-made questionnaire and interview. The questionnaire was analyzed using structural equation modeling, and the conformity between the independent variables and the dependent variable was presented. In the following, a comparative study of the selected models was performed and after analyzing the interviews with the experts, the final model was designed. As a result, the use of structural and semantic analysis methods in sequence can lead to the prevention of repetitive content. This is heavily needed in order to achieve a systematic ethical responsibility.

Received: 12/09/2019 Accepted: 14/09/2020

[1.] Department of Information Technology Management, Science and Research Branch, Islamic Azad university, Tehran, Iran
[2.] Department of Industrial Management, Karaj Branch, Islamic Azad university, Karaj, Iran- Corresponding Author, Email: poorebrahimi@gmail.com
[3.] Department of Information Technology Management, Science and Research Branch, Islamic Azad university, Tehran, Iran

Keywords *Intelligence Model; Audit Professional Ethics; Content; Plagiarism*

1. Introduction

With the advent of modern technologies and the dissemination of information in the form of various social media, the ability to disseminate and efficient use of information and knowledge content has increased. On the other hand, there are opportunities for plagiarism and illegal and immoral use for some people. In some societies, professional ethics have not been developed at the level of enterprises and organizations, nor have been developed in academic centers.One of the most important factors in the lack of ethics is the lack of effective communication between management and academic circles in practice. Managers do not face ethical issues in their profession, due to lack or defect in ethical concerns. So, they self-cure themselves to solve corporate ethical issues and do not get help from universities. In result, those who write about morality, either provides general and common issues, in which do not talking about objective issues or corporate problems, except the general individualistic sermons and preachers, or philosophical issues of ethics. This bitter reality reveals the motivation to address the nature of professional ethics as knowledge (Faramarz Gharamakeli, 2011). Plagiarism means getting (ideas, documents, code, pictures, etc.) of others and appointments (ideas, documents, photos, etc.) in their own name without mentioning the source and reference (Yaghoobi and Khotanlou, 2015). Plagiarism is increasingly becoming a major issue in the academic and educational domains (Vani and Gupta, 2016). Plagiarism refers to unauthorized use of text, cods, and ideas (Nava and Shakery, 2016). Identifying the similarity of texts is one of the branches of text mining, which is used to identify scientific plagiarism (Yagoobi and Khotanlou, 2015).

2. Literature Review

Lamba and Guwilkar (2017) were presented a method for plagiarism detection in the Indian language. They used the Jacquard coefficient method to compare the similarity of the texts after removing the additional words and parsing the text. In this study, Vector space modeling method (VSM) has been used to extract the frequency of words. Also, for recognizing the semantic similarity, fuzzy semantic similarity recognition method was used. After reading the literature of the subject, it was observed that the Bayesian algorithm, N-gram and natural language processing techniques may be suitable for Indian language. Also, winnowing algorithms and fingerprints are implemented for the Malayalam language. Nahas (2017) compared the famous plagiarism detection soft wares. Her article helps organizations to easily select and apply plagiarism detection software that is relevant to the goals of each organization. The result was that the best software for detecting plagiarism is Authenticate. This software can be useful for many languages in the world, such as Arabic, Turkish and Oriental languages and is used in 80% of the ISI (Web of Knowledge). Bakhtyar et al. (2017) the techniques and algorithms for plagiarism detection that can be used and developed in Arabic languages have been investigated. This method emphasizes the recognition of the similarity between abstract, introduction and outcome. It has emphasized on the use of N-Gram methods and fingerprint algorithms to adapt the teacher. The study also claimed that the results could be useful for other Persian, Urdu and Kurdish languages. Steki and Safi Esfahani (2016), a method for detecting scientific theft based on a backup vector machine (SVM) was presented in Persian texts. This new approach is based on the replacement of words used to recognize semantic similarity. To increase accuracy, an analytical tool is used to root the words. Statistical and semantic features are utilized for the use of a backup vector machine. They used the Jacquard coefficient statistical

method, which is based on the Lon Stein distance. In 2010, a research was carried out by Tsatsaronis et al. This research has shown that although statistical measures are used in scientific plagiarism detection methods and have a simpler and more effective implementation of word-for-word scientific theft, but these methods do not contribute to the semantic analysis of textual and non-textual information. In 2010, sousa-silva et al., Conducted a research. The results of this study showed that the replacement of words with semantically related words and the replacement of synonyms are the main characteristics that suggest a study on the "phrase-to-phrase" review. WordNet has been used by more than a few researchers. Al-Shamery and Gheni (2016) consider that finding synonyms (over WordNet) on the same place in comparing documents is to be the proof of semantic plagiarism. Gharavi et al. (2016) proposed a “deep learning based method to detect plagiarism” in the Persian language. In the proposed method, words are represented as multi - dimensional vectors, and simple aggregation methods are used to combine the word vectors for sentence representation. By comparing representations of source and suspicious sentences, pair sentences with the highest similarity are considered as the candidates for plagiarism. The decision on being plagiarism is performed using a two level evaluation method. According to researches, designing an intelligent model that is capable of analyzing the structural and semantic sentences seems to be necessary. Accordingly, in this research, firstly, professional ethics codes have been identified for the elaboration of a systematic ethical charter for the prevention of plagiarism. In the following, statistical methods for designing the proposed model were studied. The study of NLP methods to examine the semantic similarity of knowledge content is one of the other goals of the present research. In the next step, by comparative study of related models, the main dimensions of the proposed model were extracted. Then these dimensions were presented in the form of interviews

with experts. Finally, an intelligent model of ethical audit in producing knowledge content for Islamic Azad University of Iran was designed. After studying the relationship between independent and dependent variables, a comparative study was conducted. For this purpose, an adaptive matrix has been developed that is the main dimension of the models and software selected by the comparative study. In the adaptive matrix, the main indicators of the model are identified, which in the next step these indicators were introduced to the experts. The result of an expert interview resulted in the selection of appropriate indicators for designing the model. Finally, the overall dimensions of the model were determined based on selected indicators by the experts. In Tables 1 and 2, an adaptive matrix and a comparative study from related studies and models are presented respectively.

Table 1.

Comparative Matrix of Scientific Plagiarism Detection Software

Indicators / paradigm	**SIM Model**	**Jplag Model**	**PMDsCPD Model**	**YAP3 Model**	**Sherlock Model**	**Moss Model**
Run type	Console	Run the web, Java	Run the web, Java	---	Run Java	Web service, Perl script
Support language	C, Java, pascal,Lisp, Miranda and Natural language	Java, c#, c, c++, Scheme and Natural Language	Java, JSP, c, c++, Fortran and PHP	Pascal, c, LISP	Programming Language and Natural language	C, c++, Java, Javascript, Pascal, Ada, Lisp, Python, C#, Perl
Methods / algorithms	Tokenization	Tokenization + Grid string alignment	Tokenization	Tokenization + Grid string alignment	Incremental comparison of two files	Winnowing algorithm
GUI	No	Yes	Yes	No	Yes	No

Indicators / paradigm	SIM Model	Jplag Model	PMDsCPD Model	YAP3 Model	Sherlock Model	Moss Model
Similarity measurement method	Except in percent, the number of same rows	Except in percent, histogram, group of similar files	Number of same rows, Signs	---	Sum of Percentages, Similarity graph	Except in percentage
Online / offline	offline	Online	offline	offline	offline	Online

Table 2.

A Summary of the Comparative Study of Research done Differently from the Features and Tasks Involved

No.	Article title and year of publication	Name of Author(s)	Method / technique
1	Toward the discovery of reuse of the multilingual codebook / 2011	Enrique Flores, Alberto Barron, Code no, Paolo Rosso, Lidia Moreno	Method: Comparison of N-Grams Features Model Method used in: Comments and saved words
2	Discovery source code reuse multiple programming languages / 2012	Enrique Flores, Alberto Barron, Code no, Paolo Rosso, Lidia Moreno	Method: Natural language processing Tools: Compare source code in terms of functions and method
3	The discovery of software plagiarism with a graph approach / 2013	Dong-Kuy, Chae giwoon Ha, sang-Wook, Kim, Boo Joong, Kang, Elu Gyulm	API Tag and Control flow Graph (A-CFG)
4	Experience of discovering the division of codes in Microsoft / 2011	Yin Gong Dang, Song Ge, Ray, Huang and Dong mei Zhang	Tools: XIAO tools were created by Microsoft researchers and was shared for testing
5	Plagiarism discovery engine for Java source: an approach based on clustering / 2007	Ameera Jadalla, Ashraf Elnagar	Technique: Data analysis, Clustering, Tokenization, N-Gram

No.	Article title and year of publication	Name of Author(s)	Method / technique
6	Exclusive study of source code / 2010	Mark Gabel, Zhen dong Su	Method: Tokenization, Lexical analysis, Tabulation, Languages: Java, C and C ++
7	Discover the source code used in multiple programming languages	Enrique Flores, Alberto Barron, Code no, Paolo Rosso, Lidia Moreno	Method: Features N-Gram Language: Used in Python, Java and C ++
8	A lexical analysis for the conceptual dual evaluation of the C programing language / 2013	Akhil Gupta, Dr.sukhvir Singh	Method: Lexical analysis

According to research literature the research hypotheses are as follows: In examining the similarity of knowledge content, statistical methods are highly superior to natural language processing techniques. A tool called ethical codes has a great influence on the institutionalization of professional ethics in the research process. An intelligent model of professional ethics audit in the production of knowledge content has a great impact on the prevention of plagiarism. The semantic analysis method has a great impact on the qualitative implementation of the discovery plagiarism in the Persian language. Observing socio-cultural factors has a major impact on the prevention of repetitive content production.

3. Method

The present research is an applied research. The research method in the first stage is due to the recognition of the dimensions of model is an exploratory method, in the second phase, the comparative study of the selected models, and in the third stage, is analytic. The data gathering tool in the research includes a questionnaire and an interview prepared by the researcher.

The statistical population in the present research Including 439 faculty members and PhD candidates in the field of IT management at Islamic Azad University of Tehran, in which the number of sample size based on Cochran formula calculated as 205. The results of the questionnaire were prepared using structural equation modeling, analysis and adaptation between independent variables and dependent variable of the research. Finally, for the approval of the main model of the dissertation, an interview was conducted with a total of 23 experts.

4. Findings

Data analysis was performed using LISREL based on confirmatory factor analysis method. After fitting the obvious variables to the independent variables, the factor load of 35 questions out of 38 was higher than 0.5. Only questions 8, 11, and 18 had a factor load of less than 0.5.

Table 3.
Factor Loads and T-Value

No. of Question	Factor load	T-Value	No. of Question	Factor load	T-Value
Q1	0.62	14.25	Q22	0.57	۱۴,۹۴
Q2	0.67	15.88	Q23	0.61	۱۲,۶۵
Q3	0.65	11.78	Q24	0.74	۱۶,۶۴
Q4	0.67	15.22	Q25	.060	۱۵,۷۱
Q5	0.58	13.29	Q26	0.72	۱۶,۵۴
Q6	0.65	14.48	Q27	0.67	۱۰,۹۹
Q7	0.51	11.50	Q28	0.52	۹,۳۴
Q9	0.53	11.86	Q29	0.64	۱۰,۴۳
Q11	0.53	11.49	Q30	0.73	۱۶,۱۸
Q12	0.59	12.69	Q31	0.76	۱۶,۴۶
Q13	0.62	16.18	Q32	0.70	۱۵,۳۱
Q14	0.63	15.24	Q33	0.73	۱۲,۰۷
Q15	0.51	13.73	Q34	0.69	۱۷,۲۲
Q16	0.54	13.55	Q35	0.69	۱۷,۲۵

No. of Question	Factor load	T-Value	No. of Question	Factor load	T-Value
Q17	0.50	14.22	Q36	0.64	۱۵,۹۷
Q19	0.60	11.95	Q37	0.57	۱۴,۹۸
Q20	0.58	14.91	Q38	0.69	۱۱,۱۰
Q21	0.50	15.41			

The amount of factor load and T-value associated with the dependent variable are presented in Table 4.

Table 4.
Factor Loads and T-Value Related to the Dependent Variable

NO.	Title	Factor load	T-Value
39	The superiority of natural language processing on statistical methods	۰,77	19.33
40	Ethical cods	.068	17.43
41	Ethical audit model	۰,77	19.12
42	Semantic analysis	۰,71	17.96
43	Socio-cultural factors	0.77	19.09

The results obtained from the final model extracted from the software, which include the amount of Beta coefficients and T-Value between the variables, are presented in Table 5.

Table 5.
Beta Coefficients and T-Value in the Final Model

Independent and dependent variables		
Title	**Beta coefficients**	**T-Value**
The superiority of natural language processing on statistical methods	0.07	0.50
Ethical codes	0.46	2.13
Ethical audit model	0.38	1.97
Semantic analysis	0.18	2.15
Socio-cultural factors	0.52	3.95

Regarding the values obtained from the final fitting indices in the presented model, it shows that the model has a good fit and thus, the relationship between most independent variables and the dependent variable is confirmed. The only variable whose relationship with the dependent variable was not confirmed was the superiority of statistical methods to natural language processing techniques. In addition, the most important variable was observance of socio-cultural factors.

Table 6.
The Fitting Indices of the Final Presented Model

No.	Name of	The obtained amount
1	RMSEA	۰,۰9
2	X2/df	2.81
3	CFI	۰,۹۷
4	GFI	۰,67
5	AGFI	0.62

The result of an interview with the experts is described in the table 8 to be taken into account in the model, through each of these indicators, in order to identify the structural and semantic similarities.

Table 7.
The Results of Interviews with Experts

Run type	**Support language**	**Methods / algorithms**	**GUI**	**Check the result**	**Similarity measurement method**	**Online/ Offline**
Under the web	C or C #	Tokenization	Yes	Structural and Semantic Similarities	Except in percent, the number of same rows	Online / offline

5. Discussion and Conclusions

One aspect of innovation in this model is to examine both the structural and semantic similarities of sentences which have not been done in other related researches in the field of plagiarism. Investigating the degree of structural similarity by translating each word into its original root can be used as a way to determine the meaning of the content's semantic similarity in the designed model, which is consistent with the study of Fullam and park 2002. After the similarity of generated content has been revealed, in order to observe ethical responsibility in the research, consistent with each part of the research process in which the violation was committed, systematic ethical codes are presented to prevent the occurrence of plagiarism. This is consistent with a study by Vedadeir et al. (2008). Another controversy with other researches is the use of the context box, in which case the linguistic structure of each research to check the similarity level is determined and the model does not fall into the trap of the comparison. The use of semantic analysis methods in this study is consistent with the research of sousa-silva et al. (2010), in which vocabulary replacement with semantic vocabulary was also carried out to investigate plagiarism. After the formation of the texture and filtering the scope of the review through the context box, the removal of public terms and pre-processing occurs. In the following, documents are reviewed based on the author's name, organization and place of research. Then the vocabulary matrix will be based on the verb, subject, object, adjective, and adjective, in order to examine the degree of structural similarity of the generated content. Based on the initial presented model, an interview was conducted with experts and the main dimensions of the final model of the research were extracted and the final model of the research was designed. It is suggested that the result of the study of the structural similarity with the result obtained by examining the degree of semantic similarity be considered in the form of a model design and

the Ethics Committee held at each faculty and this committee take action about explaining and describing ethical codes, appropriate to each part of the research process.

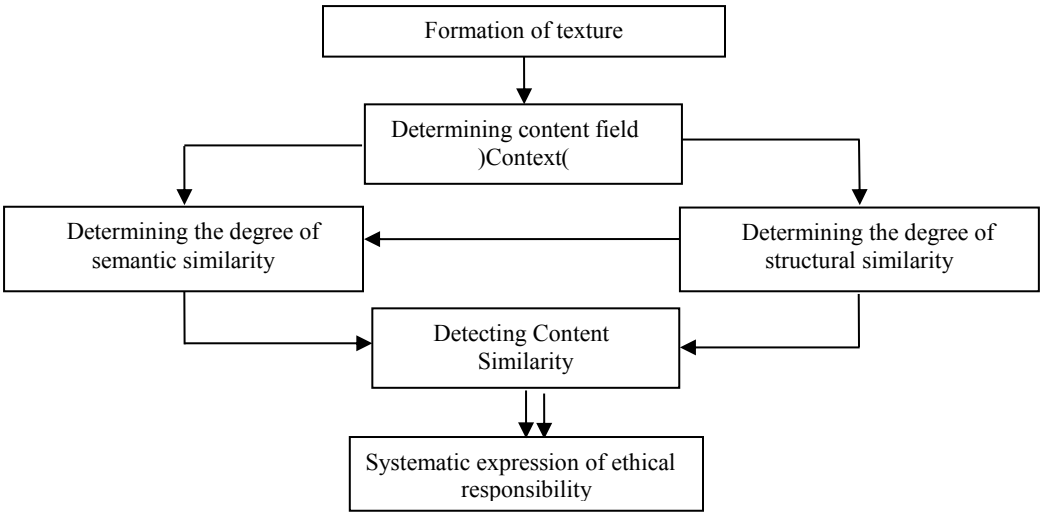

Fig. 1. Final Presented Model

In the present research, in order to identify the structural similarity, Labeling to words was used. This issue is also consistent with the research by Nojavan Aqhdargh et al. (2016), in which the research was to identify the structural similarity of words such as nouns, adjectives, and adverbs. Also, the vector-space modeling algorithm was used to detect the similarity of vector-based sentences, which is consistent with the Lamba and Govilkar research carried out in 2017. In this study, in order to recognize the structural and semantic similarity, the similarity of the sentence was made between the source and the destination document, which was consistent with the research by Souse-Silva et al. (2010).

References

Aghakardan, Ahmad, & Keihannejad, Sina. (2012). Provides a model for extracting information from text documents based on text mining in the field of e-learning. *Journal of information & communication technology*, 11 & 12, 47-54.

Al – Shamery, E. S., & Gheni, H. Q. (2016). "Plagiarism Detection using Semantic Analysis. Indian Journal of Science and Technology, 9.

Bakhtyar, Ribwar. Wakil, Karzan. Saeed, Soran (2017). Plagiarism Detection Techniques for Arabic Script Languages: A Literature Review. Kurdistan Journal of Applied Research (KJAR), Volume 2,| Issue 3.

Chong, Man Yan Miranda. (2013). A Study on Plagiarism Detection and Plagiarism Direction Identification Using Natural Language Processing Techniques.

Darueean, Soheila, & Faghihi, Mahdi. (2011). Investigating the motives and causes of plagiarism in Iran. Iranian Journal of Public Administration Mission, 2, No.1.0.

Esteki, Faeze. Safi Esfahani, Faramarz. (2016). A Plagiarism Detection Approach Based on SVM for Persian Texts," in FIRE (Working Notes), pp. 149- 153.

Faramarz Gharamaleki, Ahad; (2011). an introduction to professional Ethic. Tehran, Saramad Publication, 4th Ed.

Fullam, K, & Park, J. (2002). Improvements for scalable and accurate plagiarism detection in digital documen ts. http://www.lips.utexas.edu/~ kfullam/pdf/DataMiningReport. Pdf .

Gharavi, E., Bijari, K., Veisi, H., & Zahirnia, K. (2016). A Deep Learning Approach to Persian Plagiarism Detection. FIRE 2016 - Forum for Information Retrieval Evaluation, Kolkata.

Lamba, Harshall. Govilkar, Sharvari. (2017). A Survey on Plagiarism Detection Techniques for Indian Regional Languages, International Journal of Computer Applications (0975 – 8887). Volume 164 – No 4.0.

Nava, Ehsan, & Shakery, Azadeh. (2016). Candidate document retrieval for cross-lingual plagiarism detection using two-level proximity information. Information Processing and Management journal.

Nahas, Mohmoud Nadin. (2017). Survey and Comparision between plagiarism detection, American journal of Data Mining and Knowledge Discovery; 2 (2): 50-53.

Nojavan Aghdargh, B. Feizi Derakhshi, M. and Samad Zamini, K. (2016). Designing a system for identifying similar sentences for use in multi-document summarization, MA Thesis. University College of Nabi Akram. Faculty of Technical.

Souse-Silva, R., Grant, T. & Maia, B. (2010). I did not mean to steal someone else's words: A Forensic linguistic approach to detecting intentional plagiarism. 4th international plagiarism conference: Towards an authentic future.

Tsatsaronis, G., Varlamis I., Giannakoulopoulos, A., & Kanellopoulos, N. (2010). Identifying free text plagiarism based on semantic similarity. 4th International Plagiarism Conference.

Vani Ka, & Deepa Guptab. (2016) Detection of idea plagiarism using syntax–Semantic concept extractions with genetic algorithm. Expert Systems with Applications journal.

Vedadhir, Abuali, Farhud, Daryush, Ghazi Tabatabaee, Mahmud, & Tavasoli, Gholamabbas. (2008). Ethical implications for scientific work, contributing to the sociology of morality in the science and technology of Merton and Rosenick. The journal of ethics in science and technology, (3), volume 3, 4.

Yaghoobi, Rezvan, & Khotanlu, Hasan. (2015). Plagiarism detection in the scientific papers using semantic role labeling and Genetic algorithm. Journal of Electronic Industry, 6.