ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Text mining-based analysis of immigrant cultural dislocation: memory and identity construction in the works of Kazuo Ishiguro and Zadie Smith

Haoran Chen[1,*]

[1] School of Teacher Education, Nanjing Normal University, Nanjing, Jiangsu, 210023, China

SUMMARY: *Multicultural writing in immigrant literature presents a new narrative style of world literature in the era of globalization, showing the qualities of different cultural games and mingling, and is also an important manifestation of the evolution of the traditional European colonial relationship in the process of globalization. Taking the works of Kazuo Ishiguro and Zadie Smith as an example, the study crawls the readers' comment data of related works on relevant websites based on the text mining method, and utilizes the methods of TF-IDF algorithm, LDA theme model, and sentiment analysis to carry out the mismatch analysis of immigrant culture, and to shape the image of memory and identity of immigrant culture. The results of the study show that the works of Kazuo Ishiguro and Zadie Smith can be categorized into five themes, namely, identity and belonging, cultural conflict and integration, immigrant experience and intergenerational, social and political contexts, and emotional and psychological states, and that the readers' perception of immigrant culture of Kazuo Ishiguro's and Zadie Smith's works is dominated by positive emotions, with a positive tendency of 92.04%. Based on the results of IPA quadrant distribution, it can be obtained that the number of elements falling into the second four quadrants is the highest, i.e., the works of the two authors are able to express the content of immigrant's memory and identity well, and the readers are able to explore a lot of literary knowledge about immigrant's memory and identity construction in Kazuo Ishiguro's and Zadie Smith's works.*

KEYWORDS: *lda topic model; tf-idf algorithm; text mining; sentiment analysis; immigrant culture*

1 Introduction

Migration is a unique trend of our time, a marvelous mass occurrence of the late 20th century [1, 2]. Accompanying the phenomenon of immigration, immigrant literature has also occupied an increasingly important position in the contemporary world literary scene. Kazuo Ishiguro and Zadi Smith, as typical representatives of immigrant writers, through their exploration of "identity" in their works, reveal the memory reconstruction and identity crisis that individuals face when crossing cultural boundaries in the context of globalization [3-6].

Kazuo Ishiguro is a Japanese-British novelist whose major works include The Faint Shadow of Faraway Hills, The Painter of the Floating World, and The Long Day Will End [7, 8]. He was awarded the 1989 Booker Prize, the 2017 Nobel Prize for Literature, the Order of the British Empire, and the Chevalier de l'Ordre des Arts et des Lettres of France, and has been referred to as the “Immigrant Trio of British Literature” along with Rushdie and Naipaul [9-

*yxybrucec13@163.com
https://doi.org/10.65102/is2026013 Published: 30 April, 2026

11]. Most of Kazuo Ishiguro's works are written on the theme of memory, bridging the past and the present through memory, which not only recounts the protagonist's past events but also shows some of the protagonist's views and attitudes towards the facts of the past, reflecting how people nowadays should take the initiative to face the memories of the past with the development and changes of the society [12-15]. By taking the war as the background, the author, with the help of the fragmented memories of the main character, the retrospective narrative structure and identity construction avoids directly confronting the disasters and pains of the past, but reflects on itself through memories, reconstructs its identity, and adapts to the development of the society [16-19].

And Zadie Smith is a contemporary British writer born in London in 1975. Her father is British and her mother is a Jamaican immigrant [20, 21]. Her interracial family background has brought material and inspiration to her writing, and to date, Zadie Smith has published five full-length novels, White Teeth, Signers, On Beauty, Northwest, and Swing Time [22-24]. Smith's humorous writing style has been well received by the public, and her reflections on racial identity and multiculturalism in her works have also attracted critical attention, with critics seeing her as a representative of “race, youth, and women” and a spokesperson for multiculturalism [25-28].

With the development of artificial intelligence, text mining provides a quantitative perspective for the study of immigrant literature, which uses natural language processing techniques to extract keywords and emotional tendencies from the literary works of Kazuo Ishiguro and Zadie Smith, and quantitatively analyzes the cultural conflicts of the immigrant narratives therein, thus presenting the fractured memories and fluidity of identities in cultural dislocation [29-32].

This paper firstly introduces the algorithms and principles of text mining related methods, mainly including TF-IDF feature extraction algorithm, SnowNLP, SO-PMI and LDA algorithm. Based on the text mining method, we crawl the data of readers' comments about the works of Kazuo Ishiguro and Zadie Smith on Amazon.com, the mainstream book trading platform in the West, and construct a text database. On the basis of the processed data, the immigrant culture in the works is mismatched and analyzed using the method of this paper, specifically, feature words are extracted using the TF-IDF algorithm, the word frequency characteristics of feature words are analyzed, and the semantic network between the high-frequency feature words is constructed using the ROST CM6, to explore the correlation relationship between the feature words. After that, all the readers' comments are clustered by LDA theme model, and five optimal themes are derived Finally, the readers' emotional image perception of Kazuo Ishiguro's and Zadie Smith's works is deeply explored from the aspect of emotional image, and the ideas and values of Kazuo Ishiguro's and Zadie Smith's works are clarified.

2 Method

2.1 Chinese Segmentation and Deactivation Processing

Chinese word segmentation refers to cutting Chinese text into individual words or word string sequences through word segmentation technology and separating them by spaces. Chinese word separation is the basis and premise of deep text mining, after the Chinese word processing text data can be transformed into the form of mathematical vectors, to facilitate the subsequent analysis. At present, the mainstream Chinese lexical tools are: JieBa, THULAC, pyltp, SnowNLP and so on. Because JieBa has higher efficiency and accuracy in word separation, and can import customized dictionaries such as “immigrant culture”, “identity” and other proper nouns for Chinese word separation, this study chooses the JieBa toolkit for Chinese word

separation in Python. The main functions of JieBa library include word segmentation, custom dictionary, keyword extraction and lexical annotation, etc. It supports three word segmentation modes: precise mode, full mode and search engine mode. Deactivation of words in the text processing process will produce greater interference, the deactivation of words itself carries less useful information, but also on other words to produce a certain inhibition, to a large extent, will affect the efficiency of text processing and the accuracy of the language. Generally speaking, stop words can be roughly divided into two categories: generic stop words and proprietary stop words. Common stop words are the most widely used in text, appearing in almost all documents, such as degree adverbs, modal particles, prepositions and other words that do not contain useful information themselves but have no practical meaning, like "although", "if", "but", etc. These words usually appear more frequently, but have no practical meaning by themselves, and they need to be eliminated in text mining so as not to affect the research results. The other category is proprietary deletion words, which are deletion words applicable to specific domains. Therefore, this study intends to construct a new deactivation lexicon for filtering irrelevant words on the basis of Chinese participles.

2.2 Semantic network modeling

Semantic network model is widely used in the field of text mining. The semantic network analysis model is mainly composed of two parts: nodes and links. Nodes are used to represent words that appear more frequently or are more important in the text, and links represent the connections between different nodes. Semantic network graph is a visual graph generated by semantic network analysis, which can show the strength of network connections between different nodes, and is used to explore the location of nodes, node density and connections between nodes. The degree of a node is used to represent the number of lines connected to a node, usually, the higher the degree of a node, the more lines it has, and it can be considered as a key node.

2.3 TF-IDF feature extraction algorithm

TF-IDF also known as Word Frequency-Inverse Document Frequency is a widely used weighting technique, which is a feature extraction algorithm mainly used to evaluate the importance of words in a collection of documents and is commonly used in the field of natural language processing and text mining.

The TF-IDF algorithm measures whether a word is universally important or not by introducing the inverse document frequency of the word. In this approach, the importance of a word is proportional to the word frequency and inversely proportional to the frequency of its occurrence in other documents. In this way, some common high frequency words are filtered.The formula for TF (word frequency) is shown in equation (1):

$$TF\left(t_i, d_j\right) = \frac{N_{t_i}}{\sum_{i=1}^{n} N_{t_i}} \tag{1}$$

where $TF\left(t_i, d_j\right)$ denotes the word frequency of the word t_i in the document d_i, and N_{t_i} denotes the total number of document words.

The formula for IDF (Inverse Document Frequency) is shown in equation (2):

$$IDF\left(t_i\right)=\log\left(\frac{N}{df\left(t_i\right)}+1\right) \tag{2}$$

where $IDF\left(t_i\right)$ represents the inverse document frequency of the word, $df\left(t_i\right)$ tabulates the number of documents containing the word in the dataset, i.e., document frequency, and N represents the total number of documents in the dataset. In order to avoid a word appearing in all documents at the same time, resulting in 0, therefore, smoothing is done.

The value of TF-IDF is composed of the product of TF and IDF, and the following equation (3) is used to calculate the $TF-IDF$ value of the terms in the documents in the term-document matrix:

$$TF-IDF\left(t_i,d_j\right)=TF\left(t_i,d_j\right)*IDF\left(t_i\right)=\frac{N_{t_i}}{\sum_{i=1}^{n}N_{t_i}}*\log\left(\frac{N}{df\left(t_i\right)}+1\right) \tag{3}$$

2.4 LDA model

The Latent Dirichlet Allocation (LDA) model is widely used in research fields such as online public opinion analysis, research hotspot extraction, and library intelligence because it has a clear intrinsic structure and is suitable for large corpora. In recent years, scholars have gradually begun to use LDA model to analyze accident texts.LDA model is a document generation model, which can be regarded as a three-layer Bayesian model and belongs to the unsupervised machine learning method, and it can effectively mine the latent theme information in the text. This study uses this model to analyze and process the text in the works of Kazuo Ishiguro and Zadie Smith, expecting to discover different themes by mining the text, and finally presenting the results by means of visualization.

The schematic diagram of LDA model is shown in Figure 1. Where α,β is the Dirichlet distribution hyperparameters, θ_i represents the distribution of themes of the accidental text i, $z_{i,k}$ denotes the k th theme of the text of the works of Kazuo Ishiguro and Zadie Smith, $w_{i,j}$ denotes the i th feature word of the text, φ_k denotes the distribution of feature words representing the i th topic. M represents the number of texts in the text set, N is the number of feature words in each text, and K represents the total number of topics.

Combined with the specific object of this study, the main generation process of the accident causation LDA model is as follows: ① For the text of Kazuo Ishiguro and Zadie Smith's work i, the number of feature words in the text N_i is generated according to $N\sim Poisson(\xi)$. ② For the text, generate the parameter θ_i of the polynomial distribution of the text about the topic according to $\theta_i\sim Dir(\alpha)$. ③ For the subject k, generate the parameters of the subject's polynomial distribution $\varphi_k\sim Dir(\beta)$ about the feature words in the text base, according to φ_k. ④ For the i th feature word $w_{i,j}$ of the text, sample the topic $z_{i,k}$ to which the feature word term belongs according to the polynomial distribution $z_{i,k}\sim Multi\left(\theta_i\right)$. The specific feature words are sampled according to the polynomial $w_{i,j}\sim Multi\left(\varphi_k\right)$.

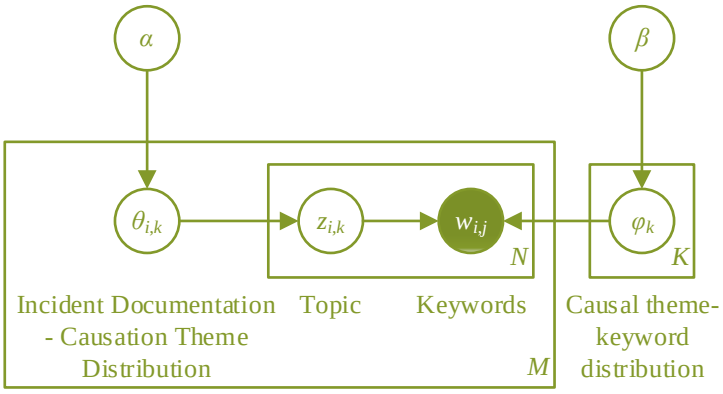


Figure 1: Schematic diagram of the LDA model

The joint and posterior probabilities of the LDA model, given the Dirichlet distribution hyperparameters α, β, are shown in equations (4) and (5), respectively:

$$p\left(\varphi_k, \theta_d, z_{i,k}, w_{i,j}\right) = \prod_{i=1}^{K} p\left(\varphi_k\right) \prod_{d=1}^{D} p\left(\theta_d\right) \left(\prod_{n=1}^{N} p\left(z_{i,k} \mid \theta_d\right) p\left(w_{i,j} \mid \varphi_k, z_{i,k}\right) \right) \tag{4}$$

$$p\left(\varphi_k, \theta_d, z_{i,k} \mid w_{i,j}\right) = \frac{p\left(\varphi_k, \theta_d, z_{i,k}, w_{i,j}\right)}{p\left(w_{i,j}\right)} \tag{5}$$

2.5 Analysis of the emotional tendency of the text

2.5.1 SO-PMI algorithm

SO-PMI is the Sentiment Orientation Point Mutual Information Algorithm. The central idea of this algorithm is to determine the degree of association between unfamiliar words and positive and negative benchmark words. If the degree of association with the positive benchmark word is large, the word is judged to be positive. If the degree of association with the negative benchmark word is large, the word is judged to be negative. If it is independent of both the positive and negative benchmark words, the word is judged to be neutral. The formula is as follows:

$$SO-PMI(word) = \sum_{i=1}^{num(pos)} PMI\left(word, pos_i\right) - \sum_{i=1}^{num(neg)} PMI\left(word_n neg_i\right) \tag{6}$$

Here num(pos) refers to the total number of positive benchmark words and num(neg) refers to the total number of negative benchmark words. The formula brings the following results:

SO-PMI>0, words are judged as positive words.

SO-PMI=0, words are judged as neutral words.

SO-PMI<0, words are judged as negative words.

Finally, 600 positive emotion words and 600 negative emotion words are intercepted and added to the Knowledge Sentiment Dictionary, and de-emphasized, and finally aggregated into the exclusive sentiment dictionary of Kazuo Ishiguro and Zadie Smith's works consisting of 5,366 positive emotion words and 4,912 negative emotion words, and the SpowNLP model is trained with the final obtained exclusive sentiment dictionary of Kazuo Ishiguro and Zadie Smith's works.

2.5.2 SnowNLP Sentiment Analysis

SnowNLP Sentiment Analysis is a natural language processing technique based on Chinese text, which aims to extract the sentiment features in the text in order to determine the sentiment tendency of the text. The basic principle is: firstly, according to the content of words, the text is divided into positive and negative sentiment distinctions. Secondly, according to the construction of sentences, the text is classified into positive and negative sentiment distinctions. Finally, based on the correlation relationship between sentences, the text is classified into positive and negative sentiment distinctions. In this study, the Bayesian principle of "document-topic-word" is used for training and prediction of data. The formula is as follows:

$$p = p(C)\prod_{i}^{n} p\left(d_i \mid C\right) = p(C)\prod_{i}^{n}\left(\frac{count\left(d_i, C\right)}{T_c}\right) \tag{7}$$

$count\left(d_i, C\right)$ denotes the frequency of vocabulary d_i appearing in the text C, T_C denotes the total number of positive texts, and n represents the number of phrases in the categorized text. SnowNLP module when P is greater than 0.5 indicates that the text is positive, and P is less than 0.5 indicates that the text is negative.

3 Results and Discussion

3.1 Text data processing

3.1.1 Text acquisition

We used the R language to collect readers' comments on the works of Kazuo Ishiguro and Zadie Smith on Amazon using the search terms "memory," "culture," "history," and "identity," respectively. Reviews were collected from December 22, 2021, to January 3, 2024, with the following information: reviewers, ratings, and reviews. The collection includes the reviewer, rating star, time of review, region of the reviewer, title of the review, content of the review, and number of raters. The preliminary data collected were deposited into csv files for text data analysis.

3.1.2 Text cleaning

Firstly, the initial cleaning is performed, and non-English comments, duplicate text, no text comments, irrelevant content, pure emoticons, etc. are manually deleted, and a total of 258 valid comments are obtained after cleaning. Secondly, text noise reduction is performed, using R code for text segmentation, letter case conversion, word shape reduction, and setting deactivated words. Finally, based on the cleaned and noise-reduced text, R is used to visualize the comment text of Kazuo Ishiguro and Zadie Smith's works.

3.2 Text characterization

3.2.1 Text Word Frequency Feature Analysis and Visualization

TF-IDF algorithm can effectively highlight the importance of core words in a document, making the extracted words more meaningful. Therefore, this study uses the TF-IDF algorithm to extract the top 40 feature words in the works of Kazuo Ishiguro and Zadie Smith, and the ranking and frequency of feature words in the works of Kazuo Ishiguro and Zadie Smith are

shown in Table 1. Among the top 40 key feature words extracted by the TF-IDF algorithm, most of the key words correspond to a feature of Kazuo Ishiguro's and Zadie Smith's works, and the accuracy rate is good. Among them, “memory” has the highest frequency of 13,473, and the high-frequency words mainly emphasize the idea of the complex experience of immigrants in identity, culture and society expressed in the works of Kazuo Ishiguro and Zadie Smith.

Table 1: Ranking and frequency of feature articles

Ranking	Feature Words	Word Frequency	Ranking	Feature Words	Word Frequency
1	Memory	13473	21	Adaption	2766
2	Past	11515	22	Assimilation	2693
3	Identity	9269	23	Dissimilation	2675
4	Belonging	8459	24	Dignity	2550
5	Homeland	6071	25	Self-Deception	2489
6	Culture	5571	26	Wound	2326
7	Tradition	5529	27	Second Generation	2332
8	Loss	5488	28	Immigrant Community	2302
9	Alienation	5406	29	Mixability	2260
10	Solitude	5232	30	Englishness	2157
11	History	4492	31	Root	2021
12	Race	4426	32	Migration	2008
13	Religion	3508	33	Boundary	1883
14	Generation Gap	3330	34	Language	1800
15	Melding	3059	35	Food	1719
16	Conflict	3002	36	Name	1578
17	Multivariate	3003	37	Ritual	1545
18	Scatter	2924	38	Silence	1407
19	Rooting	2849	39	Oppression	1278
20	Homesickness	2780	40	Hope	1267

3.2.2 Text Feature Correlation Analysis and Visualization

Through the text characterization of TF-IDF above, we can get the characteristics of literary works and emotional attitudes that readers pay attention to, but we can't see the relationship between each high-frequency feature word and the relationship between feature words and emotional attitudes. If you want to explore the connection between high-frequency text features more deeply, you can use the method of covariant semantic network analysis to do the research. Co-word semantic network analysis can count the frequency of a pair of words appearing together in the text, if the frequency of appearing together is higher, the closer the connection between the pair of words is. This method can uncover the correlation between a topic word and another topic word in the same domain, and show the closeness of the relationship between high-frequency words and the layer and relationship between them.ROSTCM6 is a commonly used software for semantic network analysis, and this paper uses this software to explore the correlation between the feature words most frequently mentioned by the readers in the reviews of Kazuo Ishiguro's works and Zadie Smith's. The covariance matrix is shown in Table 2. The co-word matrix is shown in Table 2, and the co-occurrence semantic network is shown in Figure 2. From the co-occurrence matrix and co-occurrence semantic network, it can be seen that the

connection between high-frequency feature words, such as memory, identity, belonging, culture, tradition, etc., is well represented. In the co-occurrence matrix, the feature words that readers focus on have a high number of co-occurrences with each other, such as “culture” and “memory”, which have 3493 co-occurrences. In the semantic network of co-occurrence, the feature words that readers pay attention to, such as “hope”, “religion” and “loss”, are closely related to other feature words, and they are not only the core feature words, but also the focus of readers' satisfaction and emotional attitudes in the semantic network of the collected comments. In the semantic network of the collected comments, they are not only the core feature words, but also the focus of readers' satisfaction and emotional attitudes, and the interconnection between these feature words makes the whole semantic network connected, connecting the most important parts of different commentaries, and these core feature words are the most likely to have an impact on readers' satisfaction.

Table 2: Co-word matrix

	Memory	Identity	Belonging	Culture	Tradition	Alienation	Solitude	History	Race	Generation Gap	...
Memory		2253	2513	3493	2494	1640	1726	1575	1609	710	...
Identity	2253		2504	1851	709	1282	1713	1101			...
Belonging	2513	2504		1926			1653	2583			...
Culture	3493	1851	1926		1653	1222	1173	944	973		...
Tradition	2494	709		1653		1205			1152	896	...
Alienation	1640	1282		1222	1205				1222		...
Solitude	1726	1713	1653	1173				953			...
History	1575	1101	2583	944			953				...
Race	1609			973	1152	712					...
Generation Gap	710				896						...
...	...	...	...	...	...	...	...	...	...	...	...

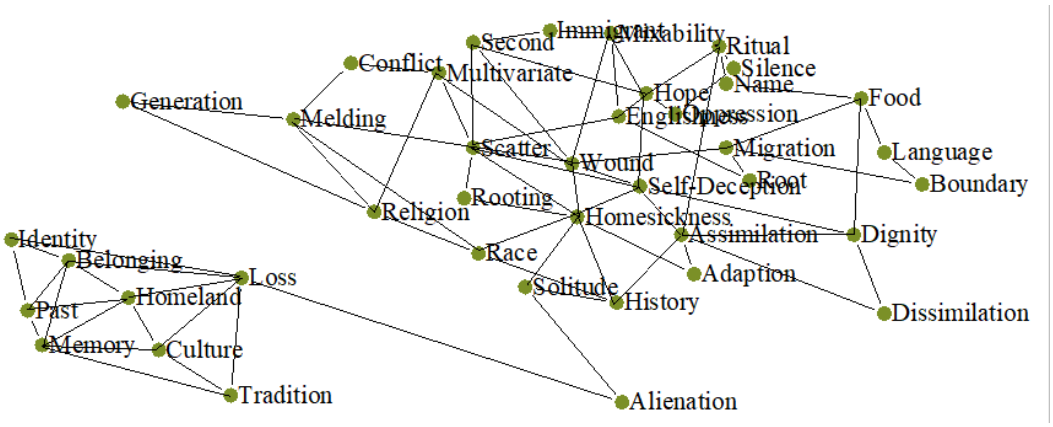


Figure 2: Co-occurrence semantic network

3.3 LDA Thematic Modeling Analysis

3.3.1 Clustering Theme Number K Determination

Before clustering the online user comment text for LDA model, the number of topics of the target, i.e., K-value, needs to be set manually. The perplexity can depict the prediction results under different topics of LDA model, which can help to establish a reasonable K value.The results of perplexity calculation based on online comment clustering are shown in Fig. The

perplexity decreases with the increase of topics, rises rapidly after reaching the lowest point, and then fluctuates continuously, but there is no lower trough. Because of the general size of the text data set, and also want to carry out subsequent research in the case of a reasonable number of topics, so ignore the performance of the LDA model after the number of topics K is greater than 16. The topic confusion is shown in Figure 3. The figure shows that it reaches the lowest point when the number of topics is 5. In order to verify the reasonableness of the number of topics is 5, the model runtime score is introduced, and the model scores for different number of topics are shown in Fig. 4. When the number of topics reaches 4, the overall model score is the highest, and when the number of topics is 5, the model score can reach the second highest level, so the number of topics K is taken as 5.

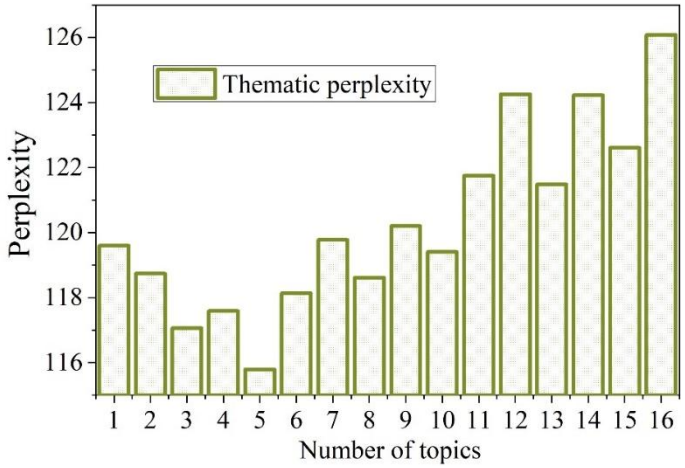


Figure 3: Thematic perplexity

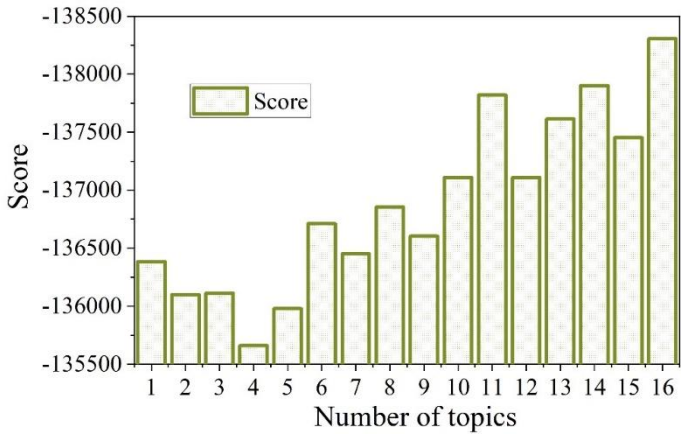
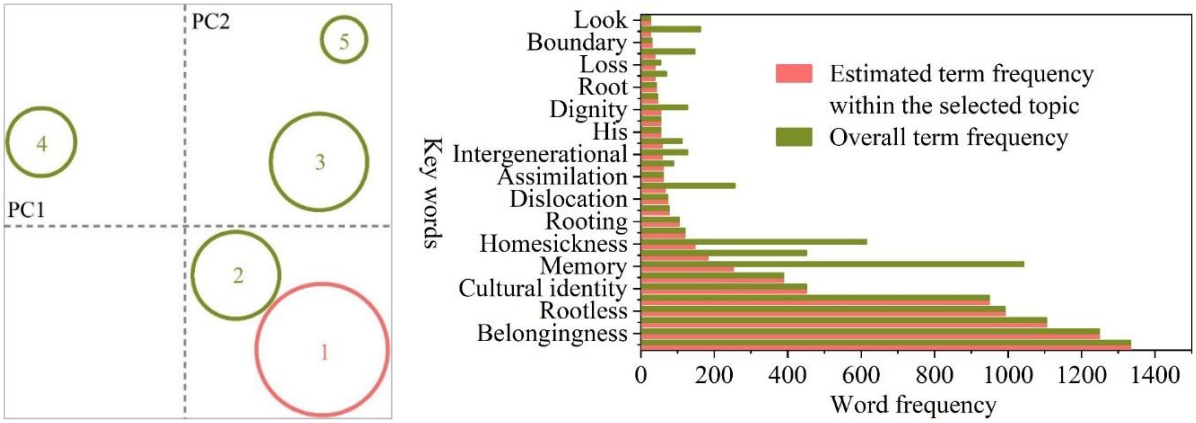


Figure 4: Model scores for different themes.

3.3.2 Visualization of online comment threads

pyLDAvis is a Pyhton visualization library that displays the distribution of clustered topics and the word frequency ordering of keywords within different topics. through the interface and interaction of the topic classification visualization, it can help to intuitively determine the reasonableness of the clustering grouping, and through the word frequency ordering it can sort out the similarities of the text under the topic to infer the meaning of the topic. pyLDAvis visualization is shown in Figure 5. Figure a shows the inter-topic distance, and Figure b shows the related words of topic 1. The size of the input parameter can be adjusted to change the relevance of the words to the theme, λ the closer to 1 the output will be more relevant to the

theme, and vice versa, it will show the higher uniqueness of the words under the theme. Since the purpose is clustering, we take λ=1 to get 5 themes, the larger the area of the theme circle means the larger the proportion of text under the theme, and the distance between each theme circle represents the difference, the farther away means the more different. Click on the circle with the number 1 at the center. Figure (b) shows the high-frequency keywords under this theme. As shown in the above figure, the main keywords within theme 1 include "identity recognition", "sense of belonging", "alienation", "rootedness", etc.



(a) The distance between themes

(b) Topic 1 related words

Figure 5: Pyldavis visualization

Different themes are defined according to the logical nomenclature, in order to explain the relationship of keywords within the group, theme naming is based on the logical relationship between the keywords with greater word frequency in the theme, this paper is also based on this method of naming the five themes after clustering, which are five themes of Identity and Belonging, Cultural Conflicts and Integration, Immigrant Experiences and Intergenerational, Social and Political Contexts, and Emotional and Psychological States, after the above After the verification and sorting above, the results fully show that most of the readers shared their comments around these five themes after reading the work. The contents of each theme and the categorized document examples are shown in Table 3. Key words that appeared in Topic 1 (Identity and Belonging) included “identity”, “cultural identity”, “sense of belonging”, “homelessness” and “self-deception”. All these descriptions express the readers' evaluation after reading the works, reflecting that the immigrant groups in the works have failed to win recognition for themselves from the mainstream society and within the family, and even lost their self-identity, whether they resisted with all their might or surrendered their weapons. Overall, from these five themes, we can conclude that the works of Kazuo Ishiguro and Zadie Smith reveal the active and narrative nature of memory, and at the same time deny the fixity and purity of identity, expressing the idea that identity is fluid, constructed, and is always in a state of “in progress”.

Table 3: LDA topic clustering results

Cluster topic number	Topic classification	Related feature words
Topic 1	Identity and attribution	Identity, cultural identity, sense of belonging, sense of homelessnessSelf-deception, memory, homesickness
Topic 2	Cultural conflict and fusion	Cultural shock, cultural collision, cross-cultural, hybridization, cultural integration, assimilation, multiculturalism
Topic 3	Immigration experience and generation	First-generation immigrants, second-generation immigrants, intergenerational conflicts, immigration journeys, trauma. Settlement, dispersion
Topic 4	Social and political context	Post-colonial, globalization, race, religion, tradition and modernity, social class
Topic 5	Emotional and psychological state	Alienation, dislocation, hope and despair, freedom of search, better life

3.4 Emotional image analysis

In this section, the trained SnowNLP sentiment analysis model is used to obtain the sentiment scores of the document set, to identify the emotional tendency of the reader's comments, and to introduce the IPA model to refine the analysis of the positive and negative factors affecting the reader's sentiment, and finally to complete the recognition of the emotional image perception of the works of Kazuo Ishiguro and Zadie Smith.

3.4.1 Identification of readers' emotional tendencies

The sentiment score of the document set is obtained by calling SnowNLP library from 0 to 1. The sentiment polarity of the document set is identified and the sentiment analysis is realized to reflect the reader's satisfaction with the five topics, and an example of the comment sentiment analysis is shown in Table 4.

Table 4: Comments on emotional analysis examples

Topic	Comment	Emotional score	Emotional polarity
Topic 1	Belonging becomes a task that needs to be continuously sought	0.8177	Positive
	The pursuit of belonging is always shaped and restricted by the grand history and political power	0.1556	Negative
Topic 2	A vibrant and mixed culture	0.8041	Positive
	The true belonging is difficult to reach	0.0376	Negative
Topic 3	Together, they have been pieced together to be their own	0.957	Positive
	There is a cultural gap that is difficult to cross within the family	0.1468	Negative
Topic 4	How is identity politics played in daily life	1.0461	Positive
	Often, the identity crisis of the individual is in the grand background of the decline of the empire	0.0628	Negative
Topic 5	Deep depiction of the internal trauma of immigrants	0.8237	Positive
	Accurate capture of outgoing immigrants	0.1426	Positive

The frequency of occurrence of each score band of the five theme emotions is shown in Fig. 6 to Fig. 10, respectively. The horizontal coordinate in the figure indicates the readers' emotion score, when the value is closer to 0, the stronger the negative emotion of the readers. The closer the value is to 1, the higher the positive sentiment. The vertical coordinate represents the number of readers' comments. The readers' emotions towards the works of Kazuo Ishiguro and Zadie Smith are shown as positive and negative emotions. Overall, 8865 positive comments were obtained at the end, and the degree of readers' positivity reached 92.04%, and readers' perception of Kazuo Ishiguro's and Zadie Smith's works was dominated by positive emotions. Measuring the mean values of the emotion scores of the five themes, it can be seen that the trend of the frequency of each score of emotion among different themes is more or less the same. That is, the mean values of identity and belonging, cultural conflict and integration, immigrant experience and intergenerational, social and political contexts, and emotional and psychological states are 0.856, 0.832, 0.828, 0.793, 0.765 respectively, reflecting the readers' emotional perceptions of each theme.

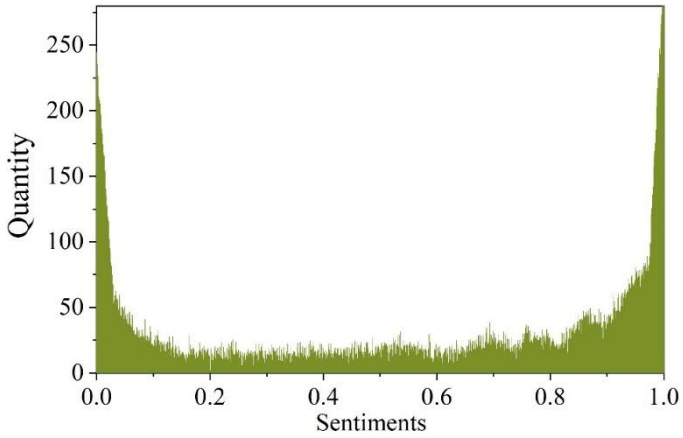


Figure 6: The frequency of the emotional score of the topic 1

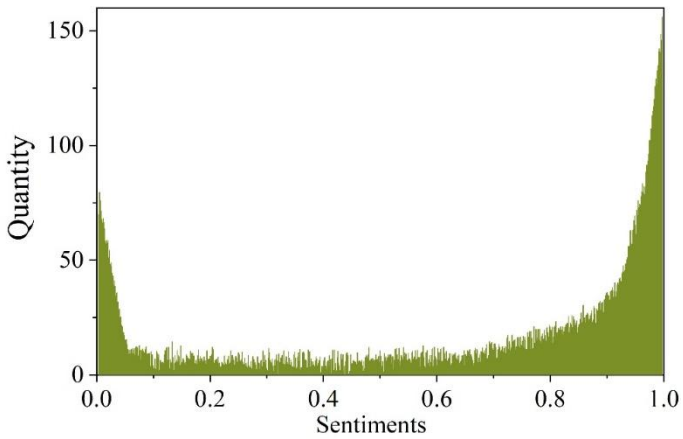


Figure 7: The frequency of the emotional score of the topic 2

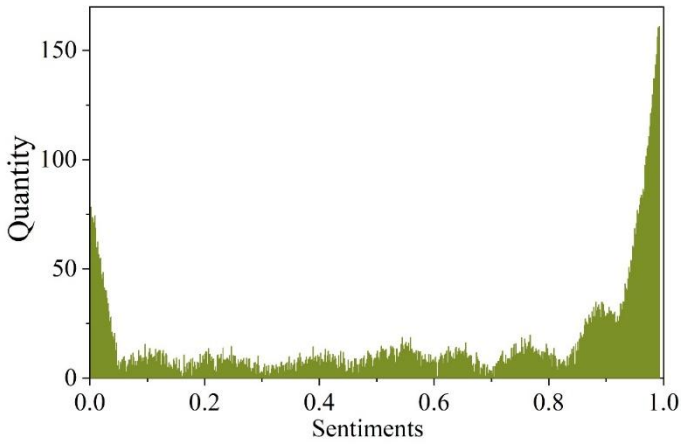


Figure 8: The frequency of the emotional score of the topic 3

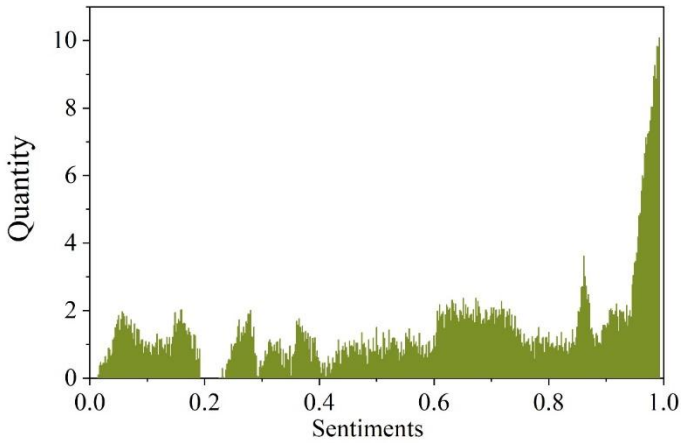


Figure 9: The frequency of the emotional score of the topic 4

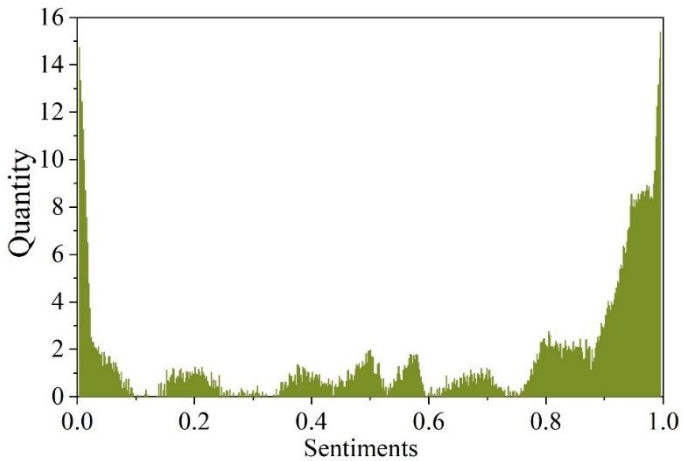


Figure 10: The frequency of the emotional score of the topic 5

3.4.2 IPA model analysis

After obtaining readers' affective tendencies through SnowNLP scores, in order to further explore the factors affecting readers' positive and negative emotions, we obtain more fine-grained analysis results of readers' perceptions, and improve the shortcomings of the current research that ignores the analysis of emotional images. In this paper, the text mining method is combined with the IPA modeling analysis method, and the importance and expressiveness are

used as the basis for evaluating the perceptual elements of Kazuo Ishiguro's and Zadie Smith's works. Five themes were identified in the previous paper. In this section, the IPA model is applied to number these elements 1-12. In the IPA model, based on the dimensional analysis and emotional scoring of the elements of image perception in the previous section, a two-dimensional coordinate system is established based on the values of importance and expressiveness of all the elements. In this paper, I_n is used as an indicator of importance, i.e., the ratio of the number of comments containing a certain n element to the total number of reader comments under a specific perception dimension. P_n is a perceptual indicator, which is the reader's evaluation of each element. That is, it is the ratio of the number of comments with positive affective tendency of a certain n element in the text to the total number of reader's comments containing that element. The structure of perceptual elements is shown in Table 5.

Table 5: Structure of perceptual elements

Serial number	Element	All frequencies	"Positive" frequency	I value	P value
1	Screening and reorganization of memory	564	523	0.1003	0.9273
2	The suppression of traumatic memory	2214	1835	0.3768	0.8288
3	The incompatibility of memory as a defense mechanism	939	884	0.1596	0.9414
4	Transmission and burden of generational memory	874	883	0.1542	1.0103
5	The performance of memory	2183	1873	0.3798	0.8580
6	Conflict of cultural memory	800	712	0.1359	0.8900
7	The material carrier of memory	1654	1176	0.2871	0.7110
8	The imaginary past	1521	1376	0.2613	0.9047
9	The dislocation of collective memory and personal memory	1069	805	0.1778	0.7530
10	The correction and disillusion of memory	617	495	0.1001	0.8023
11	As a memory of resistance	1200	938	0.1941	0.7817
12	Future memory	782	661	0.1388	0.8453

The scores obtained for each element were entered into SPSS software, with perceptibility as the horizontal coordinate, importance as the vertical coordinate, and mean value as the origin. Through the scatter plot and the related adjustment steps, it will be divided into four quadrants, namely, the performance hidden treasure area, the core advantage area, the secondary background area and the artistic highlight area, so as to get the IPA quadrant distribution map of the works of Kazuo Kuro and Zadie Smith, and the distribution of the IPA quadrants is shown in Fig. 11. As can be seen from the figure, the number of elements falling into the second four quadrants is the largest, indicating that the works of Kazuo Ishiguro and Zadie Smith together constitute a complete literary map about the construction of memory and identity.

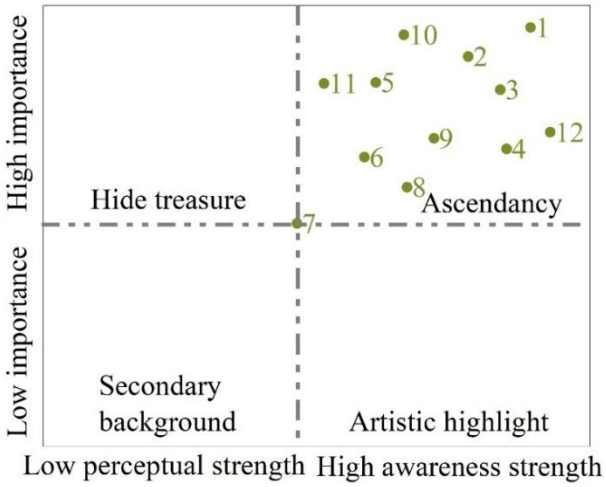


Figure 11: IPA quadrant distribution

4 Conclusion

In this paper, the reader's comments on a platform are collected to obtain a dataset of comments on the works of Kazuo Ishiguro and Zadie Smith. The TF-IDF algorithm is used for the representation of the text and the statistics of the word frequency of the readers' comments. Then the emotional image of the works of Kazuo Ishiguro and Zadie Smith is analyzed and studied by refining the emotional tendency of readers using SO-PMI and SnowNLP algorithms. The results of the study show that:

(1) In the co-occurrence matrix, “culture” and “memory” are co-occurring 3493 times. This indicates that these core feature words are the most likely to have an impact on readers' satisfaction and empathy. At the same time, these words collectively reflect the two writers' common focus on the immigrant experience, i.e., the complexity of the immigrant experience in terms of identity, culture and society.

(2) The works of Kazuo Ishiguro and Zadie Smith can be summarized into five themes, namely, identity and belonging, cultural conflict and integration, immigrant experience and intergenerational, social and political contexts, and emotional and psychological states. The two writers discuss deeply around immigration, homeland, emotion and culture, and their works reflect the pursuit and confusion of people in the immigrant experience.

(3) In terms of the emotional image of the works, readers gave roughly the same emotional score bands for the five themes, with the mean values of 0.856, 0.832, 0.828, 0.793, 0.765 respectively, reflecting most readers' recognition and empathy for the ideas conveyed by the works.

How to find one's own position in multiculturalism and how to solve the conflicts and contradictions between the “self” and the “other” are not only the problems that immigrant groups need to face, but also the problems that every individual living in multicultural environments needs to think about. In the globalized world, immigrant literature has demonstrated cultural richness and the tendency of writers to diversify their identities far beyond people's imagination. The lives and writings of Kazuo Ishiguro and Zadie Smith have enriched contemporary British literature, and their complex multicultural identities have, to a certain extent, dissolved the tendency of national identity in British literature, deconstructed the stable homogeneous character of British political identity, and fully demonstrated the complex implications of cultural identity and immigrant literature.

About the Author

Haoran Chen obtained a Master of Education degree from the School of Teacher Education at Nanjing Normal University. His research focuses on comparative literature, with a particular emphasis on postcolonial studies, diaspora and transnational studies.

References

[1] Samaddar, R. (2015). Returning to the histories of the late 19th and early 20th century immigration. Economic and Political Weekly, 49-55.

[2] Lauret, M. (2016). Americanization now and then: The “nation of immigrants” in the early twentieth and twenty-first centuries. Journal of American Studies, 50(2), 419-447.

[3] Gokieli, N. (2017). I want us to trade our skins and our experiences”: Swedish Whiteness and “Immigrant Literature. Scandinavian Studies, 89(2), 266-286.

[4] Balidemaj, A., & Small, M. (2019). The effects of ethnic identity and acculturation in mental health of immigrants: A literature review. International Journal of Social Psychiatry, 65(7-8), 643-655.

[5] Sairattanain, J., & Thawarom, T. (2022). English Children Literature for Exploring Immigrant Identity in a Language Classroom. International Journal of Language Education, 6(2), 101-112.

[6] Subhan, A., & Sulehria, S. (2025). The Language and Identity in Immigrant Literature. Pattern of Social Sciences Review, 1(2), 1-7.

[7] Lewis, B. (2024). Kazuo Ishiguro. In Kazuo Ishiguro. Manchester University Press.

[8] Dasgupta, R. (2016). Kazuo Ishiguro and ‘Imagining Japan'. In Kazuo Ishiguro in a Global Context (pp. 11-22). Routledge.

[9] Holmes, C., & Rich, K. M. (2021). On Rereading Kazuo Ishiguro. MFS modern fiction studies, 67(1), 1-19.

[10] Yi, C. C. (2016). A study of loss and memory in Kazuo Ishiguro’s never let me go, remains of the day, and he buried giant. Quint Interdiscip Q North, 9(1), 134.

[11] Hu, J. (2021). Typical Japanese: Kazuo Ishiguro and the Asian Anglophone Historical Novel. MFS Modern Fiction Studies, 67(1), 123-148.

[12] Pachuau, M. L., & Lalrinfeli, C. (2023). Situating Kazuo Ishiguro Within the Realms of Memory and Identity. Asiatic: IIUM Journal of English Language and Literature, 17(1).

[13] Bizzini, S. C. (2013). Recollecting Memories, Reconstructing Identities: Narrators as Storytellers in Kazuo Ishiguro's" When We Were Orphans" and" Never Let Me Go"/La recuperación de la memoria en la redefinición de la identidad: la narración como estrategia literaria en" When We Were Orphans y Never Let Me Go", de Kazuo Ishiguro. Atlantis, 65-80.

[14] Ray, K. S. (2017). Memory And Kazuo Ishiguro’s Novels: A Review. Literary Herald, 2(4), 292-309.

[15] Waham, J. J. (2023). The Exploration of Trauma and Memory in Kazuo Ishiguro's Never Let Me Go and The Remains of the Day. Journal of Critical Studies in Language and Literature, 4(3), 16-21.

[16] ISHIGURO'S, I. K., & VALANČIŪNAS, D. (2018). MEMORY, TRAUMA AND IDENTITY. History, Memory and Nostalgia in Literature and Culture, 213

[17] Mureșan, D. A. (2024). Memory, Self-deception and Denial in Kazuo Ishiguro’s The Remains of the Day. Perichoresis, 65.

[18] Lalrinfeli, C. (2014). Memory and Identity in Kazuo Ishiguro's An Artist of The Floating World. Labyrinth: An International Refereed Journal of Postmodern Studies, 5(1).

[19] Jamali, N., Motiee, M., & Ebrahimi, S. R. (2024). Memory, Identity, and Amnesia in Kazuo Ishiguro's The Buried Giant: A Cultural Memory Analysis through Jan and Aleida Assmann's Theories. Assessment and Practice in Educational Sciences, 2(3), 1-11.

[20] Dubovitskaya, M. A. (2024). Shifting Communicative Roles through Subcode Switching and Blending in Works of Zadie Smith. Nauchnyi Dialog, 13(5), 67-83.

[21] Pendharkar, A. (2014). Reading Zadie Smith: The First Decade and Beyond. Transnational Literature, 7(1), 1.

[22] Marcus, D. (2013). Post-hysterics: Zadie Smith and the fiction of austerity. Dissent, 60(2), 67-73.

[23] Fenno, C. (2014). Zadie Smith On Beauty, Youth, and Aging. Tulsa Studies in Women's Literature, 33(2), 179-202.

[24] Chalk, B. T. (2024). Paying Attention with Zadie Smith. In Novel Schooling: Education, Formation, and Reading in Fiction (pp. 167-202). Cham: Springer Nature Switzerland.

[25] Bağlama, S. H. (2020). Intersectionality in Zadie Smith's fiction: race, gender and class. Critical Studies in Social Sciences and Humanities, 21-38.

[26] Jansen, B. (2018). Accidental Englishness: Zadie Smith. In Narratives of community in the Black British short story (pp. 207-249). Cham: Springer International Publishing.

[27] Kowsalya, V., & Thenmozhi, J. (2024). Psychological Disposition in the Select Novels of Zadie Smith. Theory and Practice in Language Studies, 14(3), 781-785.

[28] Khan, A. W. (2020). A linguistic and sociolinguistic appraisal of the novel White Teeth by Zadie Smith. Rupkatha Journal on Interdisciplinary Studies in Humanities, 12(5), 1-13.

[29] Mahmoudi, M. R., & Abbasalizadeh, A. (2019). How statistics and text mining can be applied to literary studies?. Digital Scholarship in the Humanities, 34(3), 536-541.

[30] Scrivner, O., & Davis, J. (2017, January). Interactive Text Mining Suite: Data

Visualization for Literary Studies. In CDH@ TLT (pp. 29-38).

[31] Muralidharan, A., & Hearst, M. A. (2012). Supporting exploratory text analysis in literature study. Literary and linguistic computing, 28(2), 283-295.

[32] Omar, A. (2021). Towards a Computational Model to Thematic Typology of Literary Texts: A Concept Mining Approach. International Journal of Advanced Computer Science and Applications, 12(12).