







Application of Automatic Scoring of English Assessment Improved by Computer-Aided Design

Xinyu Zhang¹ , Ran Cui² , Hui Li³ , and Wanyue Zhang⁴ 

¹Cangzhou Normal University, Cangzhou, Hebei 061000, China, flw@caztc.edu.cn

²Cangzhou Normal University, Cangzhou, Hebei 061000, China, Zhang000xiny@163.com

³Cangzhou Normal University, Cangzhou, Hebei 061000, China, lihuicangzhou@163.com

⁴Foreign Language Department Qinghai Normal University, 810000, China, zhwyqinghai@126.com

Corresponding author: Zhang000xiny@163.com

Abstract. Because the automatic scoring of English assessment is limited by English grammar and semantics, it is not intelligent enough. In order to improve the automatic scoring effect of English assessment, based on computer-aided technology, this study used latent semantic analysis to extract features from English content and identify its specific meaning. Moreover, this paper used the syntax analyzer StanfordParser to analyze the diversity of the sentence structure. In addition, this paper projected the document vector and the word vector into a low-dimensional space by singular value decomposition, so that the compositions that are related to each other can obtain an approximate vector representation even if the same word is not used, and the relevance of the composition context is obtained. Research shows that the method proposed in this paper has certain effects.

Keywords: computer aided; English; assessment; automatic scoring.

DOI: <https://doi.org/10.14733/cadaps.2020.S1.57-67>

1 INTRODUCTION

At present, the development of oral English test has been universally based on computer network technology, such as the authoritative foreign new TOEFL English machine test and the domestic college English four or six oral machine test. In the study of computer-aided scoring monitoring for English exams, in China, the scoring curve is mainly generated by comparing the scores of multiple scorers, including the exit rate and the average score, and controlling the scoring speed of the scorers [1].

The CELST system is a semi-direct English oral examination system. It simulates the traditional examination organization and management mode, adopts the AS framework, and records the content of the candidate's oral expression through computer multimedia equipment in a human-computer interaction manner to generate the candidate's recording answer sheet. After the test, the candidate's recording answer sheet will be uploaded to the server and played back in the scoring stage. After the rating is over, users with different permissions can refer to the corresponding

information [2]. At present, there are two kinds of scoring modes supported by different evaluation types: computer-aided scoring and computer automatic scoring. In the scoring process, the total scores of students are counted according to the set scoring reference standards. The scorer scores the candidate's recorded answer sheet through the client in the campus network, and the system can support simultaneous online scoring of no less than 20 people [3].

Computer-aided scoring means that the scorer combines the rich experience accumulated by the artificial assessment with modern high-tech according to the established scoring standards, scoring rules, and spoken score scales. Its main purpose is to control the scoring error, achieve the fairness of marking and the fairness of the test. The method refers to the description of the performance of the oral tester or the grade of the speech sample on the computer-based test platform, and the subjective judgment and assignment process of the oral level of the subject [4].

Scoring process of computer-assisted scoring process: First, the scorer logs in to the rating interface, and it can be seen that it consists of three parts: left, center, and right. On the left is the name of the exam, and the menu bar is the name of the rater and the progress of the score, which helps the scorer to check the basic exam information before the score. After that, a candidate's recording answer sheet was selected. It can be seen that the middle part of the rating interface displays the test paper content of the candidate and the spoken answer recording content, and the scorer can obtain complete information, such as the text and picture of the test paper, through the scroll bar and the play button. When playing the candidate's recording answer sheet, the scorer can operate the play progress scroll bar to adjust the playback progress or repeatedly listen to a certain recording. The right side of the rating interface is the rating criteria, rating area, and submission grade. The six score levels from 0 to 5 of the score sub-region in turn represent the scores from low to high. After selecting the score results and submitting the scores, the score is completed [5].

The oral test system can have a positive impact on the scoring results. For example, it can save the candidate's answer recording data and change the irreversibility and repeatability of the scoring process in the traditional spoken language score. Computer-aided scoring has been widely applied in various computer-based test systems. However, certain human resources still need to be invested in the scoring process. On the basis of reducing the demand for scoring work, some computer-based test systems can complete the automatic scoring of pure text questions, and the score results significantly exceed the performance of teacher scores in face-to-face oral test scores [6]. In terms of automatic Chinese scoring, Schillingmann L and others have conducted in-depth research on the reading of the Putonghua proficiency test. The research shows that the scoring results have exceeded the scores of experienced professional scorers, and it has been put into the market. Zhang Jie did a study on the reading and closing questions of the National Chinese Proficiency Test. The results show that computer automatic scoring can replace manual scoring to a certain extent, and can improve the efficiency of oral test scores, and the scores of reading aloud are slightly better than closed questions [7]. The system studied in this paper has achieved automatic scoring of reading questions, listening questions and retellings, and broke the dependence of traditional automated scoring on text. The computerized automatic scoring of spoken language refers to the integration of the intelligent speech scoring system into the computer-based test system, and based on the information characteristics of accurate manual scoring spoken words, the candidates' recording data of the candidates are evaluated and scored, and the candidate's spoken recordings to be scored are classified to greatly improve the efficiency and objectivity of the colloquial scoring. After the end of the test, combined with speech recognition, natural language processing and other techniques, according to the text coverage, speech rate, grammar evaluation, keyword coverage, etc., which need to be examined in the scoring standard, the computer can automatically complete the scoring of the candidate's voice data [8].

2 RESEARCH METHOD

2.1 Selection of eigenvalues

In the Term-Document matrix, each element represents the importance of a word in the corresponding document, and the vocabulary that better reflects the subject of the composition will be given a greater weight. The weight calculation formula directly affects the ability of the selected feature words to represent the composition content. The most common method currently used is $TF \times IDF$, which was proposed by Salton of Cornell University in 1988. A standard word frequency weight f_i^k is multiplied by a factor inversely proportional to the word document frequency d_k to obtain a product. If d_k is the document frequency of the word item k, the inverse document frequency IDF_k of the word item can be expressed as [9]:

$$IDF_k = \lceil \log_2 n \rceil - \lceil \log_2 d_k \rceil \quad (1)$$

The weight of a word item will be proportional to $f_i^k \times IDF_k$, that is, those words that appear more frequently in a single text and appear less frequently in all training set texts are given a higher weight.

Mutual information measures the statistically independent relationship between a word and a category. When considering a word t and a category c, the mutual information is defined as equation (2) [10]:

$$I(t,c) = \log(p(t \wedge c)) - \log(p(t) \times p(c)) \quad (2)$$

Among them, $p(t \wedge c)$ is the probability that the word t and the category c appear simultaneously, $p(t)$ is the probability of occurrence of the word t, and $p(c)$ is the probability of occurrence of the category c. If a word and a category are statistically independent of each other, and $p(t \wedge c) = p(t) \times p(c)$, $I(t,c)$ is 0. That is to say, the occurrence of the word t has no amount of information for predicting the category c. In actual calculations, these probabilities can be approximated by the corresponding frequency of occurrence in the training set. The frequency at which both t and c occur simultaneously in the training set is defined as A. N is the number of texts in the training set, B is the frequency of texts that t appears in the training set, and C is the frequency of the categories in which c appears in the training set. Then the mutual information g can be approximated by the formula (3) [11]:

$$I(t,c) = (A/N) / [(B/N) / (C/N)] = (A/N) / (B/C) \quad (3)$$

The basic idea of the chi-square test is to determine the correctness of the theory by observing the deviation between the actual value and the theoretical value. When the chi-square test is taken, it is often assumed that the two variables are indeed independent, and then observe the degree of deviation between the actual value and the theoretical value. If the deviation is small enough, the error is considered to be a natural sample error, which is caused by the inaccuracy of the measurement means or accidentally, and it is considered that the two are indeed independent, and the null hypothesis is accepted at this time. If the deviation is so large that such deviation is unlikely to be accidental or measurement inaccurate, it can be considered that the two are actually related, that is, the original hypothesis is rejected, and the alternative hypothesis is accepted. So, what is the measure of the degree of bias? We assume that the theoretical value is E (mathematical expectation) and the actual value is x. If only the sum of the difference $x - E$ between the observed values and the theoretical values of all samples is used, it is likely that when there are multiple

observations x_1, x_2, x_3 , the values of $x_1 - E, x_2 - E, x_3 - E$ are positive and negative and cancel each other, so that the final deviation is zero. But in fact, there is a deviation between the observed value and the theoretical value, and the deviation is not small. The straightforward idea at this point is to use the variance instead of the mean, thus solving the problem of positive and negative offset. That is, formula (4) is used [11]:

$$\sum_{i=1}^n (x_i - E)^2 \tag{4}$$

This brings up a new problem: for a mean of 500, the difference of 5 is actually very small, and for a mean of 20, there is a difference of 25%. Therefore, the above formula is improved, so that the size of the mean does not affect the judgment of the degree of difference. The improved formula is shown in equation (5) [12]:

$$\sum_{i=1}^n \left((x_i - E)^2 / E \right) \tag{5}$$

If there are several samples of observation x_1, x_2, \dots, x_n , substituting them into equation (5) will give the chi-square value. After that, this value is compared with a preset threshold. If the value is greater than the threshold, the null hypothesis is considered not to be true, whereas the null hypothesis is established [13].

In the feature selection phase of the text classification problem, the main concern is whether a word t (a random variable) and a category c (another random variable) are independent of each other. If they are independent of each other, it can be said that the word t has no characterization effect on the category c , that is, it is impossible to judge whether a document belongs to the category c or not according to whether t is present or not. Unlike the normal chi-square test, it does not need to set a threshold because it is difficult to say to what extent the word t and the category c are related. Therefore, just borrow this method to select some of the most relevant ones. The term "word t is not related to category c " is generally used as the null hypothesis. The selection process also becomes the calculation of the chi-square value of each word with category c , and the chi-square values are sorted from large to small, and the first k words are taken [14].

For example, there are now N documents, and we want to examine the correlation between a word t and category c : the relationship between them is shown in Table 1.

<i>Feature selection</i>	<i>Belongs to category c</i>	<i>Does not belong to category c</i>	<i>Total</i>
Contains the word t	A	B	A+B
Don't contain the word t	C	D	C+D
Total	A+C	B+D	N

Table 1: Relationship between word t and category c .

Taking the number of documents containing t and belonging to category c as an example, if the null hypothesis is true, that is, the text of t and category c has no relevance, in all the texts, the word t should appear with equal probability, regardless of whether the text is of class c or not. This probability is shown in equation (6). $A + B$ is the number of texts containing the word t . It is divided by the total number of documents to get the probability of the occurrence of the word t [15]:

$$(A + B) / E \tag{6}$$

The number of texts belonging to category c is $A + C$. In these documents, there should be E_{11} documents containing the word t . Among them, E_{11} is as shown in equation (7) [16]:

$$E_{11} = (A + C) \times (A + B) \quad (7)$$

However, in fact, only the text of the A piece contains the word t in the text of the category c, so the difference value of this case is as shown in the formula (8) [17]:

$$D_{11} = (A - E_{11})^2 \quad (8)$$

Similarly, we can calculate the difference between the remaining three cases D_{12}, D_{21}, D_{22} . Among them, $D_{12} = (A - E_{12})^2 / E_{12}$ (including the word t but not belonging to the category c), $D_{21} = (A - E_{21})^2 / E_{21}$ (does not contain the word t but belongs to category c), $D_{22} = (A - E_{22})^2 / E_{22}$ (does not contain the word t and does not belong to category c). Therefore, the chi-square value of the word t and the category c is as shown in the formula (9):

$$\chi^2(t, c) = D_{11} + D_{12} + D_{21} + D_{22} \quad (9)$$

By simplifying $D_{11} + D_{12} + D_{21} + D_{22}$, equation (10) can be obtained:

$$\chi^2(t, c) = N \times (AD - BC)^2 / (A + C)(A + B)(B + D)(C + D) \quad (10)$$

According to the study of plain text in English, the currently accepted best feature selection method is the chi-square test. The performance of word frequency inverse document frequency method is roughly the same as that of the former, and the performance of mutual information method is the worst. Therefore, the chi-square test is used to select feature words.

2.2 Weight function

After using the feature vector selection algorithm to select some feature words in the composition, the next step is to generate the feature vectors of each composition in the n piece of compositions. Each feature of the feature vector has a weight to measure the contribution of the feature to the composition content score. The weight of the i-th feature of the j-th composition is expressed by a_{ij} . This weight function consists of two parts, one is the local weight function $L(i, j)$, and the other is the global weight function $G(i)$. The local weight function $L(i, j)$ represents the weight of the feature word i in the document j. The global function $G(i)$ represents the weight of feature i in the entire composition set.

The local weight function selects the logarithmic form of the document frequency of the feature word i, as shown in equation (11):

$$L(i, j) = \log(tf(i, j) + 1) \quad (11)$$

Among them, $tf(i, j)$ is the number of occurrences of the feature word i in the jth essay.

The global weight function is determined by the information entropy (as shown in equation (12)) determined by the distribution of the feature in all of the composition sets.

$$G(i) = H(d|i) = -\sum_j p(i, j) \log p(i, j) \quad (12)$$

$H(d|i)$ represents the conditional entropy given the occurrence of a given feature word i. $p(i, j)$ is the probability that the feature term i appears in the j-th essay. Equation (13) is used to indicate the importance of the feature word i in the composition.

$$W(i, j) = L(i, j) / G(i) = \log(tf(i, j) + 1) / \left(-\sum_{j=1}^n p(i, j) \log p(i, j) \right) \quad (13)$$

2.3 Vector space model

In general, a document vector model can be used to represent a piece of text, and each document is represented by a vector. In text space, document D_i is represented by one or more features T_j . These items can assign weights between 0–1 to individual features based on their importance. A typical three-dimensional vector space is as shown in Figure 1: each vector is marked by three different feature items. Moreover, when the feature items of the t number appear, the three-dimensional example can be extended to t-dimension. In this case, for each document D_i can be represented by the t-dimensional vector $D_i = (d_{i1}, d_{i2}, \dots, d_{it})$. Among them, d_{ij} is the number of times the j-th feature item appears in document i.

After the feature vectors of the two documents are given, the cosine similarity $s(D_i, D_j)$ between them can be calculated, which reflects the similarity between the corresponding items and the items. All vector lengths are normalized to 1, and then the projection point of the vector in the unit sphere is considered. The similarity between the two documents can be calculated by the distance between the two projection points, that is, the angle cosine after the normalization of the two vectors. The larger the cosine value, the smaller the angle, and the more similar the two documents are. Conversely, the smaller the cosine value, the larger the angle, and the greater the difference between the two documents.

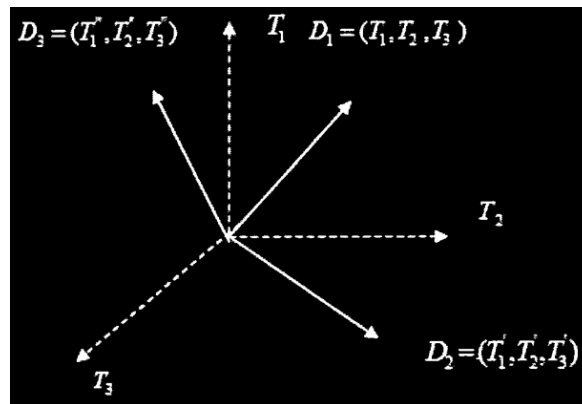


Figure 1: Three-dimensional document vector.

The advantage of the vector space model is that the unstructured text is represented as a vector form, which makes various subsequent mathematical processing possible. However, the basic assumptions (orthogonal assumptions) in the vector space model that the inter-word relations are independent of each other are difficult to satisfy in the actual environment. Moreover, the words appearing in the text tend to have a certain correlation, so the space vector model will affect the calculation results to some extent.

In keyword-based text processing, processing is mainly based on the frequency information of the vocabulary. For example, the degree of similarity between two texts depends on the number of common vocabularies they have, which may result in so-called synonyms and polysemous

phenomena. For the same thing, different people will have different expressions according to individual needs, environment, knowledge level and language habits, so the vocabulary used is also very different.

3 SYSTEM CONSTRUCTION

The content scoring module first converts the text of the composition into a word stream, and its class diagram is designed as shown in figure 2. Among them, Document Loader represents the mechanism that the text is loaded, loading refers to opening a byte stream of a specified document source, and Loader refers to the corresponding implementation class. InputFilter represents a stream of characters converted to a character stream according to a given byte stream. TextInputFilter, PDFInputFilter respectively correspond to an implementation class that converts a byte stream of a text file and a PDF file into a character stream. The Tokenizer interface converts a stream of characters into a stream of words, removing all separators. TokenEnumeration is an enumeration interface for word streams. SimpleTokenizer implements the TokenEnumeration and Tokenizer interfaces and implements the conversion of character streams into word streams.

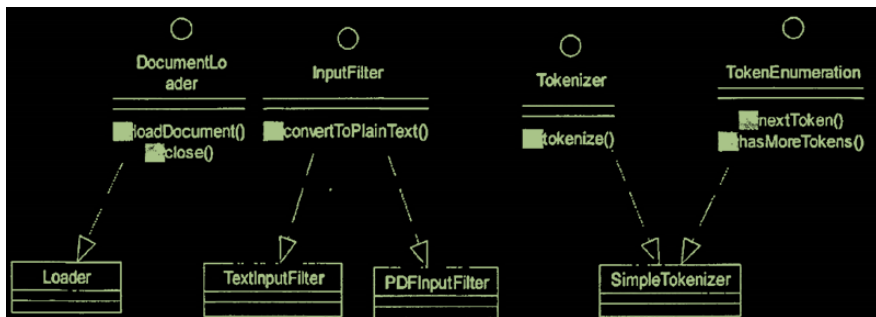


Figure 2: Class diagram design.

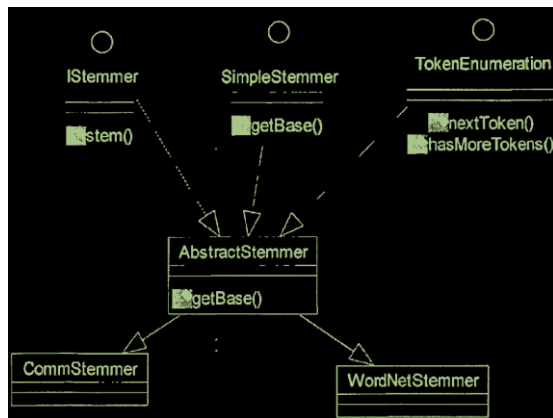


Figure 3: Class diagram of Selecting the root module.

Figure 3 is the class diagram of Selecting the root module CommStemmer is a general method of taking roots, while WordNetStemmer uses the WordNet semantic dictionary to query the concept of the form. Among them, the stem method is used to convert the word stream into a corresponding root stream. The method of getting the root of getBase in SimpleStemmer is different depending on whether the method of taking the root is based on WordNet.

In the oral test section, even if the same spoken English, different readers will get different pronunciation effects. To this end, the non-uniform linear predictive cepstral coefficient NLPC is extracted from the reader's spoken English signal, which is robust and independent of the speaker. Therefore, the reader's spoken signal is preprocessed, and then the spoken language pronunciation is divided into vowels, consonants and inactive frames using endpoint detection and Viterbi methods. Later, Bark bilinear transformation is used to spectrally bend the linear spectrum, so that the bent spectrum has non-uniform characteristics of auditory perception. Then, the NLPC is calculated by linear prediction of the non-uniform spectrum, and the reference NLPC is extracted to perform the GMM model training to obtain the training model of the standard spoken pronunciation feature parameters. Then, the GMM training is used to model each type of characteristic parameter distribution of spoken English pronunciation. Similarly, after extracting the NLPC feature parameters of the reader's spoken language, the parameters are input into the reference model, and the consistency measure with the standard spoken language GMM model is calculated, and the English speaking is scored according to the consistency measure. Then, the objective score of English is calculated according to the calculated consistency measure by the MARS algorithm, and the estimated value of the spoken English score is obtained by using the mapping relationship. The algorithm principle is shown in Figure 4.

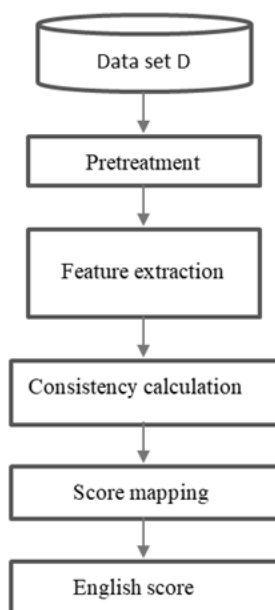


Figure 4: Principle of English scoring.

It can be seen from Fig. 4 that the NLPC of the standard spoken English signal is first extracted to perform the GMM model training, and a reference model of the standard spoken English signal is obtained. The NLPC vector of the reader's spoken language is calculated and the consistency measure between them is obtained. Then, the fractional mapping method is established by using the multivariate adaptive regression spline function (MARS) algorithm. Finally, the system is tested. The confirmed software is incorporated into the actual operating environment and combined with other system components for testing, as shown in Figure 5.

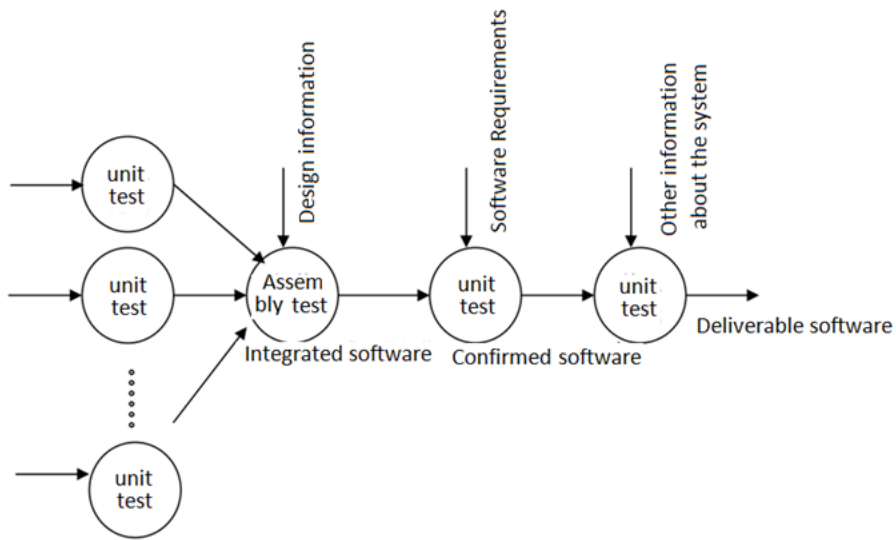


Figure 5: System software testing process.

The system operation interface is shown in Figure 6.

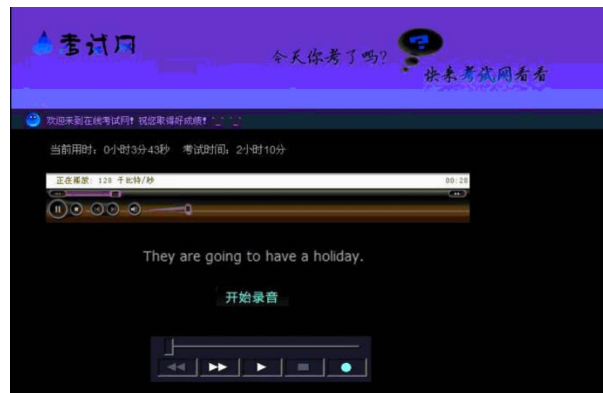


Figure 6: System operation interface.

4 ANALYSIS AND DISCUSSION

After the pre-processing of the composition is completed, the obtained word stream still contains a large number of different words, and only by extracting words that can reflect the compositional feature information can the dimension of the spatial matrix be reduced. The idea of feature extraction is to construct different evaluation functions according to different feature extraction methods, and to independently evaluate each feature in the initial vector to obtain an evaluation score. Then, all the features are sorted according to the size of the evaluation score, and a specified number of feature subsets are selected.

The WordNet semantic dictionary is used to take the root class, and WordNet Stemmer uses JwNL. JwNL is an API for accessing the WordNet semantic dictionary. It not only provides data access

functions, but also provides word-recognition discovery and language processing. In the semantic dictionary (Dictionary), a plurality of index word sets (Index Word Set) corresponding to one-word form can be found, and each element in the set corresponds to an index word (Index Word) under the part of speech, and each index word corresponds to a line in the index file (pos. index). An index word can correspond to multiple concepts, each of which is a synonym set (Synset). Each Synset represents a line in the data file (pos.data) in WordNet, which represents a concept that contains a collection of words and the words are synonymous. Synonyms point to a synonym (Target) associated with it through a pointer, which constitutes a network of synonyms. Moreover, depending on the type of relationship (synonymous relationship, antisense relationship, upper and lower relationship, partial overall relationship), Pointer has different types. Therefore, it is associated with a different synonym set (synset). If we want to know what index word a word corresponds to, we can use the part of the word in the context to determine. The first concept of the index word is selected, and the same concept can correspond to multiple words (Word), and the first word is used to represent the concept. We take the word body as an example. When its part of speech is a noun, there are 9 concepts, that is, 9 synonym sets. Its first synonym set contains two words (Word), which are organic structure, physical structure, and its upper word is natural object.

The test process is carried out in 4 steps, namely unit test, assembly test, confirm test and system test. The beginning is a unit test, which focuses on testing each program unit of the source code implementation to check whether each program module correctly implements the specified functions. Then, the tested modules are assembled and assembled for testing, which is mainly to test the construction of the software architecture related to the design. For this reason, in the process of assembling a unit test and ensuring that the correct program modules are assembled into a software system, the correctness and program structure are checked. The validation test checks whether the software that has been implemented meets the various requirements identified in the requirements specification and whether the software configuration is completely correct.

5 CONCLUSION

This article scores the composition from both content and language. The content aspect uses the LSA to extract the feature items of the composition, and language uses the existing natural language processing package Stanf-ordParser to analyze the diversity of the syntactic structure. When Stanf-ordParser extracts the syntactic tree of a sentence, it also identifies the component of the word or phrase in the sentence. Comparing the occurrence of words in a sentence, it pays more attention to the semantic composition within the context. Such an analytical method is more in line with human reading and understanding. The core idea is to project the document vector and the word vector into a low-dimensional space by singular value decomposition, so that the essays associated with each other can obtain an approximate vector representation even if the same word is not used, and the association of the composition context is obtained. At present, in China, the scores for English essays have been developed, such as some essay correction websites have appeared. However, its application is not mature, or it is still in its infancy. Therefore, there are still many works to be done in the follow-up.

6 ORCID

Xinyu Zhang, <https://orcid.org/0000-0003-1775-5218>

Ran Cui, <https://orcid.org/0000-0001-7719-8148>

Hui Li, <https://orcid.org/0000-0002-7514-6712>

Wanyue Zhang, <https://orcid.org/0000-0003-1975-2280>

REFERENCES

- [1] Johnson, D. O.; Kang, O.; Ghanem, R.: Improved automatic English proficiency rating of unconstrained speech with multiple corpora, *International Journal of Speech Technology*, 19(4), 2016, 755-768, <https://doi.org/10.1007/s10772-016-9366-0>
- [2] Lilley, J.; Spinu, L.: Automatic classification of English fricatives using cepstral coefficients, *Journal of the Acoustical Society of America*, 139(4), 2016, 2016-2016, <https://doi.org/10.1121/1.4949939>
- [3] Ewald, V. D. W.; Niesler, T.: Automatic speech recognition of English-isiZulu code-switched speech from South African soap operas, *Procedia Computer Science*, 81, 2016, 121-127, <https://doi.org/10.1016/j.procs.2016.04.039>
- [4] Lyashevskaya, O.; Panteleeva, I.: Automatic dependency parsing of a learner English corpus Realec, HSE Working papers, 2017, <https://doi.org/10.2139/ssrn.3089660>
- [5] Liu, M.; et al: Automated scoring of Chinese engineering students' English essays, *International Journal of Distance Education Technologies*, 15(1), 2017, 52-68, <https://doi.org/10.4018/IJDET.2017010104>
- [6] Wang, C.; et al: Letter-sound integration in native Chinese speakers learning English: Brain fails in automatic responses but succeeds with more attention, *Cognitive Neuroscience*, 2018, <https://doi.org/10.1080/17588928.2018.1529665>
- [7] Schillingmann, L.; et al: AlignTool: The automatic temporal alignment of spoken utterances in German, Dutch, and British English for psycholinguistic purposes, *Behavior Research Methods*, 2018, <https://doi.org/10.3758/s13428-017-1002-7>
- [8] Ahn, T. Y.; Lee, S. M.: User experience of a mobile speaking application with automatic speech recognition for EFL learning, *British Journal of Educational Technology*, 47(4), 2016, 778-786, <https://doi.org/10.1111/bjet.12354>
- [9] Behravan, H.; et al: i-Vector modeling of speech attributes for automatic foreign accent recognition, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(1), 2016, 29-41, <https://doi.org/10.1109/taslp.2015.2489558>
- [10] Chua, C. C.; et al: Analogical-based translation hypothesis derivation with structural semantics for English to Malay example-based machine translation, *Advanced Science Letters*, 24(2), 2018, 1263-1267, <https://doi.org/info:doi/10.1166/asl.2018.10729>
- [11] Kaity, M.; Balakrishnan, V.: An automatic non-English sentiment lexicon builder using unannotated corpus, *The Journal of Supercomputing*, 75(4), 2019, 2243-2268.
- [12] Johnson, D. O.; Kang, O.: Comparison of algorithms to divide noisy phone sequences into syllables for automatic unconstrained English-speaking proficiency scoring, *Artificial Intelligence Review*, 2017, <https://doi.org/10.1007/s10462-017-9594-y>
- [13] Susanti, Y.; et al: Evaluation of automatically generated English vocabulary questions, *Research and Practice in Technology Enhanced Learning*, 12(1), 2017, <https://doi.org/10.1186/s41039-017-0051-y>
- [14] Li, X.; Liu, J.: Automatic essay scoring based on Coh-Metrix feature selection for Chinese English learners, 2016, https://doi.org/10.1007/978-3-319-52836-6_40
- [15] Lee, G. G.; et al: Automatic sentence stress feedback for non-native English learners, *Computer Speech & Language*, 41, 2017, 29-42, <https://doi.org/10.1016/j.csl.2016.04.003>
- [16] Sidgi, L. F. S.; Shaari, A. J.: The effect of automatic speech recognition EyeSpeak software on Iraqi students' English pronunciation: A pilot study, *Advances in Language and Literary Studies*, 8(2), 2017, 48, <https://doi.org/10.7575/aialc.all.v.8n.2p.48>
- [17] Zhao, D.; Sun, J.: Research on the automatic error correction model combined with artificial intelligence for college english essays, 2016, https://doi.org/10.1007/978-3-319-60744-3_5