Journal of Information Hiding and Multimedia Signal Processing ©2017 ISSN 2073-4212
Ubiquitous International Volume **8**, Number **6**, November **2017**

A Two-way Parallel Query Correction Approach Based on Semantic Analysis and Reverse Hidden Markov Model in Chinese Information Processing System

Quan-You Yu[1], Jing-Rui Pei[2*]

[1]Liberal College
Shenyang Normal University
Shenyang 110034, China
[*]Corresponding author:Jing-Rui Pei,360321978@qq.com

[2*]Software College
Shenyang Normal University
Shenyang 110034, China

Received March, 2017; revised July, 2017

Abstract. *Traditional query correction methods cannot better solve the question of mixed language and Chinese long query in Chinese retrieval system, which can bring trouble for human-machine multimodal interaction. Query correction approach is not a simple signal processing problem. To solve the problem, this paper proposes a two-way parallel query correction approach based on semantic analysis and reverse hidden Markov model for mixed language in computer language. This new approach is divided into four steps. First, this method conducts uniform coding for mixed language. Additionally, it works out corresponding editing rule based on semantic analysis to manage the query words uniformly. Then, we put forward two-way parallel query correction model for Chinese long query. At last, we propose reverse hidden markov model to handle parallel query sentences. Finally, experiments are conducted to demonstrate this new approach. Experiment training language database consists of high quality texts extracted from user query log, web click log and web link information etc. Results show that our new method has high accuracy rate. And the Recall rate greatly decreases with a low convergence time. What's more, the query speed is faster than other methods.*

Keywords: Semantic analysis, Correction processing, Two-way parallel query correction, Mixed language, Reverse hidden markov model

1. Introduction.

Query correction[1,2] is a hot topic, used for correcting the spelling error of query sentence in information retrieval system. It can be divided into two categories: non-word errors and real-word errors detection in English query. In Chinese information retrieval system, it contains three types: Chinese, English and Chinese phonetic alphabet. In Chinese query, it may appear homophone error, similar Chinese Character error, lacking of the Chinese characters, loss of alphabet in phonetic alphabet etc. Therefore, many information retrieval systems conduct correction processing for query sentences to satisfy the requirement of users and improve the retrieval efficiency.

Automatic completion method[3] is based on the simple edit distance model, which can be used to find the completion of the database. However, when the data volume increases, there is a bottleneck between edit distance and linear search time, which leads to poor robustness and extensibility of the algorithm, which cannot be applied to the actual online error correction query completion system[4].

Semantic analysis[5-7] is a logical stage of the compilation process. The task of semantic analysis is to review the nature of the contextually correct source program and perform a type review. Semantic analysis is to examine whether the source code has semantic errors for the code generation phase collecting type information. For example, a job of semantic analysis is a type review, which examines whether each operator has an object that is allowed by the language specification. When the language specification is not met, the compiler should report an error. If the compiler to the real number as an array subscript report error. For example, some of the procedures specified by the operation object can be forced, then when the two operations are applied to a single type and a real object, the compiler should be converted to real type and can not be considered as a source error. In a social network, there is always communication between the nodes. A powerful technique for analyzing and understanding textual information that can be used to analyze such social networks called semantic web messaging analysis (semantic analysis). As a method of artificial intelligence and computational linguistics, it provides a structure and process for knowledge reasoning and language. Colin [8] presented two different possible solutions. In one, he indexed an unannotated version of the PIKES collection using Latent Semantic Analysis (LSA) retrieving relevant documents using a combination of query coordination and automatic relevance feedback. Although he outperformed prior work, this approach was dependent on the underlying collection, and was not necessarily scalable. In the second approach, he used an LSA Model generated by SEMILAR from a Wikipedia dump to generate a Term Similarity Matrix (TSM). Queries were automatically expanded with related terms from the TSM and were submitted to a term-by-document matrix Vector Space Model of the PIKES collection. Yi [9] proposed a novel method to discover the popular subtopics for a given query. The new method first constructed a search behavior tripartite graph based on the search log data. Then, it utilized a subtractive initialized Non-negative Sparse LSA model to mine subtopics from the tripartite graph.

This paper proposes a two-way parallel query correction approach based on semantic analysis and reverse hidden Markov model for mixed language and long sentences problem in Chinese information processing system. Firstly, this paper puts forward a pseudocode method to solve the problem of mixed language correction. This method uses English words and common characters assembly as independent words to code, which can manage unified coding for Chinese and English. Secondly, it presents a parallel error correction model based on semantic analysis to rapidly deal with Chinese long sentences. This model divides query sentence exceeding threshold value into two equal sentences. Then it executes reverse hidden Markov model to process query sentences using special editing rule at the same time. It can generate several candidate statuses in the process of handling sentences. And each candidate status can keep candidate item and operation cost. Finally, it will provide the query sentence according to operation cost.

This new method adopts characters dictionary tree and language model to judge the word itself information and word in the context information. Training language database consists of high quality texts extracted from user query log, web click log and web link information etc. To process reverse query sentence, we propose reverse character dictionary tree and reverse language model based on reverse hidden Markov model respectively. According to the character of English and Chinese, it works out relative processing rule. In this paper, the new error correction experimental results show that it has large potential of rectifying the wrong and better correction effect. This paper is organized as follows. Section 2 is the related work. Section3 detailed introduces the new method. Followed by section4, it gives the experiments for new method. There is a conclusion in the last section.

2. **Related works.** Error correction is mainly used for analyzing input query error. It can return the correct form of the error in reasonable time. So when it executes error correction, response time and correctness of the return result should be taken into consideration. In current Chinese information retrieval system, English and Chinese query correction are the two main types.

English query correction method contains spelling correction method based on word and correction technology based on context information. Spelling correction method based on word pays attention to the single word spelling correction. And many methods are proposed such as editing distance method[10-12], K-gram contact ratio method[13,14] and spelling correction method based on phoneticize[15]. Correction technology based on context information not only is used for the error words not in dictionary, but for the impertinent words used in the context. This technology utilizes language model to evaluate every keyword in this query and select the optimal combination form including query correction based on noise channel model[16,17], query correction based on Bayesian classifier[18,19] and query correction based on maximum entropy model[20,21].

In the English correction technology, editing distance method and K-gram contact ratio method only correct single word error, they can rapidly find the standard word similar to the error word. But they cannot handle the inappropriate words in the context. Another three methods can deal with this situation. In English correction technology, the query error contains spelling error in English word, misuse and space loss etc. In Chinese search engine, there are many query errors, the above methods cannot better dispose.

The Chinese query correction methods always transform the Chinese character in query word into phoneticize. Then it searches the similar phoneticize in dictionary. Finally, it determines the correction result through word frequency or language model.

Zhang[22] proposed a unified framework called HANSpeller++ based on previous HANSpeller for Chinese spelling correction. The framework consisted of candidate generating, candidates re-ranking and final global decision making. Experiments showed good performance on the test data of the task. Xiao[23] illustrated two kinds of semantic input method for Chinese word senses such as Word-based word sense input and Sentence-based word sense input. These two kind of word sense input methods were based on the statistic word sense representation and disambiguation. The prototype system showed both Word-based and Sentence-based word sense. Pinyin methods were promising in the text edit system. These two kinds of Chinese word sense input methods were designed for semantic document edit: syntactic file and semantic file aliment system, designed for the semantic document exchange for e-business. To improve the correction speed, Guo[24] put forward a kind of Chinese text classification oriented Structural Auxiliary Word algorithm. The algorithm used the special space effect of Chinese text where words had an implied correlation between text information mining and text categorization for high-correlation matching. However, the above proposed methods only can deal with query single word item. And they do not consider the Chinese, pinyin, English, numbers in Chinese information retrieval system. So we propose a two-way parallel query correction approach based on semantic analysis and reverse hidden Markov model used for mixed language in computer language.

3. **New two-way parallel query correction approach.**

3.1. **Mixed language coding.** Chinese query correction deals with Chinese character which is determined by Pinyin. Each Chinese character can correspond to a string or strings of pinyin letter combination [25]. English query correction deals with English words which is determined by English alphabet. Each English word corresponds to strings of alphabets. Through coding English word, the Chinese characters and English words can be mapped into unified coding region. Then it can manage the English and Chinese uniformly. In unicode, the English alphabet coding range is: 0x0041-0x007a. Chinese character coding range is: Ox4E00-0x9FFF. For the Chinese search engine, the unicode character beyond 0x9FFF can be ignored.

In this paper, the unicode character beyond 0x9FFF is used as English word to code. So English word and Chinese character can be transformed into character sequence to uniformly manage. In Chinese search engine, user query involves not much English words, so it only needs to count the English words in query log. Except English words, it can code the common alphabet, number and their combination form, such as 'Tianmao' and 'sina'.

In the new parallel query correction method, we use the above coding way to preprocess the training data. Then we use the coded training data to build the proposed character dictionary tree and language model. Therefore, error correction of English, Chinese and other common character combination can be uniformly handled.

3.2. **Two-way parallel correction error model based on semantic analysis.** In order to redress the correction error, it needs to set up a corresponding language model. According to the different language model, there are different correction methods, such as statistical language model, lexical functional grammar error correction method. In this paper, we present a correction method based on the sentence semantic analysis. According to the concept of hierarchical network language model, a statement is processed with semantic analysis, it can get sentence category. According to other sentence types knowledge, a semantic block can be acquired. Semantic block includes feature semantic block and generalized object semantic block. For the Chinese long query, if it adopts single direction query correction model, the correction time will increase and accuracy of correction will decrease. So this subsection presents a two-way parallel error correction model based on semantic analysis.

Two-way parallel error correction method based on semantic analysis indicates that it conducts error correction from the both ends of query word with semantic analysis. For the forward query correction model (i.e. From the left to right of the sentence), it needs to build reverse hidden Markov model. Then according to semantic analysis, it executes correction from left to right for query word. For the reverse query correction model, it needs to build reverse character dictionary and reverse language model with semantic analysis too. Then it executes correction from right to left for query word. When forward and reverse correction meets at point D, it joints overlap part and gets a whole correction sentence. The process of two-way parallel error correction method based on semantic analysis is as the following.

Input. Dictionary tree corpus DTC, language model corpus LMC, pseudocode corpus PSE and user query word QW_1.

Output. Word QW_2 after correction.

1. Executing Reverse Query Correction.
2. Using PSE builds pseudocode words list T_1.
3. Using pseudocode words list T_1 and LMC builds forward language model FL_1 and reverse language model FL_2 with semantic analysis.

4. Using pseudocode words list T_1 and FL_1 builds forward character dictionary tree $FCDT_1$ and reverse character dictionary tree $FCDT_2$ with semantic analysis. It means that storing language model information in character dictionary tree nodes avoiding repeat calculation in correction process and reducing time cost.
5. If $|QW_1| \leq M_1$. M_1 is the threshold value of query word length. Once exceeding this threshold, it will adopt two-way parallel error correction based on reverse hidden Markov model model.
6. Executing Forward Query Correction.
7. Invoking forward processing function $EditModel$ and getting candidate set R. In R, candidate item editing distance cost is $edit-cost$, language model cost is $lang-cost$.
8. Else
9. Computing mutual information, segmentation query words q_1, q_2.
10. q_1 is handled in forward correction model. Invoking forward processing function $EditModel$. q_2 is handled in reverse correction model.
11. Merging results and getting candidates set R.
12. If $|R| > 0$ and $\alpha \cdot edit - cost + \beta \cdot lang - cost < M_2$. Where M_2 is cost threshold.
13. QW_2 is equal to the optimal item in R.
14. Else
15. $QW_2 = QW_1$.
16. Return QW_2.

3.3. Building dictionary tree and language with reverse hidden Markov model.

- Building dictionary tree. Character dictionary tree is a fast retrieval multi-way tree structure[26]. It utilizes the common prefix of character string to save storage space. Root node of character dictionary tree does not contain character. Except root node, each node only includes one character. If a character string constructed by root node and node A is a legal word processed by pseudocode, then A can be remarked as complete node and A stores corresponding word information. The whole nodes in character dictionary tree are divided into two parts: complete state nodes and incomplete state nodes.

  Then we propose the reverse character dictionary tree based on forward character dictionary tree. Its aim is to reverse Pinyin string corresponding to Chinese character and store the character information.
- Building language model with reverse hidden Markov model. Hidden Markov model is a doubly stochastic process. From initial state to transfer, ended with finial state. In this process, all the vectors are defined as hidden state items. The forward language model denotes the probability after a series of activities[27]. Chen [28] presented a machine learning approach to discover the agent dynamics that drived the evolution of the social groups in a community. It set up the problem by introducing an agent-based hidden Markov model for the agent dynamics. Meanwhile, n-gram language based on reverse hidden Markov model denotes that the probability of $n-th$ word is determined by n-1 words. Take ternary grammar model for example, assuming that the probability of one word only depends on its previous two words. When $n > 2$, the probability of sentence s composed of word or character with length L can be expressed as:

$$p(s) = \prod_{i=1}^{L+1} p(a_i|a_{i-n+1}^{i-1})(0 \leq i \leq L). \tag{1}$$

Where a_i^j denotes words $(a_i, a_{i+1}, \cdots, a_j)$. It can use maximum likelihood estimation method to compute $p(a_i|a_{i-n+1}^{i-1})$ as formula (2):

$$p(a_i|a_{i-n+1}^{i-1}) = \frac{c(a_{i-n+1}^{i})}{\sum_{a_i} c(a_{i-n+1}^{i})}. \tag{2}$$

Because the data from training corpus are obtained in web log and query log, it cannot cover all nature language, $p(s) = 0$ may appear. It needs to use Witten-Bell smoothing method. In this method, n-order smoothing recursion model is defined as the linear interpolation between n-order maximum likelihood model and n-1 order smoothing model.

$$p(a_i|a_{i-n+1}^{i-1}) = \lambda_{a_{i-n+1}^{i-1}} p_{ML}(a_i|a_{i-n+1}^{i-1}) + (1 - \lambda_{a_{i-n+1}^{i-1}} p_{WB}(a_i|a_{i-n+2}^{i-1})). \tag{3}$$

$$N_{1+}(a_{i-n+1}^{i-1}\epsilon) = |a_i : c(a_{i-n+1}^{i-1}a_i)|. \tag{4}$$

$$1 - \lambda_{a_{i-n+1}^{i-1}} = \frac{N_{1+}(a_{i-n+1}^{i-1}\epsilon)}{N_{1+}(a_{i-n+1}^{i-1}\epsilon) + \sum_{a_i} c(a_{i-n+1}^{i-1})}. \tag{5}$$

Where $N_{1+}(a_{i-n+1}^{i-1}\epsilon)$ denotes the different words number behind a_{i-n+1}^{i-1}. ϵ is the free variable.

The reverse reverse hidden Markov language model denotes the probability before a series of activities. Reverse language model is constructed from the end of sentence. Forward language model estimates the appearing probability of current word according to history information. So reverse language model is transformed from forward language model and estimates the appearing probability of current word according to prospective information. The probability of sentence s composed of word or character with length L can be expressed as:

$$p(s) = p(a_L)p(a_{L-1}|a_L) \cdots p(a_1|a_L a_{L-1} \cdots a_2). \tag{6}$$

Where the probability of $a_i (0 \le i \le L)$ is determined by prospective information $a_{i+1}a_{i+2} \cdots a_L$. We also use Witten-Bell smoothing method to solve zero probability problem.

3.4. **Determining cut-off point and editing rule.** In new two-way parallel error correction model, cut-off point P is obtained by computing the mutual information between words. For Chinese long query, it needs to scan character and compute mutual information between two words. Mutual information between M and N can be defined as:

$$T(M, N) = log_2 \frac{P(M, N)}{P(M)P(N)}. \tag{7}$$

Where $P(M, N)$ is the appearing probability of Chinese character MN. $P(M)$ and $P(N)$ are the appearing probability of Chinese character M and N respectively. If $T(M, N) > 0$, then M and N are positive correlation. When $T(M, N)$ is greater than threshold, M and N can form a word. If $T(M, N) = 0$, then M and N are irrelevant. If $T(M, N) < 0$, then M and N are mutually-exclusive. M and N cannot form a word.

For query sentence s, it firstly is transformed into character sequence(Chinese is transformed into Pinyin, English and number are unchanged). Then each character is edited. If character is Chinese, then it uses Chinese rule to edit. Otherwise, it uses English rule. As we all know, Chinese editing process is very strict. In that replacing or inserting Pinyin will produce a lot of Chinese characters, which increases storage space and generates lots of inaccuracy correction results too.

Chinese rule editing contains homophones match, polyphony match, similar sound replacing, deleting and similar Chinese character replacing etc. Homophones match denotes that when Chinese character only corresponds to one Pinyin, the character will be replaced by its Pinyin. Polyphony match denotes that when Chinese character corresponds to several Pinyin forms, the character will be replaced by its Pinyin forms. Similar sound replacing denotes that Chinese character is replaced by its similar Pinyin. English rule editing contains match, replacing, inserting, deleting and character exchange. Match denotes that edited character is the English character. Replace denotes that this character is replaced by other 25 English alphabets. Inserting denotes that it inserts 26 English alphabets behind this character, and generates 26 character strings as editing result.

After editing, it searches for the character dictionary tree. If sequence has the corresponding Chinese character in character dictionary tree, then current state is remarked as complete state and added into complete and incomplete state queue (The reason why adding it into the two queues is that it is to save all the states generated by minimum and maximum match and ensures the accuracy of classification of words). Otherwise, it is remarked as incomplete state and added into incomplete state queue.

4. **Experiments and result analysis.** In the experiments, training data is extracted from query log, user click log and web link log in a Chinese commercial search engine. Training data of character dictionary tree is nearly 1.35 million, pseudocode is about 0.21million. And language model training data is approximately 0.14 billion. We randomly select 3800 query data from search engine as testing data. Alphanumeric combination is 40. Alphabet combination is 100. Number combination is 50. Chinese, alphabet and number combination is 3610. Error words in 3800 query words are 200 accounting for 5.26%. Setting length threshold of query sentence is 6. And 2000 query words are long query taking up 52.63%.

We take Precision, Recall and F-measure to evaluate the system.

$$P = N_1/N_2. \tag{8}$$

$$R = N_1/N_3. \tag{9}$$

$$F_\delta = \frac{(\delta+1)PR}{\delta P + R}(\delta \geq 0). \tag{10}$$

Where N_2 denotes query number in query correction model. N_1 denotes the number of output correction. N_3 denotes error query number in testing set. In this paper, δ is set as 0.5.

Query Correction(QC) is one query correction method used for Chinese search engine mixed language. Reverse Query Correction(RQC) conducts correction from the end of query sentence based on reverse character dictionary tree and reverse language model. Two-way Parallel Query Correction(TRQC) is this paper's new scheme. Table1 is the comparison results with the three methods.

Table 1. Experiment results with different methods

Method	N_2	N_1	N_3	Time	$F_{\delta=0.5}$	R	P
QC	166	125	180	130.28s	73.98	73.82%	75.49%
RQC	178	132	180	138.17s	75.02	76.75%	73.56%
TRQC	141	121	180	81.72s	76.77	70.97%	84.38%

From table1, we can know that single correction can weaken operation cost of each word for Chinese long query. So it may put the correct query words into error words set. However, our new method can transform long query into short query, which enhances the operation cost of each word and improves the accuracy.

We also make comparison to error checking model n-gram model, SpEQ [29] (a Machine Learning based approach that generates corrections for misspelled queries directly from the user's own mail data), CICBF (causal inference technique under contextual-bandit framework) [30] with our new method TRQC under the same conditions. The results are shown in table2.

Table 2. Experiment results with different methods

Method	$F_{\delta=0.5}$	R	P
n-gram	74.53	73.62%	75.57%
SqEQ	76.09	72.56%	76.56%
CICBF	75.78	72.56%	77.38%
TRQC	77.64	69.25%	82.67%

In table2, TRQC has the biggest value 77.64 than any other methods. Recall reaches to 69.25% which is the lowest value than CICBF 72.56%, SpEQ 72.56% and n-gram 73.62%. In addition, precision of TRQC is 82.67%. Therefore, our method is the best choice for mixed language query.

However, when we use new method to conduct query correction, the context information of query word will be decreased. Some local right but global wrong query sentence can not be corrected. Then the right correction number will reduce too. If the long query sentence is segmented inappropriately. It can result in the same problem. Overall, the Recall reduces by 3% and correction speed increases 40% through our new method.

5. Conclusions.

In this paper, we propose a two-way parallel query correction approach based on semantic analysis and reverse hidden Markov model for mixed language in computer language, which is very useful for human-machine multimodal interaction. This method can uniformly deal with mixed language query (Chinese, Pinyin and English etc,). And reverse hidden Markov model are proposed based on forward character and language dictionary tree. It obtains the candidate set of query words by editing. Candidate item can be gotten through linear combination between editing distance weight and language model probability. Two-way parallel query correction with semantic analysis is introduced to deal with Chinese long query. Finally, experiments show that new parallel correction method for mixed language is more accuracy than single direction query. What's more, the speed and convergence time is improved too. In the future, we will study multi-channel parallel error correction method to further improve query correction system.

REFERENCES

[1] S. Tian,Y. Cai, Z. Hu, A Parity-Based Data Outsourcing Model for Query Authentication and Correction, *Conference on IEEE Distributed Computing Systems (ICDCS)*, pp. 395-404, 2016.

[2] J. Duan, P. Mi, H. Liu, Error Checking for Chinese Query by Mining Web Log, *Mathematical Problems in Engineering*, vol. 2015, Article ID 985204, 5 pages, 2015. doi:10.1155/2015/985204.

[3] D. Jiang , W. T.Leung, W. Ng , et al., Beyond Click Graph: Topic Modeling for Search Engine Query Log Analysis, *International Conference on Database Systems for Advanced Applications*. pp.209-223, 2013.

[4] T. Liu, S. Yin An improved particle swarm optimization algorithm used for BP neural network and multimedia course-ware evaluation, *[J]. Multimedia Tools and Applications*, pp. 1-14, 2016.

[5] Frommholz D, Linkiewicz M, Poznanska A M. Inlining 3d Reconstruction, Multi-Source Texture Mapping and Semantic Analysis Using Oblique Aerial Imagery, *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2016, XLI-B3:605-612.

[6] Hao S, Yanyan X U, Dengfeng K E, et al. SCESS: a WFSA-based automated simplified chinese essay scoring system with incremental latent semantic analysis[J]. Natural Language Engineering, 2016, 22(2):291-319.

[7] Sun Y, Teng L, Yin S, et al. Study a Join Query Strategy Over Data Stream Based on Sliding Windows[C]// International Conference on Data Mining and Big Data. Springer, Cham, 2017:334-342.

[8] Layfield C, Azzopardi J, Staff C. Experiments with Document Retrieval from Small Text Collections Using Latent Semantic Analysis or Term Similarity with Query Coordination and Automatic Relevance Feedback[C]// Semanitic Keyword-based Search on Structured Data Sources. Springer International Publishing, 2016:25-36.

[9] Yi D, Zhang Y, Wei B. Query Subtopic Mining via Subtractive Initialization of Non-negative Sparse Latent Semantic Analysis[J]. Journal of Information Science & Engineering, 2016, 32(5):1161-1181.

[10] R. Chaurasiya, N. Londhe, S. Ghosh, A Novel Weighted Edit Distance-Based Spelling Correction *Approach for Improving The Reliability of Devanagari Script-Based P300 Speller System*, PP. 1-1, 2016.

[11] M. Li, Y. Zhang, M. Zhu , et al., Exploring distributional similarity based models for query spelling correction, *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics. Association for Computational Linguistics* pp. 1025-1032, , 2006.

[12] Q. Chen , M. Li, M. Zhou, Improving Query Spelling Correction Using Web Search Results, *Conference on EMNLP-CoNLL*, vol. 7, pp. 181-189, 2007.

[13] J. Feng , S. Bangalore, Effects of word confusion networks on voice search, *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, pp. 238-245, , 2009.

[14] M. J. Zaki , C. D. Carothers , B, K. Szymanski, VOGUE: A variable order hidden Markov model with duration based on frequent sequence mining, *Journal of ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 4, no. 1, pages 5, 2010.

[15] J. Gao , X. Li , D. Micol , et al., A large scale ranker-based system for search query spelling correction, *Proceedings of the 23rd International Conference on Computational Linguistics. Association for Computational Linguistics*, pp. 358-366, 2010.

[16] I. Elawady , A. M. Lakhdar, K. Mustapha The Noise Reduction over Wireless Channel Using Vector Quantization Compression and Filtering, *Journal of International Journal of Electrical and Computer Engineering (IJECE)*, vol. 6, no. 1, pp. 130-138, 2016.

[17] W. Naiqi,C. Zili , L. Junwei , et al., Analysis on Channel Capacity of Transform Domain Communication System, *Journal of Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 4, pp. 2790-2796, 2014.

[18] B. O. Mainsah , K. D. Morton, L. M. Collins , et al., Moving away from error-related potentials to achieve spelling correction in P300 spellers, *Journal of IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no.5, pp. 737-743, 2015.

[19] J. Liu , S. Yin , L. Teng, an improved multiple extended target tracking algorithm based on variational bayesian cardinality equilibrium multi-objective bernoulli filtering, *Journal of Icic Express Letters Part B Applications An International Journal of Research & Surveys*,7, 2016.

[20] C. Chelba , T. Mikolov , M. Schuster , et al. One billion word benchmark for measuring progress in statistical language modeling, *Journal of arXiv preprint arXiv*, pp1312, 3005, 2013.

[21] F. Peng , S. Roy , B. Shahshahani, et al., Search results based N-best hypothesis rescoring with maximum entropy classification, *Conf. on ASRU*. pp. 422-427, 2013.

[22] S. Zhang ,J. Xiong1 ,J. Hou , et al., HANSpeller++: A Unified Framework for Chinese Spelling Correction, *Journal of ACL-IJCNLP* , pp. 38, 2015.

[23] G. Xiao ,J. Guo , Z. Gong , et al., Semantic input method of Chinese word senses for semantic document exchange in e-business, *Journal of Journal of Industrial Information Integration*, vol. 3, pp. 31-36,2016.

[24] X. Guo , H. Sun ,T. Zhou , et al., *SAW Classification Algorithm for Chinese Text Classification-Journal of Sustainability,* vol. 7, no. 3, pp. 2338-2352, 2015.

[25] M. Li, M. Zhou, Distributional similarity-based models for query correction, *U.S. Patent* 7,590,626[P]. 2009-9-15.

[26] C. Shi , C. Wang ,B. Xiao , et al., Scene text recognition using part-based tree-structured character detection, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* pp. 2961-2968, 2013.

[27] K. Heafield, I. Pouzyrevsky, J. H. Clark, et al., Scalable Modified Kneser-Ney Language Model Estimation, Conference on ACL, vol. 2, pp. 690-696, 2013.

[28] Chen H C, Goldberg M, Magdonismail M, et al. Reverse engineering a social agent-based hidden markov model–visage.[J]. International Journal of Neural Systems, 2008, 18(6):491-526.

[29] A.Bhole, R. Udupa, On Correcting Misspelled Queries in Email Search, *Conf. on AAAI. 2015: 4266-4267. Li L, Chen S, Kleban J, et al. Counterfactual estimation and optimization of click metrics in search engines: A case studyConf. on Proceedings of the 24th International Conference on World Wide Web. ACM*, pp. 929-934, 2015.

[30] L. Li , S. Chen , J. Kleban, et al., Counterfactual Estimation and Optimization of Click Metrics in Search Engines, *A Case Study* , vol. 25, no. 2, pp.929-934, 2015.