Requirement Consistency and Integrity Verification Method based on Natural Language Processing

Chao Wu, Zhong Hua Huang, Yu Ting Yang* and Yue Liu

Beijing Institute of Computer Technology and Applications, BeiJing 100854, China

*E-mail: casic_zp@163.com

Abstract. In the iterative process of the update of the information service system, the continuous improvement of the system is restricted due to problems such as non-standard, incomplete, and strong ambiguity in the description of requirements. It makes it difficult to verify the consistency and integrity of requirements. In this paper, we focus on the huge requirements of complex information service system, and propose research on requirement itemized extraction and structured description technology based on natural language processing and requirement item keyword extraction and standardized expression technology. The library management system is used as the verification object to verify the effectiveness of the method proposed in this paper.

Keywords: requirement extraction, itemized, structured, key word extraction, consistency, integrity

1. Introduction

With the development of information technology, the version of the information service system is updated faster and more complex. Requirement changes have become the norm in the information service system development process. This is a challenge to the research of information service system requirement management. If the requirements cannot be updated and extracted in time, the requirements may be inconsistent and incomplete, which will lead to the development of systems that do not meet the needs of users, increase the workload of developers, consume time, and increase costs. Therefore, it is important to consider the dynamic evolution and change of requirements and maintain the continuous consistency and integrity of requirements. Traditional requirement description is based on multi-source, unstructured natural language, which makes the expressed requirement irregular, incomplete, and ambiguous. All problems can lead to incomprehensible requirements, deviations in implementation, and even project failure. Therefore, it is necessary to research requirements extraction and standardization technology based on natural language processing, as well as consistency and integrity verification technology.

In summary, we research requirements consistency and integrity verification methods based on natural language processing and knowledge mining technology. First, we research the requirement extraction and structuring methods for information service systems; Then, we propose keyword extraction and standardization technology for requirement items; Finally, we carry out research on the consistency and integrity of requirements. Based on this, it assists users to understand demand items, thereby ensuring the successful development and iterative evolution of the information service system.
2. Related work
This paper introduces related work from two aspects: formal requirements description methods and requirements integrity and consistency detection methods.

2.1. Formal requirements description methods
Currently, there are usually three forms of requirements documents. They are requirements documents based on natural language, semi-formal representation method [1], and formal representation method [2]. Documents based on natural language are easier to read and build, but their definitions are often not strict enough and ambiguities. Semi-formal requirements documents are not easy for formal reasoning and verification. Formal representation methods are generally based on mathematics and have strict formal definitions that are convenient for reasoning and testing [3].

Bertrand Meyer elaborated the advantages of applying formal description methods to software development in the literature [4], especially the significance and advantages of the software requirements phase. Z method [5] is a formal method based on set theory, which defines the mapping relationship and operation rules from the perspective of set. The VDM method [6] is a model-oriented formal method that establishes a series of data types. Temporal Logic method [7] is an attribute-oriented method, which is mainly applied to software description of concurrent and distributed systems. The Event-B model [8] is also a formal requirement description method. A general model of Tabular expression is proposed in the literature [9], which can also be used for formal requirements description.

The main reason for the formalization of requirements is that natural language has the following characteristics: ambiguity, disorder, contradiction, inintegrity, inaccuracy and omission.

There are mainly the following tools for formal expression of natural language: SPECIFIER [10], RA [11], KBRA [12], KATE [13], NL2ACTL [14] and FORSEN [15], etc.

2.2. Requirements consistency and integrity verification methods
Requirements verification is one of the current hot issues in the field of software engineering. The basic idea is to transform the original requirement text into a corresponding requirement verification model, and perform verification work based on the characteristics of the requirement model. The current main requirements verification models include: semantic-based verification model, semi-formal verification model, mathematical verification model, and ontology-based verification model [16].

The semantic-based verification model detects inconsistencies in requirements through lexical, grammatical and semantic analysis of the original requirements text. For example, Nathan Weston [17] uses RDL to describe requirements, and relies on requirements classification semantic analysis to verification the consistency of requirements. Literature [18] uses Bayesian networks to analyze requirement semantics to verify requirement. This method of semi-formal verification model mainly uses state machine [19], UML model [20] or feature tree requirement model [21], focus model [22] and other requirement models for direct analysis and verification. The mathematical verification model mainly detects and deals with the inconsistencies in the distributed requirements specification [23]. In terms of using ontology to verify the requirements model, Chi-Lun Liu [24] defined a set of ontology-based requirements conflict analysis methods. This method defines the requirement action diagram based on ontology, and establishes corresponding conflict detection rules to analyze the conflicts in the requirement.

3. Requirement consistency and integrity verification solution
The overall scheme is shown in Figure 1. It includes three parts: 1) natural language processing-based requirement itemized extraction and structured description technology; 2) TextRank-based requirement item keyword extraction and standardized expression technology; 3) ontology model-based requirement consistency and integrity verification technology.
3.1. Natural language processing-based requirement itemized extraction and structured description technology

In order to achieve intelligent requirement management and intelligent analysis of the impact of requirement changes, requirement items information needs to be extracted from unstructured documents. This part mainly includes two parts of research content: requirement itemized extraction technology based on requirement template and requirement itemized classification and structured description technology based on CNN.

3.1.1.Requirement itemized extraction technology based on requirement template. Users describe requirements in different ways and expression habits, but usually the internal expression logic of requirements is consistent. Through data mining, common patterns of user expression logic can be extracted, which becomes a requirement template. The premise of establishing a mining model is to formally express the user's expression logic.

In this paper, first, the domain word segmentation and part-of-speech tagging models are studied to transform the requirement text into a part-of-speech tagging sequence. Then, use a concise and effective way to model the logic in the requirement description to extract the requirement template. Finally, the requirement itemized extraction is realized based on the requirement template.

(1) New words recognition and adaptive word segmentation method

New words are words that do not appear in the dictionary, and are different from OOV (Out of Vocabulary). New words are a dynamic concept that incorporates time, while OOV is relative to the dictionary. Their classification methods are different, and the relationship diagram is shown in Figure 2.
In this paper, a rule-based method is used to identify new words. On the one hand, build a new vocabulary based on the knowledge of word-formation of new words, and build a special vocabulary database based on the characteristics of words on the other hand.

(2) Requirement pattern extraction method based on statistical analysis
In this paper, we extract and analyze domain requirement sentence patterns, obtain common requirement description rules, and construct requirement sentence patterns. A semi-automatic requirement sentence pattern extraction method is adopted. First, extract the domain sentence pattern rules from the domain requirements after the word segmentation is marked, and combine expert opinions to summarize the requirement sentence pattern. Then, according to its description rules, write regular expressions to realize each type of requirement sentence pattern. Finally, extract the new requirements, and modify the requirement sentence pattern based on the accuracy of the extraction.

3.1.2. CNN-based requirement classification and structured description technology. CNN makes full use of its advantages in semantic analysis to transform the problem of extracting requirement items into a problem of requirement text classification. The requirement classification model established in this paper includes convolutional layer and pooling layer. In this paper, the length of the convolution kernel is set as the dimension of the word vector, and the convolution operation is performed to extract the semantics of the overall word vector, so as to avoid the problem of incomplete use of information due to partial operations on the word vector. Different requirement information corresponds to different word sequence lengths. Using the above method, texts of different lengths can be generated into feature vectors of the same size, and simple classifiers (Sigmoid, LR, etc.) are used for requirement information classification.

The method steps are:
(1) Using domain word segmentation processing technology, the requirement information in text format is transformed into word sequences.
(2) Use the Skip-gram model to train word vectors, obtain the distributed representation of each word, and then build an index table of words to vectors.
(3) By mapping words into word vectors, the word sequence is transformed into a two-dimensional feature matrix.
(4) On the basis of the two-dimensional feature matrix, CNN use the different convolution kernels to act on the text word vector to extract the semantic information of the required text.
(5) Use the pooling operation to splice the local semantic information into sentence vectors, and then use the classifier to train the requirement classification model.

The requirement ontology model has defined the characteristics of the requirement and the attributes that the requirement contains. To realize the structuring of requirements, it is necessary to extract sentence information corresponding to different requirement attributes from the description of a certain requirement based on requirement itemization, to realize the transformation from requirement itemization to requirement structured.

3.2. TextRank-based requirement item keyword extraction and standardized expression technology
The TextRank algorithm is a concrete realization of the graph ranking method in the field of natural language processing. First, the word weight is calculated according to the TF-IDF algorithm, and the words with larger weights are selected as candidate keywords. Then, analyze the relationship between candidate keywords in the document, build a TextRank model to generate keyword sequences, and extract keywords. Its overall framework is shown in Figure 3.

**Figure 3.** Framework structure diagram of requirement keyword extraction method based on TextRank.
The Trie Tree is used to organize the keyword information of the massive requirement documents, and the leaf nodes store the requirement attribute subject information corresponding to the keywords, to realize the fast search of keywords and the positioning of the requirement subject. Introduce the fast retrieval technology based on optimized double array Trie Tree to compress the storage structure of Trie Tree. The structure and query method of Trie Tree are optimized based on the sorting strategy. On the one hand, it guarantees quick retrieval of the requirement attribute theme corresponding to the requirement information, on the other hand, it solves the serious waste of space caused by the Tire Tree algorithm. Finally, the rapid and accurate positioning and analysis of keywords in the requirement text is realized.

4. Requirements consistency and integrity verification

In order to realize the requirement consistency and completeness verification based on the domain ontology model, it is necessary to build a requirement ontology model library. The specific process is as follows: First, integrate domain standards, historical documents, evolutionary data and expert experience, and build a domain requirement knowledge system through knowledge collection and knowledge mining technology. Then, based on this system and the principles of ontology construction, a requirement ontology model is constructed, and finally a requirement ontology model library is formed. According to the consistency constraint and integrity constraint in the requirement ontology model, with the help of ontology reasoning method, the consistency and integrity verification of the extracted requirements is carried out.

4.1. Requirements consistency

In the field of information services, the requirements document is described in unstructured natural language. The inherent ambiguity of natural language can lead to difficult understanding of requirements and even project failure. Therefore, the consistency of requirements needs to be verified. At present, the main requirements verification models include semantic-based verification models, formal requirements analysis models, mathematical verification models and ontology-based verification models. In this paper, we adopt an ontology-based requirement consistency verification model. The verification process is as follows: Identify structured requirement items from the requirement specification, extract requirement keywords, and express them in a standardized manner. Based on the consistency constraint rules, it is compared with the requirement ontology model, and the requirement consistency verification result is obtained.

4.2. Requirements integrity

The integrity of the review requirements should be considered as follows:

1. Are there missing functions or business processes?
2. Is there an interface between each defined function?
3. Are messages or messages passed between defined functions?
4. Is the user of the function defined?
5. Is the interaction between users and functions clearly defined?
6. Are interfaces with external processes and systems defined?
7. Can the described functions be mapped to the business process?
8. Are there any references to requirements to be determined in the document?
9. Are there undefined terms and references in the document?
10. Is every part of the document complete?

The integrity analysis of non-functional requirements includes whether software performance is considered whether security requirements are considered, whether reliability is considered, and whether system capacity is considered.

The integrity constraint mainly includes the limit of the number of concept instances; The indispensable attribute of the concept instance; Properties between the properties of a concept instance.
4.3. Ontology model-based requirements consistency and integrity verification

This paper uses the requirement ontology model library established above, and on the basis of structuring requirement entities for requirement entries, introduces requirement consistency verification and integrity analysis based on ontology models, and establishes ontology models for different types of requirements. The ontology models contain attributes. Information and dynamic expansion can be added constraints. Requirement entities of the same ontology can directly verify the consistency of constraints based on ontology reasoning, and use the integrity constraints in the ontology model to analyze the integrity of the requirement entities.

The content of requirements verification [26] based on domain ontology is shown in Table 1:

| category                  | Description                                                                 |
|---------------------------|-----------------------------------------------------------------------------|
| security constraints      | Certain states or state transitions should not appear in the requirements description. Certain states will inevitably appear in the requirement description; certain state transitions will inevitably appear in the requirement description. |
| activity constraints      | Express the triggering relationship between states. Indicates that the appearance of a certain state will inevitably lead to the appearance of another state. The trigger relationship between multiple states constitutes a trigger sequence relationship. |
| requirements within nodes | First, determine whether there is overlap in the requirement descriptions of the two nodes, and secondly, determine whether the description of the same entity has the same state, and finally, on this basis, verify the requirement between multiple nodes. |

According to the above categories of requirement verification, based on the requirement ontology model library in the domain, the consistency and integrity of the requirements are verified. Based on the above constraints, the integrity of the requirement ontology and the standardized requirement entity is verified. On this basis, the consistency verification of the relationship between the requirement entity and the entity is carried out, and finally the requirement consistency and integrity verification results are obtained, which guides the development and evolution of the system.

5. Case verification

In this section, for a specific case, test the consistency and integrity between the requirement keywords extracted by this method and the requirement based on expert knowledge, and give the analysis results. Given a series of requirements texts for large, medium and small library management systems. These standardized requirements texts can be used as requirements analysis cases. The requirement text describes the specific requirements of the library management system. In terms of functional requirements, it includes the requirements of borrowers and administrators. Non-functional requirements include user interface and security. The specific process is shown in Figure 4:

First, based on the requirement template and matching rules, the requirement items are extracted from the requirement documents of the three library management systems. Second, CNN is used to classify the requirement items and express them in a structured manner. Third, the Trie Tree is used to identify the keywords in the requirement, and TextRank is used to rank the importance of the
Finally, the keywords extracted in this paper are verified for consistency and integrity with the ontology model based on expert knowledge.

Figure 4. The specific process of case verification.

5.1. Requirement itemized extraction technology based on requirement template
The requirement item extraction process based on the requirement item is shown in Figure 5: First, extract some rules from the requirement text of the library management system, combine expert opinions, and summarize the requirement sentence pattern. Then, for each type of requirement sentence pattern, according to its description rules, write regular expressions to get the corresponding requirement sentence pattern. Finally, use the requirement sentence pattern to extract the new requirement and get the requirement item.

Figure 5. The requirement item extraction process.

5.2. Requirement item classification and structured technology
The classification process of requirement items is as follows: First, the requirement items are segmented, and the Skip-gram model is used to construct an index table of words to vectors, and texts of different lengths are generated into feature vectors of the same size. Next, define the types of requirements, including target requirements, business requirements, functional requirements, performance requirements, constraints and restrictions, etc. Finally, the CNN network model is trained with the processed feature vector as input and the requirement type as classification basis.

When training the network, the loss function adopts the cross-entropy loss function. Use Adam optimizer for parameter learning. The learning rate of the model is $10^{-4}$. This paper uses accuracy, precision and recall to evaluate the effectiveness of the method. Comparing the method proposed in this paper with the method in the literature [26], the results are shown in Table 2:

Table 2. Classification result comparison table.

| Experimental method | Accuracy | Precision | Recall  |
|---------------------|----------|-----------|---------|
| SVM                 | 0.812    | 0.801     | 0.831   |
| CNN                 | 0.878    | 0.823     | 0.910   |
It can be seen from Table 2 that the classification accuracy of CNN for the 5 types of requirements reaches 87.8%, which is 6.6% higher than that of SVM. The CNN method also has a significant improvement in accuracy and recall. The accuracy and recall have increased by 1.2% and 7.9%, respectively.

Use the requirement document of the book management system to test, then the requirement items extracted by the above method (take 10 as an example) are shown in Table 3.

**Table 3. Requirement item classification table.**

| number | requirement items                                      | requirement               |
|--------|--------------------------------------------------------|---------------------------|
| 1      | Support to store book information                      | target requirements       |
| 2      | Provide a certain safety mechanism                     | performance requirements  |
| 3      | Support to add, delete and modify book information      | functional requirements   |
| 4      | Support remote booking and continue borrowing of books  | functional requirements   |
| 5      | Realize the management functions of book storage, borrowing, return, retrieval, reader management, and query | business requirements  |
| 6      | Require friendly user interface and provide help        | constraints and restrictions |
| 7      | Provide information backup services                    | performance requirements  |
| 8      | Support the storage and management of reader information | target requirements       |
| 9      | Support users to perform retrieval operations          | functional requirements   |
| 10     | Support librarians to print and statistical operations | functional requirements   |

Based on the above classification results, the structured requirement items are shown in Table 4:

**Table 4. Structured requirement items.**

| requirement items                                      | requirement type                                      |
|--------------------------------------------------------|-------------------------------------------------------|
| target requirements                                    | Support to store book information                     |
|                                                        | Support the storage and management of reader information |
| performance requirements                              | Provide a certain safety mechanism                     |
|                                                        | Provide information backup services                    |
| functional requirements                               | Support to add, delete and modify book information      |
|                                                        | Support remote booking and continue borrowing of books |
|                                                        | Support users to perform retrieval operations          |
|                                                        | Support librarians to print and statistical operations |
| business requirements                                 | Realize the management functions of book storage, borrowing, return, retrieval, reader management, and query |
| constraints and restrictions                          | Require friendly user interface and provide help       |

5.3. Requirement keywords recognition and extraction technology
The Trie Tree is used to quickly locate the keywords in the structured requirement items, and then the TextRank technology is used to sort the importance of the keywords to achieve the extraction of keywords to evaluate the effect of keyword extraction. This paper chooses three indicators: accuracy (P value), recall (R value) and F value. Comparing the method proposed in this paper with the method
proposed in [27], for the library management system, the index value comparison results of the number of keywords 3, 5, 7, 10 extracted are shown in Table 5, Table 6, Table 7.

**Table 5. P value comparison.**

| feature             | 3   | 5   | 7   | 10  |
|---------------------|-----|-----|-----|-----|
| TFIDF               | 0.5224 | 0.4310 | 0.3742 | 0.3097 |
| Tire Tree+TextRank  | 0.7197 | 0.7092 | 0.6697 | 0.6236 |

**Table 6. R value comparison.**

| feature             | 3   | 5   | 7   | 10  |
|---------------------|-----|-----|-----|-----|
| TFIDF               | 0.1873 | 0.2379 | 0.2521 | 0.3287 |
| Tire Tree+TextRank  | 0.2070 | 0.3400 | 0.4495 | 0.5685 |

**Table 7. F value comparison.**

| feature             | 3   | 5   | 7   | 10  |
|---------------------|-----|-----|-----|-----|
| TFIDF               | 0.2324 | 0.2467 | 0.3180 | 0.3549 |
| Tire Tree+TextRank  | 0.3216 | 0.4596 | 0.5379 | 0.5805 |

Table 5 - Table 7 are the comparison between the Tire tree+TextRank algorithm proposed in this paper and the classic TFIDF algorithm. It can be seen that the method proposed in this paper is superior to the classic method in terms of extraction effect. Its advantage is to construct a co-occurrence network through sliding windows, iteratively calculate the importance of words, and output the top words as keywords.

The extracted keywords are shown in Figure 6:

![Figure 6. Keyword extraction result of requirement items.](image_url)
As shown in Figure 6: the keywords with different importance levels extracted by TextRank technology are expressed in different colors. Red, yellow, blue, and purple indicate that the importance of keywords decreases in order, and the three red keywords indicate the most important keywords.

5.4. Requirements consistency and integrity verification

Two indicators, accuracy rate and recall rate, are selected to evaluate the consistency and integrity of requirements. These two indicators are defined as follows:

\[
\text{Accuracy} = \frac{\text{The number of similar keyword pairs correctly identified}}{\text{Total number of similar keyword pairs identified}} \tag{1}
\]

\[
\text{Recall} = \frac{\text{The number of similar keyword pairs correctly identified}}{\text{Total number of existing similar keyword pairs identified}} \tag{2}
\]

The consistency and integrity of the same requirement text and different requirement texts are verified, and the results are in Table 8:

| Text type                  | Accuracy   | Recall   |
|----------------------------|------------|----------|
| the same requirement text  | 71.21%     | 71.21%   |
| the different requirement text | 79.41%     | 82.11%   |

It can be seen from Table 8 that the method proposed in this paper can extract the requirement keywords from requirement text documents. And its consistency and integrity are better; The accuracy and recall of different requirement text calculations are significantly higher than the same requirement text. The main reason is that some operations between different texts are in the same text and are calculated repeatedly. A certain degree of relief can be obtained by adjusting the recall rate, but the evaluation value of different texts will still be greater than the same text.

6. Summary

In this paper, we propose a requirement consistency and integrity verification method based on natural language processing. First, extract and structure requirement items; Then, extract and standardize the keywords of the requirement items; Finally, the requirements consistency and integrity are verified based on the ontology model. It can support data-based requirement extraction and evolution analysis and requirement correlation analysis for requirement changes.

The method proposed in this paper can be applied to the requirement management of information service system, avoiding the time-consuming, labor-intensive and error-prone situation of manual management of requirement. In addition, it can also be used to quickly retrieve and intelligently recommend the requirement data of the information service system.

References

[1] L Musat, M Hubl and A Buzo. Semi-formal representation of requirements for automotive solutions using SysML[M]/ Languages, Design Methods, and Tools for Electronic System Design, 2016.

[2] H Jie. Requirements analysis based on sof formal method[J]. Computer Applications And Software, 2007.

[3] HH Liu. Research on some issues in requirements engineering [D].

[4] B Meyer. On formalism in specifications[J]. IEEE Software, 1985.

[5] Spivey and J. M. Understanding Z:a specification language and its formal semantics[M]. Cambridge University Press, 1988.

[6] Jones C B. Systematic software development using VDM[M].Englewood Cliffs: Prentice Hall, 1990.
[7] Koymans R, Vytopil J and W P De Roever. Real-time programming and asynchronous message passing[J]. 1983:187-197.
[8] Abrial J R. Refinement, decomposition, and instantiation of discrete models: application to Event-B[J]. Fundamenta Informaticae, 2006.
[9] Jin Y and Parnas D L. Defining the meaning of tabular mathematical expressions[J]. ence of Computer Programming, 2010, 75(11):980-1000.
[10] Miriyala, Kanth and Harandi. Automatic derivation of formal software specifications from informal descriptions[J]. IEEE Transactions on Software Engineering, 1991.
[11] Reubenstein H B and Waters R C. The requirements apprentice: automated assistance for requirements acquisition[J]. IEEE Transactions on Software Engineering, 1991, 17(3):226-240.
[12] Czuchry A J J and Harris D R. KBRA: A new paradigm for requirements engineering[J]. IEEE Expert, 1988, 3(4):21-24.
[13] Harandi and Mehdi T. Workshop on software specification and design[J]. Acm Sigsoft Software Engineering Notes, 1988, 13(1):29-29.
[14] Fantechi A, Gnesi S and Ristori G. Assisting requirement formalization by means of natural language translation[J]. Formal Methods in System Design, 1994, 4(3):243-263.
[15] Vadera S and Meziane F. From english to formal specifications[J]. The Computer Journal. 1994, 37(9):753-761P.
[16] Seok Won Lee and R A. Gandhi. Ontology-based active requirements engineering framework[C]/Proceedings of the 12th Asia-Paciﬁc Software Engineering Conference. Washington: IEEE, 2005:481-490.
[17] Weston N, Chitchyan R and Rashid A. Formal semantic conﬂict detection in aspect-oriented requirements[J]. Requirements Engineering, 2009, 14(4):247-268.
[18] Sardinha A, Chitchyan R and Weston N. EA-Analyzer: automating conflict detection in a large set of textual aspect-oriented requirements[J]. Automated Software Engineering, 2013, 20(1):111-135.
[19] Joao Araujo and Jon Whittle. Modeling and composing scenario-based requirements with aspects [C]. Proceedings of the 12th IEEE International Requirements Engineering Conference(RE ’04), Washington: IEEE, 2004:122-131.
[20] Mehner K, Monga M and Taentzer G. Interaction analysis in aspect-oriented models[C]// IEEE International Requirements Engineering Conference. IEEE, 2006.
[21] Roubtsova, EE Ella, Roubtsov and SA Serguei. A feature computation tree model to specify requirements and reuse[C]// Icise - Eighth International Conference on Enterprise Information Systems: Databases & Information Systems Integration. DBLP, 2006.
[22] Kasirun, Zarina Mohd, Salim and Siti Salwah. Focus group discussion model for requirements elicitation activity[C]// International Conference on Computer & Electrical Engineering. IEEE Computer Society, 2008.
[23] Kedian Mu, Qi Zhang and Zhi Jin. Verifying Software Requirements Based on Answer Set Programming[C]// Knowledge Science, Engineering & Management, Third International Conference, Ksem, Vienna, Austria, November. DBLP, 2009.
[24] Liu C L. Ontology-based requirements conflicts analysis in activity diagrams[J]. 2009.
[25] Si-Li W, Hua-Ping Z and Bin W. Research of optimization on Double-Array Trie and its application[J]. Journal of Chinese Information Processing, 2006, 20(5):24-30.
[26] X F Zhu. Domain ontology-based software requirements inconsistency management[D]. 2007.
[27] Zhong-Bao M A and Guan-Rong L. Research on chinese text classification model based on SVM[J]. Computer Technology and Development, 2006.