

# System of Automatic Chinese Webpage Summarization Based on The Random Walk Algorithm of Dynamic Programming

Feng Wang<sup>\*1</sup>, XiaoMing Qin<sup>2</sup>, Yizhen Wang<sup>3</sup> and Xinjiang Wei<sup>1</sup>

<sup>1</sup>School of Mathematics and Statistics Science, Ludong University, Yantai 264025, China. Key Laboratory of Language Resource Development and Application of Shandong Province; Yantai 264025, China; <sup>2</sup>College of Computer and Information Engineering, Jiaozuo Teachers College, Jiaozuo 454000, China; Institute of Computer Application Technology, Jiaozuo Teachers College, Jiaozuo 454000, China; <sup>3</sup>Tilburg University, Tilburg, The Netherlands

**Abstract:** As the Internet becomes more and more deeply connected with our life, the Internet has brought together mass text material, and it is still in explosive growth. In order to quickly and accurately to help users find the required content, the traditional solution is to use a search engine. However, the results of existing automatic webpage summarization systems for search engine are of low quality. Because they just based on statistical method, gather some sentences in the web document beside the search phrases. Neither symbolizes the subject of the document, nor take into account the user search phrases. According to the shortages, An automatic webpage summarization systems is realized.

On the basis of the work done, this paper proposed an automatic text summarization method based on relation graph and text structure analysis. This method firstly segment text into semantic paragraphs. For each semantic paragraph, a subject term discover method based on relation graph analysis is proposed. At last, both search phrase and document subject are take into account, it extracts summary according to the guidance of the subject terms.

**Keywords:** Automatic summarization, random walk, semantic relatedness, seme based graph, webpage.

## 1. INTRODUCTION

As a basic issue in natural language process (NLP) [1], calculation of lexical semantic relevancy [2] is frequently involved in researches in this field to achieve success in semantic comprehension, classification and disambiguation. However, owing to the limited researches on semantic relevance of Chinese language and the statistical methods based on linguistic data which involves a large quantity of corpus as basis as well as long-term training which could be subject to the sparsity and imbalance of corpus, dissatisfaction may come as a result. As such, in this paper some sememe figures based on CNKI have been proposed and furthermore, based on sememe figures, some evaluation models of word-relevance are presented. To be specific to this model, a random walk algorithm based on dynamic programming is hereby proposed to calculate direct and indirect relevance between sememes. According to the experiment, the proposed methodology in this paper enjoys the following advantages: 1) The model proposed in this paper is more consistent with the cognition of human being; 2 ) It is highly related to the results of rating by human being ; 3) It is and easy to achieve. The new algorithm for lexical relevance has been applied in the automatic summarization system for web page in Chinese by the author. Compared to other algorithms based on methods by Q.Liu [3], the accuracy of summarization by this algorithm is greatly improved.

## 2. TECHNOLOGY FOUNDATION

### 2.1. Sememe and CNKI

Sememe which is one of concepts in Chinese Linguistics is the minimum semantic unit of word sense. Using sememe to describe word sense can turn complexity into simplicity and be convenient for formalization. In linguistics, each word may contain several sememe (transmission contains move, change, etc.) which is the minimum unit of language for independent use. While, the sememe is the semantic unit of semene. Sememe analysis is the most important and basic method for research on semantic, which description of word sense formalization is convenient for computer processing.

CNKI which uses sememe analysis to describe word sense is the most detailed and complete Chinese knowledge system. It names sememe as yiyuan which is the basic unit for CNKI to describe word sense, each word sense is described by several yiyuan. CNKI also uses more than 1600 yiyuan which are classified into 10 types. Each yiyuan type forms tree structure by hyponymy among yiyuan. This paper used CNKI as word sememe dictionary to form sememe probability network and calculate word relevance.

### 2.2. Random Walk Model Based on Sememe Figure

Random walk is a classic application of Markov Chain. All states and transition relationships among them in random walk can be regarded as a directed graph that the state is a point, transmission relationship is an edge with weight, the weight stands for transition probability. In each step of

random walk, there is the same probability when current statues transfer to next statues in no relation to previous walk path and transmission relationship [4]. This paper presented an extended random walk model which used particle to simulate human mind. When considered the situation of human evaluating correlation among words, people commonly would observe two words at the same time, and then consider the semantic differences between these two words. However, sememe analysis uses sememe to describe word semantic, the lexical correlation can be obtained by analyzing semantic relevant tightness among sememe.

Furthermore, we believe that the overlap ratio of two words sense can be reflected by finding quantity of direct or indirect semantic relation on average between two words within a certain time, and then this overlap ratio can be used to evaluate semantic correlation among words [5]. Based on this consideration, we presented an extended random walk model that two particles randomly walked at the same time along the edge on sememe figure formed on CNKI from two specific words. Because, the particle's each encounter case will form a path to connect two words, which stands for a type of semantic relation, thus semantic correlation among words can be evaluated by encounter probability among particles.

Firstly, the probability defined in this paper showd that a walking particle which begins to walk from  $n_j$  reaches  $n_i$  by passing an edge from any node.

$$n_{in_j}^{(t)} = \sum_{n_x \in V} n_{xn_j}^{(t-1)} P(n_i | n_j) \tag{2-1}$$

Secondly, calculated encounter probability of two particles separately beginning to walk from  $n_a$  and  $n_b$  ( $n_a$  and  $n_b$  included):  $P_{ENCOUNTER}(n_{y n_a}^{(x)} n_{y n_b}^{(t-x)})$

$$P(n_{y n_a}^{(x)} n_{y n_b}^{(t-x)}) = \sum_{n_y \in V} \sum_{m=1}^t n_{y n_a}^{(m)} n_{y n_b}^{(t-m)} \tag{2-2}$$

Secondly, calculated encounter probability of two particles beginning to walk separately from  $n_a$  and  $n_b$  ( $n_a$  and  $n_b$  included):  $P_{ENCOUNTER}(n_{y n_a}^{(x)} n_{y n_b}^{(t-x)})$

In addition, for the simulation of human correlation evaluation process, we induced a parameter t as random walk step limit for semantic contact path length between two words as a reason that no semantic correlation existed nodes which were far from beginning words and beyond limit of human analysis ability.

After the consideration above, we introduced  $P_{before}(n_a, n_b, t)$  which stands for the walk encounter probability of two particles walking separately from  $n_a$  and  $n_b$  with the step less than t.

By method of introducing calculation of two particles encounter probability, this paper not only avoided that there still needed to use vector difference measurement algorithm with indistinct implication when we should calculate only a particle probability's stationary distribution, but also formed a semantic relevance calculation model based on human intelligence simulation.

### 2.3. Sememe Formation

This paper abstracted sense and sememe from CNKI as node, explanation by sense for sememe as semantic contact among nodes formed sememe figure with weight.

Sememe node: each sememe of CNKI corresponded to each node in sememe figure. For example, a node in figure corresponds to sememe '罪' (means crime) in CNKI.

Sense node stands for node of lexical sense. For example, '罪犯(zuifan)#1' stands for the first sense of '罪犯(zuifan)' (word criminal) which means people who commit a crime.

Edge from sememe to sense stands for explanation relationship for sense by sememe, for example, three sememe of '罪'、'人'、'莠' (crime, human and bad people (or weeds)) are used to explain sense '罪犯(zuifan) #1'.

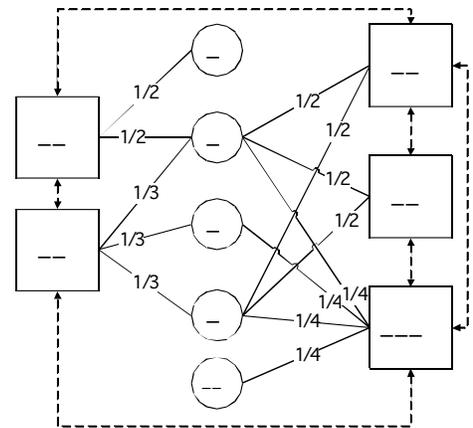


Fig. (1). Sememe graph.

Fig. (1) shows parts of nodes and edges highly related to two words of crime and offense in the complete sememe figure.

Above all, the sense figure formed in this paper contains 1618 sememe nodes, 66181 sense nodes, 397086 edges. Aware of probably existing edges between sememe and sense in sememe figure, thus the finally formed sememe figure is very sparse.

### 3. IMPROVED RANDOM WALK ALGORITHM

Edge in sememe figure can be showed by transition matrix  $E \in \tau^{N \times N}$ . Inside, N stands for total edge quantity in sememe figure. Element  $E_{[i,j]}$  shows condition probability of transition from node  $n_i$  to  $n_j$ . As a reason, there is no effect information to distinguish importance among different sememe, therefore all edges are equally treated in this paper. With more specifics, probabilities of transition from a spe-

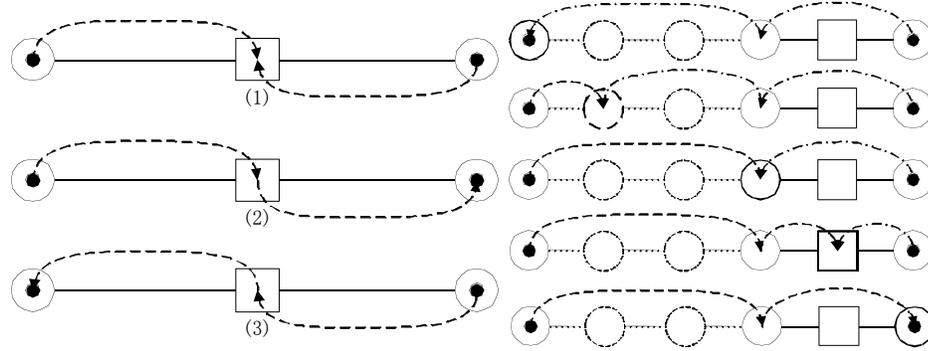


Fig. (2). 3 encounter modes and 5 transition situations with step=1 (left).

cific node to different adjacent nodes are same. Calculation method of element  $E_{[i,j]}$ , as follows:

$$E_{[i,j]} = outdeg(n_i)^{-1} \quad (3-1)$$

$outdeg(n_i)$  means out-degree of  $n_i$ .

A basic calculation can be obtained from the beginning of formula 2-1: calculate  $n_{in_j}^{(t)}$  for all needed  $n_i$ 、 $n_j$ 、 $t$  and then calculate encounter probability using formula 2-2, finally obtain the result using formula 3. However, this calculation's time complexity is quite high to  $O(n^3 * t^2)$  inside  $n$  means total node quantity,  $t$  means time parameter in formula 2-3.

Be noticed of sparse matrix  $E$  and any path connecting two sense node passes sememe node which quantity is low, thus early dealing with sememe node will reduce the scale of issues. Therefore, this paper induced sememe encounter probability matrix  $S[mode, n_i, n_j, d](n_i, n_j \in V_{sememe})$  which shows the encounter probability of two particles separately walking from the beginning of node  $n_i$  and node  $n_j$  after  $d$  steps. Mode means two particles encounter modes which have three types: 1) encounter in a middle sense node after beginning to walk at the same time; 2) encounter at the second particle's beginning position with the situation that the first particle moves and the second stay where it is; 3) instead of 2) encounter at the second particle's beginning position with the situation that the second particle moves and the first stay where it is. Be aware that the shortest path between two sememe nodes are formed by two edges: one is from sememe node to sense node, the other is from sense node to sememe node. This situation is named as step=1, as shown in Fig. (2) left. The encounter probability with step=1 can be calculated using following formula:

$$S[1, n_i, n_j, 1] = |outdeg(n_i) \cap outdeg(n_j)|^{-1} \quad (3-2)$$

$$S[2, n_i, n_j, 1] = outdeg(n_i)^{-1} \quad (3-3)$$

$$S[3, n_i, n_j, 1] = outdeg(n_j)^{-1} \quad (3-4)$$

After analysis, matrix  $s$  calculation is consistent with overlapping sub-problems and optimal substructure, so the dynamic programming can be used to optimize calculation. For this issue, there are five different existing states transition situation, as shown in Fig. (2) (right). These five situations can be induced into three encounter modes ( $n_i, n_j, n_k \in V_{sememe}$ ):

$$S[1, n_i, n_j, l] = \sum_{n_k} \max( \quad (3-5)$$

$$S[1, n_i, n_k, l-1] * S[2, n_k, n_j, 1],$$

$$S[3, n_i, n_k, l-1] * S[2, n_k, n_j, 1],$$

$$S[3, n_i, n_k, l-1] * S[1, n_k, n_j, 1])$$

$$S[2, n_i, n_j, l] = \sum_{n_k} S[2, n_i, n_k, l-1] * S[2, n_k, n_j, 1] \quad (3-6)$$

$$S[3, n_i, n_j, l] = \sum_{n_k} S[3, n_i, n_k, l-1] * S[3, n_k, n_j, 1] \quad (3-7)$$

The time complexity of stated state transition formula above is the square of sememe node quantity which leads to an improved efficiency. After adding of walk step limit  $t$ , we obtained a formula with only consideration on sememe  $P_{before}(n_a, n_b, t)(n_a, n_b \in V_{sememe})$ :

$$P_s(1, n_a, n_b, t) = \sum_{d=1}^t S[1, n_a, n_b, d]$$

$$P_s(2, n_a, n_b, t) = \sum_{d=1}^t S[2, n_a, n_b, d] \quad (3-8)$$

$$P_s(3, n_a, n_b, t) = \sum_{d=1}^t S[3, n_a, n_b, d]$$

When  $P_{before}(n_a, n_b, t)(n_a, n_b \in V_{sememe})$  is obtained, the encounter probability  $P_{before}(n_a, n_b, t)(n_a, n_b \in V_{sememe})$  among senses can be calculated by following formula:

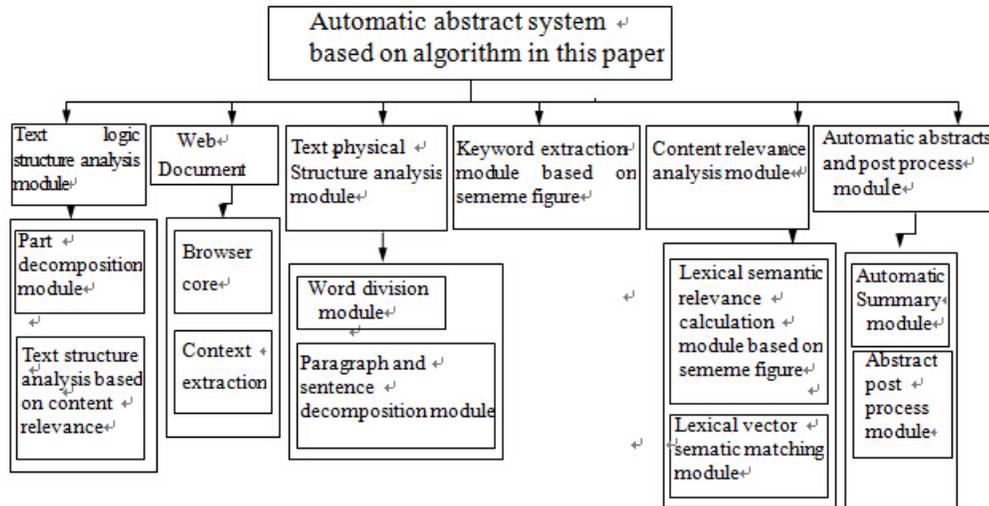


Fig. (3). System main function modules.

$$P_{before}(n_a, n_b, t) = \sum_{(n_a, n_i) \in E} \sum_{(n_b, n_j) \in E} \max($$

$$outdeg(n_b)^{-1} * P_s(2, n_i, n_j, t) * outdeg(n_i)^{-1},$$

$$outdeg(n_b)^{-1} * P_s(2, n_i, n_j, t) * outdeg(n_a)^{-1}, \quad (3-9)$$

$$outdeg(n_b)^{-1} * P_s(1, n_i, n_j, t) * outdeg(n_a)^{-1},$$

$$outdeg(n_b)^{-1} * P_s(3, n_i, n_j, t) * outdeg(n_a)^{-1},$$

$$outdeg(n_a)^{-1} * P_s(2, n_i, n_j, t) * outdeg(n_j)^{-1})$$

So far, the needed encounter probability among senses are obtained, the calculation complexity of this algorithm is  $O(N^2 * t)$ .

We can evaluate semantic relevance between senses using this encounter probability, as follows:

$$rela(n_a, n_b) = P_{before}(n_a, n_b, t) \quad (3-10)$$

#### 4. CHINESE PAGE AUTOMATIC SUMMARY SYSTEM DESIGN

##### 4.1. System Main Function Modules

As shown in Fig. (3), this system are made from six modules, such as, text logic structure analysis module, WEB document summary module, text physical structure analysis module, keyword extraction module based on sememe figure, content relevance analysis module, automatic summaries and post process module.

- (1) Text logic structure analysis module: Recognize sub-head, and divide document into several sections by sub-head.
- (2) WEB document summary module: Read webpages by external input URL address, render webpages using browser core and summary page text using text summary method based on vision analysis.

- (3) Text physical structure analysis module: Analysis (words, sentences and paragraph are included) physical structure in webpage text.
- (4) Keyword extraction module based on sememe figure: Extract keywords vectors from semantic paragraph.
- (5) Content relevance analysis module: Calculate relevance among words and sentences to provide content relevance support for other modules.
- (6) Automatic summary and post process module: Using algorithm in this paper, extract summaries from document and output to external caller after simple processing.

Processing figure of each module in automatic summary system, as shown in following Fig. (4).

As shown in Fig. (4), this system is formed by two parts of webpage extraction module and automatic summary module. By external input URL address, webpage pre-processing module reads, parses and renders web pages, also output webpages topic related content and webpages text after Web vision and label analysis, Chinese text characteristics analysis. Webpages topic related content filter webpages noise and is outputted to search engine module to index. Webpages text is outputted to automatic summary module for post automatic summary processing.

Automatic summary module are formed by six modules, such as, semantic knowledge base module, text pre-processing module, lexical relevance calculation module, semantic paragraph division module, automatic summary and post processing module, Hadoop storage interface. Semantic knowledge base module stores general vocabulary,, field vocabulary, CNKI knowledge base, webpages summary knowledge base to provide support for other modules. Hadoop is responsible for Caching intermediate results to improve speed of automatic summary.

##### 4.2. System Design

For accomplishment of dynamically automatic summary which is divided into two sub-processes as automatic sum-

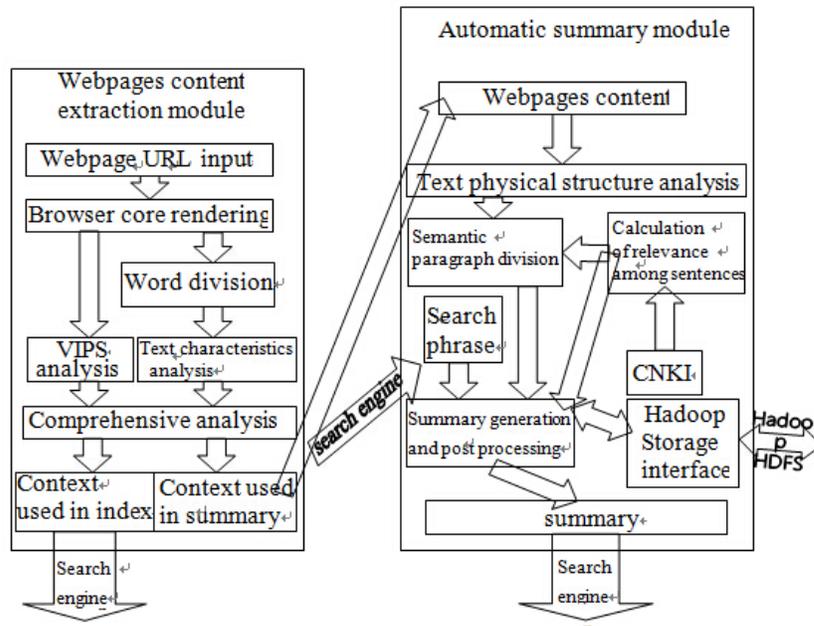


Fig. (4). System main function module.

mary pre-processing and dynamically automatic summary process.

After the division of webpages text semantic paragraph, intermediate results will be cached into Hadoop platform through Hadoop storage interface. When user inputs search words, the search engine will transmit two parameters of webpages ID and search Query to automatic summary module. Through Hadoop storage interface, the automatic summary module takes out intermediate results corresponding to this webpages ID for directly outputting summary after sentences choice and post processing module.

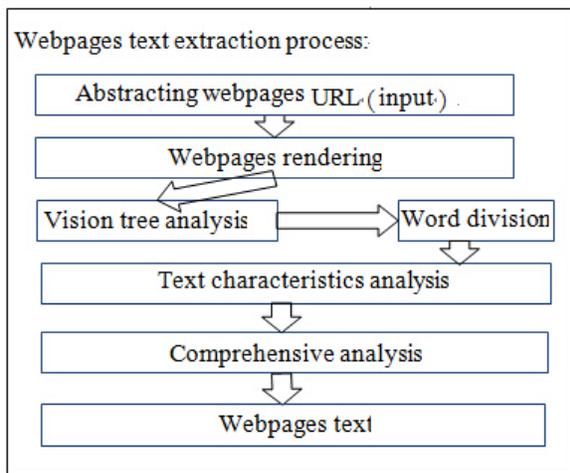


Fig. (5). Web pages text extraction module processes.

As shown in Fig. (5), webpages context extraction process, as follows:

- 1) Read input URL corresponding to HTML files;
- 2) Process HTML label and render this webpage;
- 3) Analyze webpages by vision tree and each section's position and area;

- 4) Divide words in each section and analyze text characteristics.
- 5) Totally consider the results of vision tree analysis and text characteristics analysis.
- 6) Extract out webpages topic related context and webpages context.

As shown in Fig. (6), automatic summary process, as follows:

- 1) Obtain web pages context from web pages pre-processing module and calculate the unique ID corresponding to this webpage;
- 2) Mark in words division and words speech;
- 3) Identify title and divide sentence, paragraph, subhead;
- 4) Divide semantic paragraph using text structure analysis algorithm. Lexical relevance calculation module provides support of lexical relevance calculation.
- 5) Intermeddle results after semantic paragraph division will be cached on Hadoop platform through Hadoop storage interface with taking this webpage unique ID as recognition.

Dynamically automatic summary process, as follows:

- 1) Read intermeddle results through Hadoop storage interface using obtained webpages unique ID;
- 2) Scores for sentences using the method based on the combining noumenon with TF-IDF, summary sentences with certain proportion from each semantic paragraph to form summaries;
- 3) Post processing for summaries and filtration for sentences with repeated semantic to improve accuracy and enhance readability, and so on

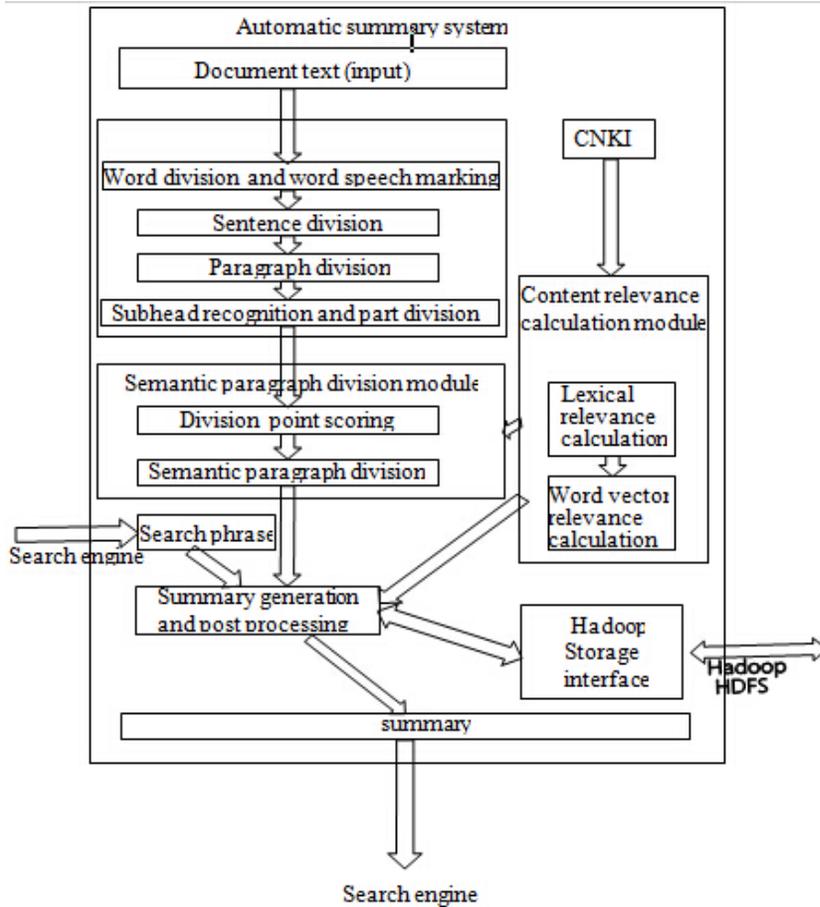


Fig. (6). Automatic summary module process.

4.3. System Interface

As shown in Table 1.

(1) Get article text

Function: Get article text for single document webpages, for example, single document webpages, return empty string.

(2) Get webpages summary interface

Function: Get webpages summary, and outputted summary length isn't over max length.

(3) Automatic summary pre-processing interface

Function: Pre-process automatic summary through this interface informing automatic summary module when webpages extraction by search engine, pre-processed webpages can call interface 5 for fast dynamic summary.

(4) Get fast dynamic summary interface

Function: Fast get dynamic webpages summary which needs to input pre-processed webpages ID, outputted summary length isn't over max length, and empty string will return if webpages corresponding to this ID isn't pre-processed by interface 4.

Interface's calling method:

- (1) Web service (JSON/XML)
- (2) C# DLL

Table 1. Interface definition.

(1-1)

Interface Name	Get Article Text
Interface input parameter (type, variable name)	String, URL
Interface output parameter (type)	String

(1-2)

Interface Name	Get Summary
Interface input parameter 1 (type, variable name)	String, URL
Interface input parameter 2 (type, variable name)	String, MaxLength
Interface output parameter (type)	String

(1-3)

Interface Name	Process Summary
Interface input parameter 1 (type, variable name)	String, URL
Interface input parameter 2 (type, variable name)	String, ID
Interface output parameter (type)	Int

Table 1. Contd.....

(1-4)

Interface Name	Get Dynamic Summary
Interface input parameter 1 (type, variable name)	String, ID
Interface input parameter 2 (type, variable name)	String, MaxLength
Interface output parameter (type)	String

**5. EXPERIMENT FORMATION AND RESULT ANALYSIS**

This paper evaluates through calculation of accuracy rate and recall rate by comparing with automatic summary and human ideal summary. For summary mission, only the right meaning is everything instead of strictly comparing whether the summary generated by system is consistent with experts summary, which is too harsh. However, it's very hard for summary formed by human to reach uniqueness. As description of the same thing in different ways, user also can form many different common summary or the acceptable focusing-on user summary supposed by them. Actually, the experiment shows that it's hard to be consistent with the issue which sentences or paragraphs can be included in a summary. Even a same summary expert, for him, there are most different among the summaries made by him in a same article after an interval time.

Thus, this paper presented a new evaluation strategies: the automatic summary is accurate if automatic summary is consistent with sub-topic of an article covered by human summary. For example, human summary and automatic summary choose two different sentences with close meaning which are treated accurate.

On detailed calculation method, totally evaluated summary accuracy using accuracy rate, recall rate and F value.

We used search engine crawlers technology to extract 2000 news pages with different style as test corpus. For avoiding the article as an important factor with too long or too short context to influent evaluation results, this paper only consider the general news with middle length that chosen web pages with 30 sentences context for experiment. To each article, the system generates automatic summary with 0.1 Compression ratio firstly, and then extract 10% context as human ideal summary. When human extraction, prior chosen sentences which have been chosen by automatic summary system, and compared summary generated by system with human summary. This paper evaluated summary in

the view of information content, using three important indexes in the field of information retrieval: accuracy rate, recall rate, totally accuracy rate and F value for recall rate.

We wrote an automatic summary automated experiment software to assist experiment, and after experiments among 2000 articles, improved algorithm's average accuracy rate to 0.502, average recall rate to 0.853, average F value to 0.727. part of experiment results, as shown in Table 2:

Through experiment, we found that the accuracy and recall rate of improved algorithm are better. While, we could find the high recall rate (85.3%) of improved algorithm, which are obviously higher than accuracy rate corresponding to the design purpose of algorithm in this paper.

**6. SEMANTIC RELEVANCE EVALUATION**

There are two classic method for evaluation of semantic relevance: one is the relevance suitable with human evaluation, the other is performance of evaluation algorithm in specific application.

This paper used two relevance evaluation method above which contains comparsion with a group of human evaluation data, and applied this algorithm into an summary system, by which observed the influences on summary. However, researchers didn't agree on how to quantify semantic relevance, this paper evaluated using spearman's p coefficient. When comparsion of these two relevance results, spearman's p coefficient only considered relative ranking of relevance value. Presently, WordSimilarity-353(WS-353) is a common human evaluation dataset of English words relevance, this research invited many participants to score with word pair relevance. For the reason that no Chinese dataset with high quality, so we chose subset (100 words) of WS-353, and translated word pair into Chinese which was called CWord-100. The principle of choosing word pair is that the English word included in word pair can be translated into Chinese words with the same meaning, and there are accurate and efficient sememe descriptions in CNKI.

The writer wrote a lexical relevance calculation software package that had improved the algorithm. Taking several classic algorithm as comparsion baseline, we calculated relevance value of word pair in CWord-100, and compared with human evaluation to get the sorting relevance which is suitable for comparsion, the results are as shown in Table 3:

Through experiment, compared with other algorithm, we found that improved algorithm could provide relevance value higher related with human evaluation. Some classic word pair were chosen in Table 4 to compare with popular method of Q. Liu [18].

Table 2. Improved algorithm results.

Improved Algorithm	1	2	3	4	5
accuracy rate	0.4	0.65	0.5	0.7	0.6
	3	5	2	1	5
recall rate	0.86	0.85	0.9	0.9	0.9
	6	5	2	0	5

**Table 3. Comparison experiment among several classic method (based on CWord-100 test set).**

Model	CWord-100
algorithm in this paper	0.803
T. Hughes [7]	0.799
Q. Liu [3]	0.654
X. Yun [6]	0.677
Z. Shuqin [8]	0.730

**Table 4. Use method in this paper and method of Q. Liu [3] to calculate relevance of classic word pair.**

Word Pair	Human Evaluation	Q. Liu [3]	Improved Algorithm
老虎/老虎	1.000	1.000	1.000
电脑/键盘	0.762	0.005	0.207
电话/通讯	0.750	0.152	0.507
书/纸张	0.746	0.005	0.207
媒体/广播	0.742	0.444	0.255
医生/护士	0.700	0.948	0.270
犯罪/罪犯	0.652	0.044	0.226
国王/卷菜	0.023	0.171	0.003
火车/汽车	0.631	1.000	0.500
面包/黄油	0.619	0.204	0.560
夏天/白天	0.394	0.722	0.326
周三/新闻	0.222	0.112	0.001

This text automatic summary system once used relevance evaluation method of Q. Liu, however, which calculated a wrong relevance value, and effected the accuracy of system summary. After changing Q. Liu method to method in this paper, system summary accuracy was improved visibly. the results are as shown in Table 5:

## 7. SUMMARY

This paper has researched totally on Chinese webpage automatic summary technology, and presented a lexical semantic relevance algorithm based on CNKI knowledge and

**Table 5. Influence on text automatic summary system accuracy using improved algorithm and Q. Liu method.**

Use	Degree of Accuracy
improved algorithm	0.705
Q. Liu [3]	0.683

computational semantics which calculate direct and indirect relevance between sememe using improved random walk algorithm. To be different with existed relevance measure algorithm using random walk model, this paper presented to use average encounter probability instead of average arriving probability, which would be in accordance with relevance recognition for human, and also avoided to use vector difference measure algorithm with blurry implication. However, there still are some disadvantages, such as, the generated summaries are not smooth and fluent. Thus, more summary post processing technologies will be added in to improve summary readability.

## CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

## ACKNOWLEDGEMENTS

This research is supported in part by National Science Foundation of China No. 61374108.

## REFERENCES

- [1] S. Banerjee, and T. Pedersen, "Extended gloss overlaps as a measure of semantic relatedness," *IJCAI*, vol. 3, pp. 805-810, 2003.
- [2] E. Gabrilovich, S. Markovitch, "Computing semantic relatedness using wikipedia-based explicit semantic analysis," *IJCAI*, vol. 7, pp.1606-1611, 2007.
- [3] Q. Liu, and S. Li, "Word similarity computing based on how-net," *Computational Linguistics and Chinese Language Processing*, vol. 7, no. 2, pp. 59-76, 2002.
- [4] P. Berkhin, "A survey on pagerank computing," *Internet Mathematics*, vol. 2, no. 1, pp. 73-120, 2005.
- [5] W. Yi and W. Xiaolin, "Algorithm for words' semantic relevancy based on modified algorithm for sememes' relevancy," *Jouranal of The China Society For Scientific And Technical Information*, vol. 31, no. 12, pp. 1271-1275, 2012.
- [6] X. Yun Xu and F. F. Zhang. Semantic relevancy computing based on hownet, *Transactions of Beijing Institute of Technology*, vol. 25, no. 5, pp. 411-414, 2005.
- [7] T. Hughes, D. Ramage, *Lexical Semantic Relatedness with Random Graph Walks*, EMNLP-CoNLL, 2007, pp. 581-589.
- [8] Z. Shuqin and W. Yangyang. The model of words relation computing based on the HowNet, *Microcomputer and ITS Applications*, vol. 31, no. 8, pp. 77-80, 2012.