

Article

# Method for Retrieving Digital Agricultural Text Information Based on Local Matching

Yue Song <sup>1,2,\*</sup>, Minjuan Wang <sup>1,2</sup> and Wanlin Gao <sup>1,2</sup>

<sup>1</sup> College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China; minjuan@cau.edu.cn (M.W.); gaowlin@cau.edu.cn (W.G.)

<sup>2</sup> Key Laboratory of Agricultural Informationization Standardization, Ministry of Agriculture, Beijing 100083, China

\* Correspondence: b20153080285@cau.edu.cn

Received: 4 June 2020; Accepted: 23 June 2020; Published: 2 July 2020



**Abstract:** In order to improve the retrieval results of digital agricultural text information and improve the efficiency of retrieval, the method for searching digital agricultural text information based on local matching is proposed. The agricultural text tree and the query tree are constructed to generate the relationship of ancestor–descendant in the query and map it to the agricultural text. According to the retrieval method of the local matching, the vector retrieval method is used to calculate the digital agricultural text and submit the similarity between the queries. The similarity is sorted from large to small so that the agricultural text tree can output digital agricultural text information in turn. In the case of adding interference information, the recall rate and precision rate of the proposed method are above 99.5%; the average retrieval time is between 4s and 6s, and the average retrieval efficiency is above 99%. The proposed method is more efficient in information retrieval and can obtain comprehensive and accurate search results, which can be used for the rapid retrieval of digital agricultural text information.

**Keywords:** local matching; digitization; agricultural text information; recall rate; precision rate; retrieval

## 1. Introduction

With the rapid development of digital agricultural information management technology, intelligent agricultural information technology has become the research hotspot in agricultural informationization. Digital agricultural texts have more and more information. How to obtain comprehensive and accurate information in massive information becomes a problem [1]. Agricultural information resources on the Internet are dynamic and unstable [2]. Understanding the characteristics of agricultural network information is beneficial for workers to purposefully collect and manage Internet information resources.

The characteristics of agricultural network information are: first, the content of information is diversified. Second, agricultural network information resources are large-scale and widely distributed. Thirdly, information has great freedom and arbitrariness. This brings great inconvenience to users when using valuable network information resources. Fourth, the distribution and composition of agricultural information lack structure and organization, which increases the difficulty of information resource management and retrieval. Many scholars have studied this. For example, Hu Yitao et al. [3] constructed a set of complete metadata standards for human object digital resources aiming at agricultural cultural heritage digital information resource management and value representation mining, which proved the advantages of metadata method. Y et al. [4] put forward new requirements and challenges for information resources in response to the arrival of the era of big data, and made their own contributions to the innovation of information resource management in the era of big data;

Wang Lijun [5] proposed a text information retrieval method based on feature clustering to improve the accuracy of text information retrieval.

Faced with the agricultural information in the existing distributed and complex network environment, the limitations of traditional retrieval methods are becoming clearer [6]. On the one hand, information overload brings about the problem of low recall rate and low precision. Users must spend a lot of time and effort to filter out the information they want [7]. On the other hand, information retrieval lacks intelligence [8]. In view of these problems, local matching method can accurately screen the required information and is more accurate. Therefore, this paper proposes a digital agricultural text information retrieval method based on local matching. Construct the agricultural text tree and query tree, generate the relationship between ancestors and descendants in the query, and map it to the agricultural text. According to the local matching retrieval method, the vector retrieval method is used to calculate the similarity between digital agricultural texts and submit queries. The similarity is sorted from large to small, so that the agricultural text tree outputs the digital agricultural text information in turn [9]. This method has high retrieval efficiency, comprehensive and accurate retrieval results, and can be used for fast retrieval of digital agricultural text information.

## 2. Material Methods

### 2.1. Method for Retrieving Digital Agricultural Text Information Based on Local Matching

#### 2.1.1. Construction of Digital Agricultural Text Tree and Query Tree

Describe the digital agricultural text information as a tree called the digital agriculture text tree [10] which is described by the Formula (1):

$$d = (r, SN, TN, T, \langle_{snh}, \tau, \sigma) \quad (1)$$

where  $r$  is the virtual root node in the tree, which represents the entire text;  $SN$  is the collection of structural nodes in the tree;  $TN$  is the collection of text nodes in the tree;  $T$  is the collection of types of all structural nodes in the tree;  $\langle_{snh}$  describes the parent–child relationship between structural nodes in the tree.

Assume that  $sn_i, sn_j \in SN$ , and  $sn_i$  is a child of  $sn_j$ , then  $sn_i \langle_{snh} sn_j$ ;  $\tau$  is the mapping from  $SN$  to  $T$ ;  $\sigma$  is the mapping from  $SN$  to  $TN \cup \{NULL\}$ .

Defining the path in the agricultural text tree  $p = (sn_1, sn_2, sn_3, \dots, sn_m)$  represents a path between the structural node  $sn_1$  and the structural node  $sn_m$ , which describes the ancestor–descendant relationship between  $sn_1$  and  $sn_m$ ;  $head(p)$  and  $tail(p)$  represent the start and end points of the path respectively; the distance between the nodes is defined as  $dist(x, y) = |p| - 1$ , where  $head(p) = x$ ,  $tail(p) = y$ , and  $|p|$  represents the number of nodes in the path.

In the agricultural text tree, the set of descendant nodes of the structural node is described by Equation (2):

$$DESC(sn) = \{sn_i | \exists p((head(p) = sn) \wedge (tail(p) = sn_i))\} \quad (2)$$

The text information set of the structure node is described by Formula (3):

$$CONTENT(sn) = \left\{ \begin{array}{l} tn_j | (sn_i \in DESC(sn) \cup \{sn\}) \wedge (tn_j = \sigma(sn_i)) \\ \wedge (\sigma(sn_i) \neq NULL) \end{array} \right\} \quad (3)$$

The subtext is defined as  $sd_n = (n, SN_n, TN_n)$ , where  $n \in SN$ ,  $SN_n = DESC(N)$ , and  $TN_n = \{tn | (tn = \sigma(sn)) \wedge (sn \in SN_n)\}$ . That is, the subtext  $sd_n$  is the subtree with the node  $n$  as the root in the text tree, and the type of the subtext is the same as the type of the node  $n$ ,  $TYPE(sd) = \tau(n)$ ; the content of the subtext is  $CONTENT(sd_n) = CONTENT(n)$ .

Similar to the description of the agricultural text, the submitted query containing the structural information is described as a query tree [11], each node of the query tree is a subquery, and there must be the description of the text information in the submitted query.

### 2.1.2. Agricultural Text Information Retrieval

In the process of searching agricultural text information using the method of this paper, each sub-query on the query tree is processed step by step from bottom to top. At the same time, the vector search is used to calculate the similarity between the sub-query and the sub-text, and finally, the similarity between the text and the whole query is obtained [12]. According to the ranking of the similarity, the retrieval result of the agricultural text information is output. The detailed search process is as follows:

First, the vector of the subquery  $q$  is  $q = (sq_1, sq_2, sq_3, \dots, sq_m)$ , and  $sq_i \in DESC(q)$ ; the vector representation of the corresponding subtext is, where  $result(sq)$  indicates the matching result that the search of the subquery  $sq$  should satisfy, which is described by Formula (4):

$$result(sq) = \left\{ n \mid \begin{array}{l} (n \in SN) \wedge ((\tau(n) = sq) \vee (sq \in CONTENT(n))) \\ \wedge relD(n, sq) \end{array} \right\} \quad (4)$$

where  $relD(n, sq)$  means the structural characteristics  $n$  should satisfy the following relationship, namely Formula (5):

$$relD(n, sq) = \exists n_d \exists sq_d \left( \begin{array}{l} (n_d \in DESC(n)) \wedge (sq_d \in DESC(sq)) \\ \wedge (n_d \in result(sq_d)) \end{array} \right) \quad (5)$$

As can be seen from  $result(sq)$ , the retrieved result text tree may not completely contain the query tree [13], which is the incomplete matching process, that is a local matching process. Therefore, as long as the relationship between the nodes  $n$  satisfies the ancestor–descendant in any sub-query, it is considered to satisfy the query requirements, thus implementing the process of local matching [14].

In the local matching retrieval process, each method needs to construct the corresponding result sub-text in the order of the root traversal in the text tree, and calculate the similarity between the sub-query and the sub-text [15]. Here, the similarity function is defined as  $SIM(sd, q)$ , which represents the similarity between the subtext  $sd$  and its matching sub-query. In order to calculate the similarity function, the value of the elements in the vector need to be further determined [16].

When performing local matching, the value of each element of the query vector should be determined according to the distance between the query  $q$  and its descendants [17]. The value of this paper is the reciprocal of the distance  $w_{q_i} = QW(dist(q, sq_i)) = 1/dist(q, sq_i)$ , which means that the greater the distance between the queries is, the lower the degree of correlation with the user query is.

Assume that  $w_{result(sq_i)}$  is the value of each element of the subtext vector on sub-query  $sq_i$ , and the subtext vector can be described by Equation (6):

$$sd = (w_{result(sq_1)}, w_{result(sq_2)}, \dots, w_{result(sq_m)}) \quad (6)$$

Using the traditional text analysis to find the value of each element of the subtext vector [18], according to the particularity of the agricultural text, use Formula (7) to calculate the weight value of the keyword sub-query in the local matching process of agricultural information:

$$w_{result(qt)} = \frac{f_{qt}}{|CONTENT(sd_{sup})|} \log \frac{n_{sup}(qt)}{N_{sup}} \quad (7)$$

where  $sd_{sup}$  is the subtext containing the keyword  $qt$  and satisfying the previous sub-query;  $f_{qt}$  is the frequency of occurrence of the keyword  $qt$  in the text information set  $CONTENT(sd_{sup})$ ;

$|CONTENT(sd_{sup})|$  means the total length of the text information contained in the subtext  $sd_{sup}$ ;  $n_{sup(qt)}$  is the number of subtexts that contain the keyword  $qt$  and matches the query  $q_{sup}$ ;  $N_{sup}$  is the number of subtexts for all matching queries  $q_{sup}$ .

Finally, according to the cosine matching coefficient method, the similarity between the subtext and the sub-query in the local matching process [19] is obtained:

$$SIM(sd, q) = \sum_{i=1}^n w_{qi} w_{result(qi)} / \sqrt{\sum_{i=1}^n w_{qi}^2 \sum_{i=1}^n w_{result(qi)}^2} \quad (8)$$

In the process of processing the sub-query from bottom to top, the similarity between the digital agricultural text and the submitted query is obtained by the above method. In order of the similarity, the agricultural text tree sequentially outputs digital agricultural text information [20]. According to the above analysis, it can be known that this method can retrieve the digital agricultural text information more accurately than the previous method, which improves the disadvantages of the previous method that the retrieval is not timely and accurate, and improves the work efficiency.

## 2.2. Experimental Materials

In order to verify the retrieval effect of this method, this paper carries out the experiment of digital agricultural text information retrieval, using Windows 7 computer and Xapian software, on the small-scale data set (set a) and large-scale data set (set B). The set A is randomly selected from the agricultural WebPages that have been classified, including 2000 agricultural texts in four aspects: agricultural science and technology, agricultural new information, agricultural products and agricultural development, which are mainly used to evaluate the retrieval effect of the model, including precision and recall rate. The set B is extracted from the 1.2 million web pages captured by the agricultural search engine, and is mainly used to evaluate the retrieval efficiency of the model.

### 2.2.1. Experimental Setup of Recall Rate and Precision Rate

The method based on double semantic space and the method based on maximum weight matching calculation are compared with the method in this paper. The specific experimental setup is as follows: the test is performed on the test set A, when the amount of data in the test set is 100, 200, 300, 400, and 500, the recall rates of the three methods are measured separately; and the set A is used to verify the precision of the three methods [21]. The test is carried out 5 times.

In order to highlight the advantages of this method, 10%, 20%, and 30% of comprehensive news information is added to the test set A to interfere with the retrieval. The above experiment is re-executed and the results were recorded.

### 2.2.2. Experimental Setup for Retrieval Efficiency

Three methods are used to retrieve 200,000 pieces of text information in test set B, and the retrieval time of the three methods is recorded.

Because the agricultural information retrieval method based on the maximum weight matching calculation takes less time than the dual semantic space method, in order to further verify that the proposed method has the advantage of low time, the former method is used again to compare with the method [22]. The specific experiment setting is as follows: the experiment is divided into 5 times, each time the number of test set pages is 50,000, 100,000, 150,000, 200,000, 250,000, and the time-consuming situation of the two methods is recorded.

The test set B is also used for the efficiency comparison experiment, and the experiment is carried out in three stages. The first stage: 400,000 agricultural web pages are retrieved; the second stage: 800,000 agricultural web pages are retrieved; the third stage: 1.2 million agricultural web pages are retrieved.

### 3. Results

#### 3.1. Comparative Test of Retrieval Effect

The recall rate of the three methods is shown in Figure 1.

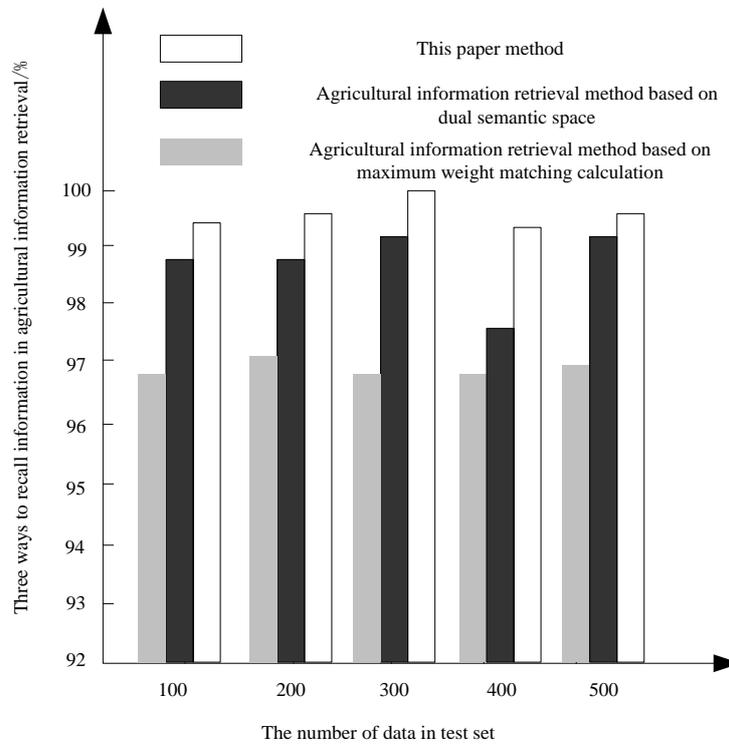


Figure 1. The recall rate of agricultural information retrieval by three methods.

#### 3.1.1. Recall Rate

Interference information is added on the basis of test set A. Three methods are used to test the recall rate, and the obtained comparison results are described in Figures 2–4.

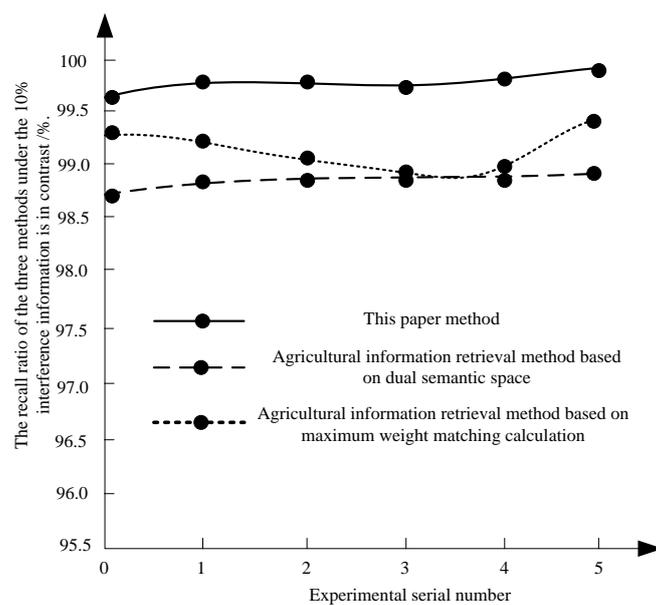


Figure 2. Comparison of recall rates of three methods with 10% interference information added.

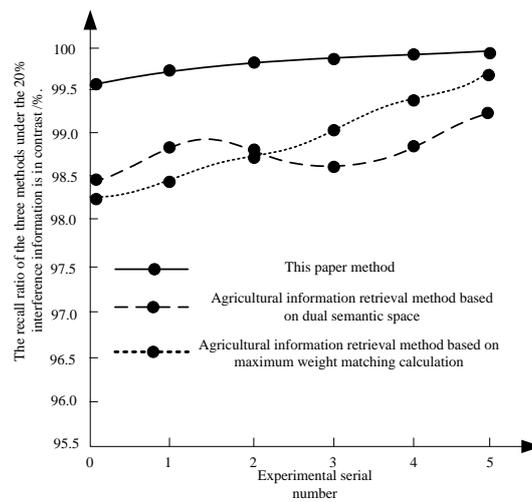


Figure 3. Comparison of recall rates of three methods with 20% interference information added.

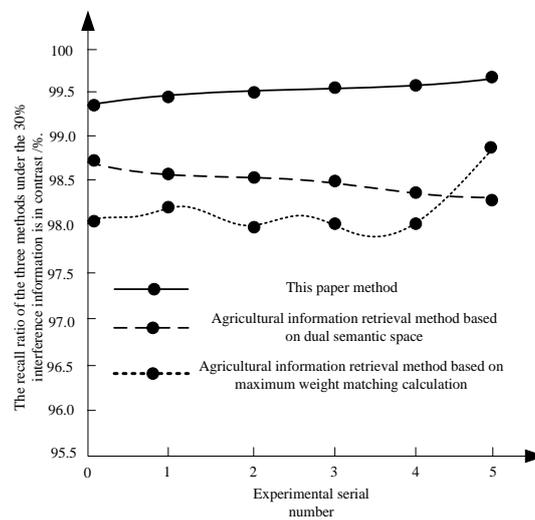


Figure 4. Comparison of recall rates of three methods with 30% interference information added.

3.1.2. Precision Rate

The precision ratios of the three methods are described in Table 1, Table 2, and Table 3, respectively.

Table 1. Precision rate of agricultural information retrieval method based on maximum weight matching calculation.

Test Set Number/Individual	Test Times				
	First Test/%	Second Test/%	Third Test/%	Fourth Test/%	Fifth Test/%
50	94.4	94.5	93.4	94.5	94.6
100	94.5	94.3	93.4	94.3	94.2
150	94.6	93.2	94.5	93.2	94.5
200	94.2	93.9	94.2	93.9	94.8
250	94.5	93.7	93.5	93.5	93.8
300	94.8	93.8	93.5	94.8	94.8
350	94.8	94.8	94.8	93.8	94.9
400	94.8	94.9	93.8	94.1	93.4
450	93.5	93.9	94.1	93.8	93.4
500	94.9	94.7	93.8	93.8	94.5
Precision ratio/%	94.5	94.1	94.0	94.0	94.3

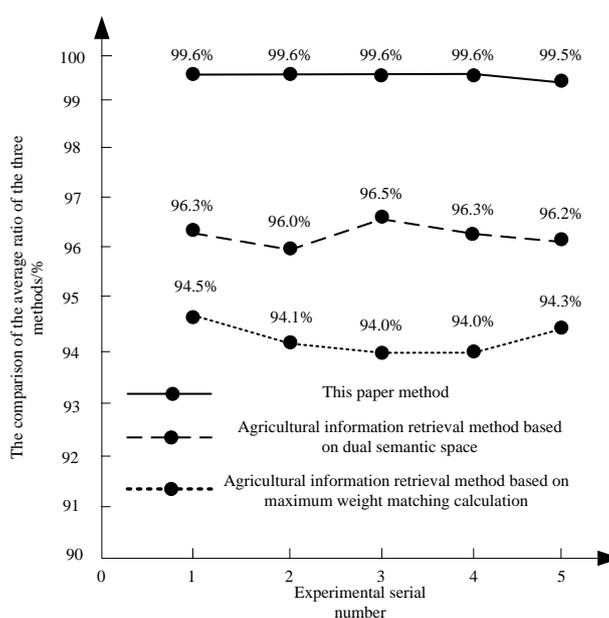
**Table 2.** Accuracy of agricultural information retrieval methods based on dual semantic space.

Test Set Number/Individual	Test Times				
	First Test/%	Second Test/%	Third Test/%	Fourth Test/%	Fifth Test/%
50	95.8	95.4	96.8	96.5	96.5
100	96.2	95.6	97.1	96.8	96.6
150	96.8	96.8	97.2	96.7	95.6
200	96.7	96.5	96.2	95.9	95.8
250	95.9	96.6	96.4	95.8	96.5
300	95.8	95.6	97.1	95.6	95.8
350	96.8	95.8	96.5	95.8	96.7
400	96.5	96.2	95.6	97.1	95.6
450	95.8	96.8	95.8	97.2	96.8
500	96.7	95.1	96.5	96.2	96.5
Precision ratio/%	96.3	96.0	96.5	96.3	96.2

**Table 3.** The accuracy of this method.

Test set Number/Individual	Test Times				
	First Test/%	Second Test/%	Third Test/%	Fourth Test/%	Fifth Test/%
50	99.5	99.2	99.3	99.3	100
100	99.5	99.8	99.6	99.4	99.5
150	99.1	99.3	99.7	100	99.1
200	99.2	99.4	99.4	99.6	99.2
250	99.5	99.6	99.8	99.8	99.5
300	99.6	99.6	99.4	99.9	99.8
350	99.8	99.5	99.6	99.5	99.9
400	99.9	99.8	99.6	99.1	99.5
450	99.7	99.7	100	99.2	99.1
500	100	99.6	99.4	100	99.2
Precision ratio/%	99.6	99.6	99.6	99.6	99.5

In order to clearly express the precision of the precision of the method, the average of the precision of the above Tables 1–3 is made into a graph as shown in Figure 5.

**Figure 5.** Comparison of the average of three methods.

### 3.2. Comparison Experiment of the Retrieval Efficiency

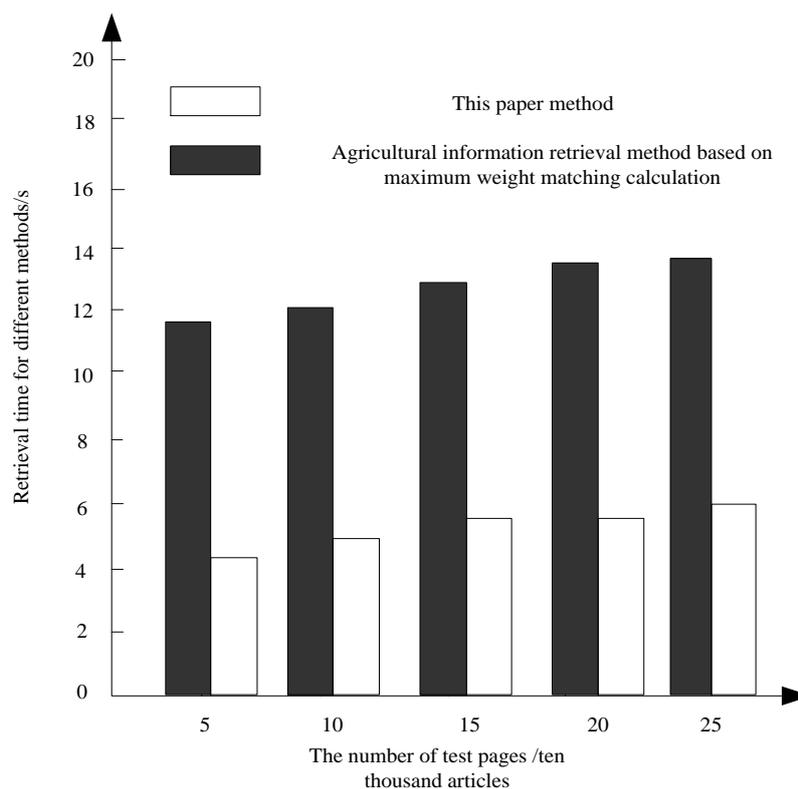
#### 3.2.1. Comparison of Retrieval Time

The retrieval time consumption of the three methods is described in Table 4.

**Table 4.** The time consumption of retrieving digital agricultural information by three methods.

Test the Number of Pages/Tens of Thousands of Pages	This Paper Method/s	Agricultural Information Retrieval Method Based on Maximum Weight Matching Calculation/s	Agricultural Information Retrieval Method Based on Dual Semantic Space/s
2	2.5	3.2	6.8
4	3.2	4.2	8.2
6	3.2	5.1	10.2
8	4.2	6.8	13.8
10	4.8	7.2	15.4
12	5.2	9.1	19.4
14	5.5	9.9	23.4
16	6.8	11.2	26.7
18	8.6	12.9	28.9
20	9.1	15.2	32.2
Mean time/s	5.3	8.5	18.5

The retrieval time obtained by the proposed method in the paper and the method based on the maximum weight matching is as shown in Figure 6.



**Figure 6.** Different ways to retrieve digitalized agricultural information.

#### 3.2.2. Comparison of Retrieval Efficiency

The retrieval efficiency of the three methods in the three stages are described with reference to Figure 7, Figure 8, and Figure 9, respectively.

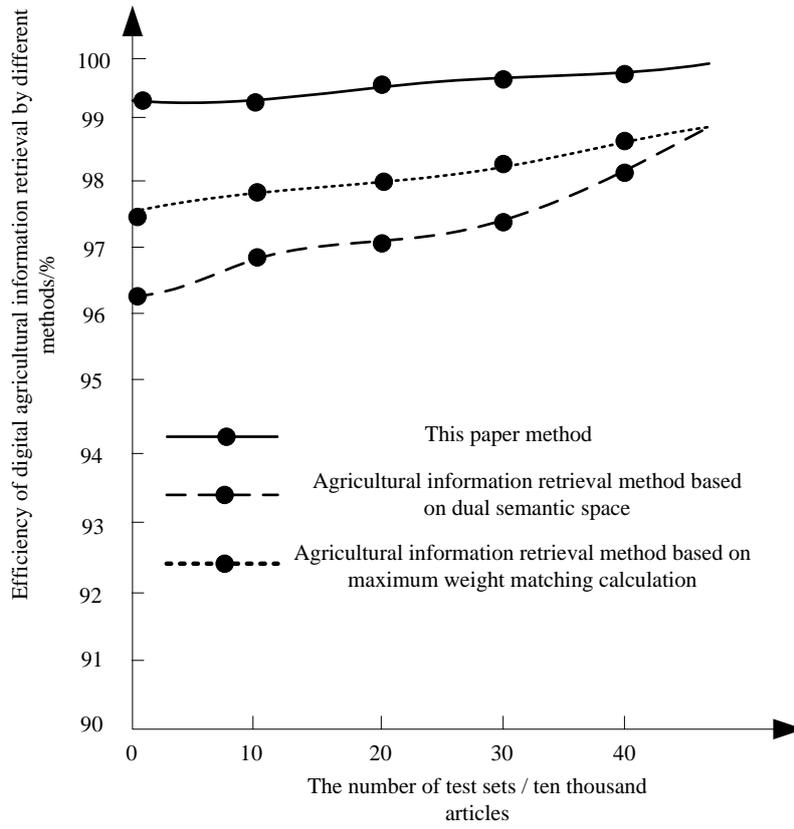


Figure 7. Comparison of the retrieval efficiency of three methods in the first stage.

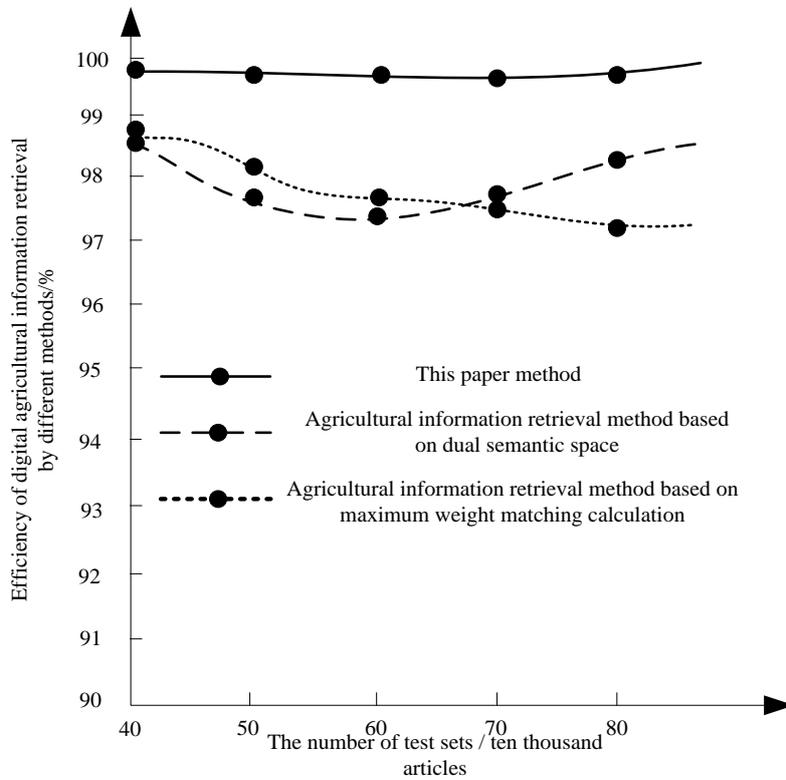


Figure 8. Comparison of the retrieval efficiency of three methods in the second stage.

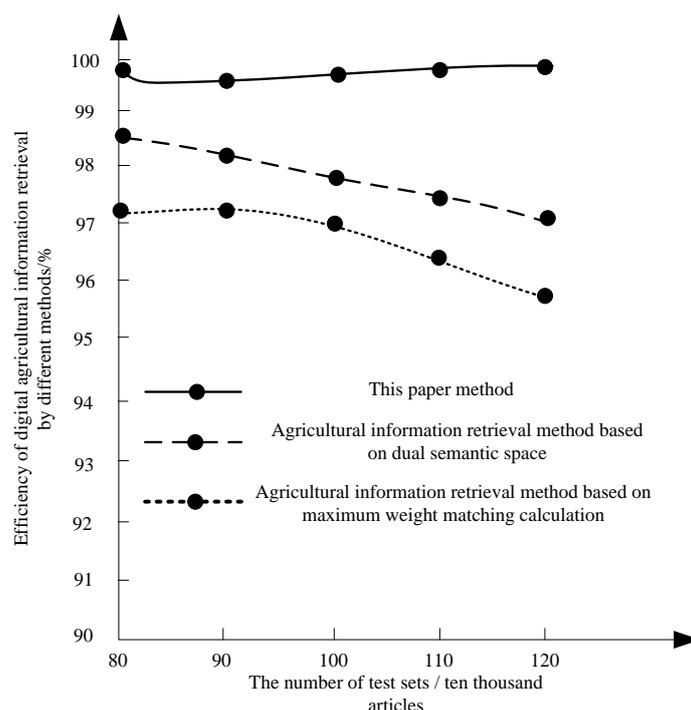


Figure 9. Comparison of the retrieval efficiency of three methods in the third stage.

## 4. Discussion

### 4.1. Discussion on the Retrieval Effect of the Three Methods

#### 4.1.1. Recall Rate

When the number of test sets is 100, 200, 300, 400, and 500, the recall rates based on the method of maximum weight matching calculation are 96.8%, 97.0%, 96.8%, 96.8%, and 96.9%, respectively. The mean value is about 96.8%; in the same case, the recall rates obtained by the dual semantic space method are 98.8%, 98.8%, 99.2%, 97.6%, and 99.2%, respectively. However, when the number of test sets is 400, the recall rate is only 97.6%, which indicates that the stability of the method is worse than the method in this paper. According to the length of the bar graph, it can be seen that the recall rate of the method is the highest among the three methods, and the mean value is above 99%. The above data indicates that the information retrieval rate of this method is the highest.

Figure 2 shows a comparison of the recall rates of the three methods with 10% interference information added. Overall, the method is located at the top of the graph, indicating that the method has the highest recall rate. The recall rates of the methods in this experiment are 99.8%, 99.8%, 99.7%, 99.8%, and 99.9%, respectively [23,24]. It can be seen from the data that the precision of this method appears to be slightly fluctuating, but it is generally rising. The results obtained in 5 experiments are all above 99.5%. This shows that the digital agricultural information retrieval of the method is comprehensive and has advantages over other methods, which can be used for the effective retrieval of digital agricultural information.

Analysis of Figure 3 shows the comparison of the recall rates of the three methods with 20% interference information added to the test data set. In this experiment, the recall rate based on the method of maximum weight matching calculation is directly proportional to the number of experiments. The recall rates obtained in 5 experiments are 98.4%, 98.6%, 99.0%, 99.3%, and 99.6%, the data of this group indicate that the method has great potential; the recall rates obtained in the five experiments in this method are: 99.7%, 99.7%, 99.8%, 99.8%, and 99.9%, respectively.

According to this set of data, it can be seen that the recall rate of this method has not decreased due to the addition of 20% interference information in the test set. The recall rate of this method is above

99.5%, showing a good recall status. Comprehensive analysis, the accuracy of this method for digital agricultural information retrieval is higher, and it has advantages compared with other methods.

Analysis of Figure 4 shows the comparison of the recall rates of the three methods with 30% interference information added to the test data set. The proportion of interference information in this experimental test set continues to increase. It can be seen from the graph that the recall rate of the three methods has decreased. In the case of adding 30% interference information, the recall rate of this method also decreased, but compared with the other two methods, the decline is small, and the recall rates obtained by 5 experiments are 99.5%, 99.5%, 99.6%, 99.6%, and 99.7%.

It can be seen from the data that the recall rate of this method is over 99.5%, which has the advantage of high recall rate. Based on the above discussion, it can be concluded that the method has the high recall rate in the process of retrieval digital agricultural text information. In view of the shortcomings of the other two methods, the methods in this paper have been made up, which greatly improved the recall rate of digital agricultural text information retrieval.

#### 4.1.2. Precision Rate

Analysis of Tables 1–3 shows that the three methods are used to compare the precision of digital agricultural information retrieval. The precision rate of agricultural information retrieval method based on maximum weight matching calculation is about 94.2%, and the precision rate of agricultural information retrieval method based on double semantic space is about 96.3%. The precision rate of this method is about 99.6%.

Analysis of Figure 5 can clearly see the comparison of the mean values of the three methods. It is not difficult to see from the data in the figure that the precision rate of this method is about 3% higher than the precision rate based on the dual semantic space method, and about 5% higher than the precision rate based on the method of maximum weight matching calculation. In summary, the method can be used to digitize the effective retrieval of agricultural information.

The experimental results show that the recall rate of this method is superior to the other two methods. In the case of adding interference information, the recall rate is greater than or equal to 99.5%, and the full result is less affected by the interference information;

The precision rate of this method is above 99.5%, which can obtain accurate agricultural information retrieval results. Therefore, the method can obtain comprehensive and accurate digital agricultural information retrieval results, which can be used for the effective retrieval of digital agricultural information.

### 4.2. Discussion on the Retrieval Efficiency of the Three Methods

#### 4.2.1. Retrieval Time

Analysis of Table 4 shows that the time of the method in this test is 2.5 s, 3.2 s, 3.2 s, 4.2 s, 4.8 s, 5.2 s, 5.5 s, 6.8 s, 8.6 s, and 9.1 s, and the average time is 5.3 s.

As the number of test sets increases, the method uses the small increase in time; while the average retrieval time of the other two methods is 8.5 s and 18.5 s, respectively, and the retrieval efficiency is significantly lower than the method.

In Figure 6, the time consumption of the method for retrieving agricultural information is 4.1 s, 5.0 s, 5.5 s, 5.5 s, and 6.0 s, respectively. At the initial stage of testing, the number of test pages is 50,000 at least 4.3 s.

In the later stages of testing, the number of test pages is 250,000 and the maximum time is only 6.0 s.

Comparing the bar graphs of the two methods, the bar graph of the method is obviously higher than the agricultural information retrieval method based on the maximum weight matching calculation by about 6%.

Although the retrieval time of both methods increases with the increase of the number of test sets, the retrieval time of this method is far less than the method based on the maximum weight matching calculation.

Based on the data results in Table 4 and Figure 6, it can be seen that the method of this paper retrieves digital agricultural information in a shorter time and has the advantage of high efficiency.

#### 4.2.2. Retrieval Efficiency

It can be seen from Figure 7 that in the first stage, the retrieval efficiency of the three methods increases with the increase of the number of test sets, showing the better state. The retrieval efficiency obtained by this method are 99.2%, 99.1%, 99.4%, 99.4%, and 99.5%, which are located at the top of the graph and have the highest efficiency among the three methods.

Analysis of Figure 8 shows that in the second stage, the retrieval efficiency of the proposed method is still above 99.5%. The retrieval efficiency of the other two methods decreases and fluctuates greatly with the increase of the number of test sets. In general, with the increase of the number of test set, the retrieval efficiency of the other two methods is significantly reduced. The retrieval efficiency of this method is still stable and has the high efficiency advantage.

It can be seen from Figure 9 that in the third stage, the retrieval efficiency of the method is still between 99.5% and 99.8%.

The experimental results show that the method of digital agricultural information retrieval is short in time and high in efficiency. The average retrieval time is between 4 s and 6 s, and the average retrieval efficiency is over 99%. It can be used for the rapid retrieval of digital agricultural information.

## 5. Conclusions

In view of the limitations of the existing digital agricultural text information retrieval methods, and in order to improve the retrieval efficiency of the digital agricultural text information, a local matching based digital agricultural text information retrieval method is proposed. Construct the agricultural text tree and query tree, generate the relationship between ancestors and descendants in the query, and map it to the agricultural text. According to the local matching retrieval method, the vector retrieval method is used to calculate the similarity between digital agricultural texts and submit queries. The similarity is sorted from large to small, so that the agricultural text tree can output the digital agricultural text information in turn.

The cost of this method is less, and the experimental results show that this method can achieve high recall and precision in a short time, with high recall and precision. The method proposed in this paper not only makes up for the shortcomings of traditional methods, but also provides an effective and scientific retrieval method for digital agricultural text information, and provides a reference for agricultural information processing and research in this field at home and abroad. At the same time, it also proves the reliability of the local matching method in data retrieval, which can be applied to more fields. However, the collected data is not comprehensive, so it is still necessary to strengthen the research in this area in the future work, so that the method can be more widely used.

**Author Contributions:** Conceptualization, Y.S.; Data curation, W.G.; Formal analysis, M.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Wu, Z.; Gao, K.; Wang, Z.L. Linear information retrieval method in X-ray grating-based phase contrast imaging and its interchangeability with tomographic reconstruction. *J. Appl. Phys.* **2017**, *121*, 23–26. [[CrossRef](#)]

2. Yang, Y.G.; Sun, S.J.; Wang, Y. Quantum oblivious transfer based on a quantum symmetrically private information retrieval protocol. *Int. J. Theor. Phys.* **2015**, *54*, 910–916. [[CrossRef](#)]
3. Hu, Y.T.; Hui, F.P. The Application of Metadata Methods in the Perspective of Digital Humanities: A Case Study of Agricultural Heritage. *Library* **2019**, *000*, 82–87.
4. Ming, Y.; Cuicui, F.; Fuchuan, M. Research on Innovation of Information Resource Management in the Age of Big Data. *Res. Libr. Sci.* **2019**, *6*, 56–61.
5. Wang, L. Text information retrieval algorithm simulation analysis under massive data. *Comput. Simul.* **2016**, *33*, 429–432.
6. Junnila, V.; Laihonon, T. Information retrieval with varying number of input clues. *IEEE Trans. Inf. Theory* **2016**, *62*, 625–638. [[CrossRef](#)]
7. Shi, J.; Liu, D.; Cui, L. Retrieval of contaminated information using random lasers. *Appl. Phys. Lett.* **2015**, *106*, 685.
8. Besson, M.; Kutas, M.; Vanpetten, C. Effect of semantic expectancy upon information-retrieval. *Psychophysiology* **2016**, *23*, 425–426.
9. Khennak, I.; Drias, H. An accelerated PSO for query expansion in web information retrieval: Application to medical dataset. *Appl. Intell.* **2017**, *47*, 793–808. [[CrossRef](#)]
10. Krause, C.; Johannsen, D.; Deeb, R. An SQL-based query language and engine for graph pattern matching. *Networks* **2016**, *20*, 345–359.
11. Liang, M.; Du, J.; Cao, S. Super-resolution reconstruction based on multisource bidirectional similarity and non-local similarity matching. *IET Image Process.* **2015**, *9*, 931–942. [[CrossRef](#)]
12. Subber, W.; Matouš, K. Asynchronous space–time algorithm based on a domain decomposition method for structural dynamics problems on non-matching meshes. *Comput. Mech.* **2016**, *57*, 211–235. [[CrossRef](#)]
13. Zhang, X.; Tan, C.L. Handwritten word image matching based on Heat Kernel Signature. *Pattern Recognit.* **2015**, *48*, 3346–3356. [[CrossRef](#)]
14. Bors, A.G.; Papushoy, A. Image retrieval based on query by saliency content. *Digit. Signal Process.* **2015**, *36*, 156–173.
15. Kotsifakos, A.; Karlsson, I.; Papapetrou, P. Embedding-based subsequence matching with gaps-range-tolerances: A Query-By-Humming application. *VLDB J.* **2015**, *24*, 519–536. [[CrossRef](#)]
16. Duch, A.; Lau, G.; Martínez, C. On the cost of fixed partial match queries in K-d trees. *Algorithmica* **2016**, *75*, 1–40. [[CrossRef](#)]
17. Tang, J.; Luo, J.; Tjahjadi, T. Robust Arbitrary-View Gait Recognition Based on 3D partial similarity matching. *IEEE Trans. Image Process.* **2016**, *26*, 7–22. [[CrossRef](#)] [[PubMed](#)]
18. Mei, K. Opportunities for women, minorities in information retrieval. *Commun. ACM* **2017**, *60*, 10–11.
19. Lyu, H.; Zhang, J.; Zha, G. Developing a two-step retrieval method for estimating total suspended solid concentration in Chinese turbid inland lakes using Geostationary Ocean Colour Imager (GOCI) imagery. *Int. J. Remote Sens.* **2015**, *36*, 1385–1405. [[CrossRef](#)]
20. Niazi, S.K.; Alam, S.M.; Ahmad, S.I. Partial-area method in bioequivalence assessment: Naproxen. *Biopharm. Drug Dispos.* **2015**, *18*, 103–116. [[CrossRef](#)]
21. Gao, W.; Zhu, L.; Guo, Y.; Wang, K. Ontology learning algorithm for similarity measuring and ontology mapping using linear programming. *J. Intell. Fuzzy Syst.* **2017**, *33*, 3153–3163. [[CrossRef](#)]
22. Gao, W.; Wang, W. A tight neighborhood union condition on fractional  $(g, f, n', m)$ -critical deleted graphs. *Colloq. Math.* **2017**, *149*, 291–298. [[CrossRef](#)]
23. Shirakol, S.; Kalyanshetti, M.; Hosamani, S.M. QSPR analysis of certain distance based topological indices. *Appl. Math. Nonlinear Sci.* **2019**, *4*, 371–385. [[CrossRef](#)]
24. Dewasurendra, M.; Vajravelu, K. On the method of inverse mapping for solutions of coupled systems of nonlinear differential equations arising in nanofluid flow, heat and mass transfer. *Appl. Math. Nonlinear Sci.* **2018**, *3*, 1–14. [[CrossRef](#)]

