



Intelligent Web Information Extraction Systems for Agricultural Product Quality and Safety System

Mohammad Ali Tofigh¹, Zhendong Mu²

¹Faculty of Information Science and Engineering, Management and Science University, Malaysia

²Jiangxi University of Technology, Nanchang 330098, Jiangxi, China

Emails: mohd_ali@msu.edu.my, zdmu123@jxut.edu.cn

Abstract:

With the development of society, people pay more and more attention to the safety of food, and relevant laws and policies are gradually introduced and being improved. The research and development of agricultural product quality and safety system has become a research hot spot, and how to obtain the Web information of the system effectively and quickly is the focus of the research, so it is essential to carry out the intelligent extraction of Web information for agricultural product quality and safety system. The purpose of this paper is to solve the problem of how to efficiently extract the Web information of the agricultural product quality and safety system. By studying the Web information extraction methods of various systems, the paper makes a detailed analysis and research on how to realize the efficient and intelligent extraction of the Web information of the agricultural product quality and safety system. This paper analyzes in detail all kinds of template information extraction algorithms used at present, and systematically discusses a set of schemes that can automatically extract the Web information of agricultural product quality and safety system according to the template. The research results show that the proposed scheme is a dynamically extensible information extraction system, which can independently implement dynamic configuration templates according to different requirements without changing the code. Compared with the general way, the Web information extraction speed of agricultural product quality safety system is increased by 25%, the accuracy is increased by 12%, and the recall rate is increased by 30%.

Key words: Quality safety system, Web information, Information extraction, For agricultural products

1. Introduction

Since the 1980s, driven by reform and opening policies, China's economy has begun to develop at a high speed [1]. With the rapid development of our country's economy and the improvement of people's living standards, food safety issues are receiving more and more attention. The 2017 Central Economic Work Conference, the Rural Work Conference and the Central Document No. 1 clearly stated that the quality of agriculture should be upheld, and the agricultural standardization strategy should be implemented. Subsequently, Shan doing Province also proposed a key development direction for food safety issues: under the leadership of the "Food Safety Shan dong" construction, continuously improving the province's food safety governance capabilities, promoting the healthy and sustainable development of the food industry, and defending the public "safety on the tongue" [2]. Under the correct leadership of the Provincial Party Committee and the Provincial Government, all agricultural departments adhere to the quality

and safety of agricultural products as the strategic direction of the development of modern agriculture in Bin Zhou. The main goal is to promote the structural adjustment of the supply side of agriculture. , Adhere to the problem-oriented, and gradually form a suitable agricultural product quality road. In addition, the frequent occurrence of agricultural product quality and safety incidents has attracted great attention from the country. In recent years, a number of related laws and regulations have been promulgated. After years of active attempts, the framework of China's agricultural product standards, testing, and certification systems has also initially formed [3]. However, from the perspective of practical effects, there is still a large gap with developed countries, and systematic theoretical research is urgently needed to improve it. From the perspective of theoretical research, the existing research is more about how to improve the quality assurance safety from the surface of the quality of agricultural products, and lacks a systematic understanding of the quality, safety, and comprehensiveness of agricultural product quality and safety management systems. The overall and related aspects of other aspects are insufficiently considered, and a systematic analysis of the agricultural product quality and safety management system is lacking. In order to solve these problems, the agricultural product quality and safety system has become a hot spot in research and development, but the information extraction method is very inefficient.

Internationally, Web information extraction has become an important research topic. Need to browse and find massive amounts of network information, including network news reports, related comments, online forums, etc., to extract useful information related to analysis from this information. Most of the information on different networks is presented in the form of web pages, but the structure of web pages is constantly changing, in addition to traditional single structure type web pages like news; there are also forums, blogs, wikis, wikis, wiki Interactive (multi-structured) dynamically typed web pages [4]. And no matter what type, the web page is full of a lot of "floating" advertisements, website navigation, unnecessary pictures, links, and copyrights that are not related to the subject matter of the web page [5]. How to obtain the information quickly and effectively you need without being affected by web page noise, how to intelligently extract large amounts of information content from the Internet, meet the needs of users of various types of research, and distribute complex and vast cross-regional distribution The transformation of research on irregular data into local structured text data has become an important research topic today. At present, most of the information extraction technologies studied at home and abroad basically extract information from web pages with fixed format or the same information organization distribution law, such as conference paper information, commodity information, and book information [6]. The purpose of these studies is not to extract the body content of web pages, but to convert unstructured or semi-structured, irregular data in web pages into structured and regular data. Web information extraction technology has been greatly developed in recent years, and the efficiency of various extraction methods is also different. Therefore, what method to use for Web information extraction of agricultural product quality and safety systems is the focus of research, and it is also the main research content of this article.

To explore how to efficiently and intelligently extract Web information from agricultural product quality and safety systems, this paper conducts a study on a scheme that can automatically extract Web information from agricultural product quality and safety systems based on templates. Among them, Han made a detailed introduction to the investigation and construction of food and agricultural product quality and safety risk assessment systems, analyzed the current problems of agricultural product quality and safety, and explained related laws and regulations. It shows that the national government attaches great importance to the quality and safety of agricultural products and indicates the importance and research significance of the current agricultural product quality and safety system [7]. In his article, ZHAO proposed the research status and development prospects of agricultural product quality and safety systems and expounded the existing problems of several agricultural product quality and safety systems, especially the inefficiency of information processing. In addition, it showed the significance and importance of agricultural product quality and safety system for food monitoring and made solutions to improvement and problems [8]. Pan expounded in detail the pursuit of agricultural product quality and safety system, the principle and implementation path of judging the quality of agricultural products, and put forward the importance and impact of Web information extraction of agricultural product quality and safety system, and the necessity of carrying out related research [9] . Mercedes proposed several common Web information extraction and extraction techniques and related algorithms and pointed out the advantages and disadvantages of various methods. It also puts forward some common problems in the current Web information extraction technology, and briefly introduces the problems encountered in various industries and their related performances [10]. Zhan proposed a templated information extraction algorithm in Web information extraction technology, briefly introduced this algorithm, and explained the feasibility and basic principles of this

algorithm. It shows the superiority of this information extraction method [11]. Lie proposed a traditional and common Web information extraction technology: a method for identifying spatial information of Web documents based on semi-supervised machine learning. The extraction efficiency of this method is introduced in detail, which contrasts the information extraction methods proposed in this paper [12].

To put it simply, this article seeks to effectively extract the web information of agricultural product quality and safety, and develops a scheme that can automatically extract web information of agricultural product quality and safety system based on templates as the main research content. Specifically, the main research content of this article is roughly divided into five parts: the first part is the introduction part, which aims to systematically review the main research content of this article from the research background, research purpose, research ideas and methods; The theoretical basis, a detailed and systematic summary of the current research status of agricultural product quality and safety systems and Web information extraction technology, and also introduced the current application of template information extraction algorithms in Web information extraction. The third part is related research, which illustrates the application status of template information extraction algorithm in Web information extraction of agricultural product safety and quality system by querying data and conducting related experiments. The fourth part is the analysis of the data, and the feasibility and excellence of the scheme are verified from the aspects of extraction speed and accuracy through specific survey data and research results; the fifth part is the summary and recommendations of this article. The summary of the results of this article and the prospect of further application of template information extraction algorithm in the web information extraction of agricultural product quality safety system.

2. Proposed Method

2.1. Agricultural Product Quality and Safety System

Many years have passed since the development of agricultural product quality and safety systems. System science includes a very wide range of research fields. Since its inception, system science has been widely used in economic, political, military, diplomatic, cultural, ecological, and economic management departments. Based on systematic thinking, research scholars in different fields and disciplines have proposed, such as: "ecosystem", "physical system", "information system", "cultural system", "social system", "economics" System "and other system concepts. The study of these system concepts has its own emphasis, but summed up, we can find that the formation of any system must have three conditions: one is that the system is a whole with a specific function; the other is that the system is composed of several parts; the third is that each component There is interaction and interdependence between parts. In short, agricultural product quality and safety management is a collective term for all management actions or rules and regulations that are conducive to the promotion and protection of agricultural product quality and safety. According to the three components of the above system and the definition of agricultural product quality and safety management, this study defines the connotation of agricultural product quality and safety system as follows: First, the agricultural product quality and safety management system is a whole with specific functions. Its functions include internal and external aspects. The internal function is to improve the quality and safety of agricultural products through the internal operation of the system, and the external function is to ensure the competitive optimization of the agricultural product market and the smooth realization of the value of agricultural products through the close connection with the external environment.

Secondly, the agricultural product quality and safety system is composed of several parts and elements, and has different division methods from different perspectives. From the point of view of vitality, these parts and elements can be divided into the main subsystem and the flow element subsystem. The main subsystem covers the human factor in the management of agricultural product quality and safety. From the perspective of functional types, it includes suppliers, logistics enterprises, Sales enterprises, consumers, regulators, etc .; the flow element subsystem refers to the various material flows, energy flows, and information flows necessary to maintain the normal operation of agricultural product quality and safety management, including information and technology. From the perspective of function realization, the agricultural product quality and safety management system can be regarded as consisting of two aspects: the agricultural product value creation subsystem and the agricultural product safety realization subsystem. The agricultural product value creation subsystem is covered by the agricultural product production and processing or agricultural product service provision process. The main function of all elements is to ensure the

realization of the value of agricultural products; the agricultural product safety realization subsystem realizes the quality and safety of agricultural products.

Thirdly, there are interactions and inter-dependencies between the various components of the agricultural product quality and safety system, that is, the subject subsystem needs the support of the flow elements subsystem to achieve the initiative of the subject. If the value creation subsystem does not cooperate with the value realization subsystem, Meaningless. In summary, a reasonable expression is: the agricultural product quality and safety management system is a main subsystem consisting of suppliers, logistics enterprises, sales enterprises, consumers, regulators, etc., and is composed of material flow, information flow, energy flow, etc. The social-economic system composed of mobile subsystems, each subsystem and its interaction with the external environment, coordinate with each other, and jointly realize the functions of agricultural product value creation, realization, and agricultural product quality and safety.

2.2. Web Information Extraction Algorithm and Technology

There are many characteristics of Web information. It can be seen from the characteristics that the semi-structured nature of Web information and noise-containing data make it difficult to extract Web information. It also explains the difference and connection between information retrieval and information extraction. Extraction. Researchers have proposed various techniques to solve this problem. The process of Web information extraction is more complicated. The following sections summarize the important information extraction techniques that have emerged. The first is information extraction technology based on regular expressions. Regular expressions are used to identify strings with certain information distribution rules. In the process of web page information extraction, the web page is first processed as a character stream file. A reasonable regular expression is configured to match the information to be extracted, and then the required information is extracted. The basic algorithm of this extraction method is as formula As shown in 1, 2, and 3:

$$C_j = \{t_i \mid f(x) = C_f, 1 \leq i \leq n, 1 \leq j \leq n \quad (1)$$

$$\tau_{ii}(t+1) = \rho \tau_{ii}(t) + \Delta \tau_{ii} \quad (2)$$

$$\Delta \tau_{ii} = \sum_{k=1}^{i+1} \Delta \tau_{ii}^k \quad (3)$$

The second method of Web information extraction is statistical web page information extraction technology. This technology is based on statistical web page information extraction technology, which uses natural language processing technology to extract information from web pages, and counts some characteristics of the information contained in web pages to extract the content of the text. This method first uses a parser to generate a corresponding DOM tree according to the structure of the HTML markup on the web page, and then counts the number of Chinese characters contained in each node in the tree, and selects the nodes containing text information from it in a certain way to extract relevant information. This method overcomes the defects of the traditional web page information extraction method and the corresponding wrapper required by different data sources, and has certain applicability. However, the extraction of web page text information in this way depends on the threshold value. The setting of the threshold value will affect the accuracy of information extraction, and the result of extraction may also contain noise. These problems also seriously affect the application of this method. In addition, when text extraction is performed, the method also sorts the obtained text in descending order according to the length of the text block string, which increases the time complexity. The basic algorithm of this extraction method is shown in formulas 4 and 5:

$$E(A) = \sum_{j=1}^V \frac{D_f}{D} * \text{Info} (D_f) \quad (4)$$

$$x = r_1 x \left[1 - \frac{x}{1 + a_1 y} \right] \quad (5)$$

The third method of Web information extraction is information extraction technology based on inductive learning. WIEN is the first Web information extraction system that uses inductive learning to extract information. It uses a unique multi-user extraction rule mechanism to extract information from all documents. The system needs to mark the sample pages in advance, and the system will automatically find suitable heuristic induction algorithms to generate different wrappers for information extraction based on the different logical structure of the pages. The system uses an inductive learning algorithm to train and generate wrappers from the query result samples. Because WIEN only considers the separators that are closely adjacent to the data to be extracted in the source code of the web page, it cannot correctly package these web pages with incomplete information or irregular distribution of data items. In this way, the accuracy of the information extraction of the entire system will vary because of the organization of the data, which will not meet the needs of those who pursue high information accuracy. The basic algorithm of this extraction method is shown in Equations 6 and 7:

$$t_i = \max\left\{\frac{\Delta x}{v_i}, \frac{\Delta y}{v_j}\right\} = \max\left\{\frac{x_2 - x_1}{v_x}, \frac{y_2 - y_1}{v_y}\right\} \quad (6)$$

$$y = r_2 y \left[1 - \frac{y}{1 + a_2 x}\right] \quad (7)$$

3. Experiments

3.1. Related Processing of Experimental Data

The object of this experiment is the agricultural product quality and safety system adopted by a local government. The feasibility of the scheme that can automatically extract the web information of the agricultural product quality and safety system based on the template was tested. During the experiment, there is a large amount of experimental data to be processed, and there must be errors in these data. It is also very important to handle the errors appropriately. Therefore, before using these experimental data for forward and reverse analysis, the error should be processed and analyzed on the original data. Generally, the errors of the experimental data can be divided into three types: system error, random error and gross error. Among them, random errors are often caused by random factors, and their signs and absolute values are irregular. However, as the number of experiments increases, random errors are generally considered to be normally distributed. The gross error mainly refers to the fact that in the statistical data, due to the observer's carelessness, or sudden changes in environmental conditions, unstable instrumentation and other factors, the observation error does not conform to a certain statistical distribution rule, which is usually a measurement error. System error is the error caused by the measurement instrument, the change of the measurement reference and the influence of external conditions. At present, the systematic error of observations is generally composed of corresponding statistics based on the statistical characteristics of observations, and then test hypotheses are made based on the characteristics of their probability distributions, and judgments are made by comparing actual calculated values with quartile values. Common test methods are: U test, variance test, t test and so on. In the measurement process, the gross error should be eliminated, and the system error should be eliminated or weakened, so that the observation value contains only the random error $I, 0$.

At present, when resolving this kind of problem at home and abroad, the least square method is usually used to process the experimental data twice. The basic idea of the least square method is to first assume that the observations only contain accidental errors, but this is basically not true in reality. Possibly, for this reason, a new theory has been developed to study systematic errors and gross errors. At present, the more effective method for processing systematic errors is the additional parameter method; there are two methods for processing gross errors. One is the data detection method that still belongs to the category of least squares, and the other is the method of robustness estimation that is different from the least squares method. Or robust estimation. In addition, in the actual situation, various social work-related links are constantly changing, and the information collection system is also in a moving state, which means that the entire collection process is dynamically changing, so there will be relative errors in experimental management. It is inevitable. Modern error theory generally believes that the measured true value cannot be determined, and the

existence of the quantum effect excludes the existence of the unique true value, so the error cannot be accurately obtained. The error used in the experiment in the past is actually a kind of deviation; the experimental error evaluated is actually unavoidable and uncertain.

3.2. Experimental Model Establishment

After obtaining the quality and safety information of agricultural products, the important work is to extract effective information from the information, so as to control the qualified quality and ensure the safety of agricultural products. In actual work, some work information is often not updated in time, and some messages are not transparent enough. Therefore, it is a very complicated problem to collect the correct service information in a timely manner and evaluate the most needed areas. The relationship between the efficiency of agricultural product quality and safety system and the extraction of Web information is the schedule model. The speed, efficiency and accuracy of progress are the keys to the quality and safety of agricultural product quality and safety systems. Therefore, in the research of agricultural product quality and safety systems, a progress arrangement model is mainly established. The purpose of establishing a progress schedule model is to establish a functional relationship between the agricultural product quality and safety system and Web information extraction, that is, to use the information obtained from various channels to determine the impact of various factors on the extraction speed and accuracy, and to establish a functional model that reflects the progress schedule. After obtaining various Web information, you can automatically generate templates to complete the information extraction according to your own dynamic adjustment.

Commonly used methods for establishing statistical models include stepwise regression, multiple regression, weighted regression, and so on. There are many factors affecting the determination of Web information extraction. In addition, the built GM model was used to predict the information in the future. The agricultural product quality and safety system is an emerging industry. Statistics on the quantitative data of the agricultural product quality and safety system market at home and abroad have only begun in recent years, so it can be used. There are very few data to predict the future market application of agricultural product quality and safety systems. The article uses the gray prediction model to predict the development of template information extraction algorithms in Web information extraction applications, and predicts the effectiveness of this extraction method from the side. On the premise of qualitative research on Web information extraction of agricultural product quality and safety systems, further enrich the quantitative research theory.

In addition, when establishing a web information extraction model for agricultural product quality and safety systems, it is necessary to determine relevant collection indicators. After establishing a statistical model, the collection indicators can be determined. From the knowledge of mathematical statistics, it can be known that when the statistical model established based on the least square method satisfies the conditions of Gaussian assumption and the normal distribution of residuals, the obtained statistical model is the best unbiased estimation. This model can be used for overall estimation and prediction. Under normal circumstances, there is no abnormality in the residual sequence obtained after the observations are fitted; if there are abnormal values, it may indicate a precursor of instability. The upper and lower limits for judging whether the collected information value is abnormal are called collection indexes. There are two common methods for establishing collection indicators based on statistical models: confidence interval method and small probability method.

3.3. Experimental Conditions and Equipment

This article discusses the web information extraction system for agricultural product quality and safety systems adopted by a local government. Its automation technology and intelligence mainly use Internet of Things sensing and identification technology, Internet of Things communication and application layer technology. These technologies and the equipment needed are the main experimental conditions and equipment for this experiment. The so-called Internet of Things perception and identification technology refers to the Internet of Thing's collection of information through perception and identification, and is the main data source of the Internet of Things. Commonly used technologies are: two-dimensional code technology, radio frequency identification RFID technology, infrared sensing technology, GPS satellite positioning technology, audio and visual identification technology, biometric identification technology, etc. Sensing technology mainly embeds sensors around or on an object, collects data of the object or the

surrounding environment, and senses various physical or chemical changes. Commonly used technologies include sensor technology, radio frequency identification technology, and so on. The sensor is the main source of information for the application of the Internet of Things. It senses the status information of the measured object, and converts the perceived information into electrical signals or other forms of information, and then outputs it. , Display and control requirements, and finally achieve automatic detection and automatic control functions. The node information table structure of the sensors used in this paper is shown in Table 1:

Table 1. Sensor node information table structure

The field names	Field type	Field meaning
INDEX_ID	int (4)	Record the serial number
NODE-TYPE	Small int (5)	Node types
SENSOR-TYPE	Small int (5)	The sensor type of the node
NWK_ADDR	Small int (5)	The network address of the node
EXT_ADDR	Small int (5)	The MAC address of the node20-27mm
TIME	Time stamp	Data information update time

The so-called Internet of Things communication and application layer technology refers to that communication technology can be divided into two categories according to the transmission medium: wired communication technology and wireless communication technology. In recent years, with the widespread use of mobile communication equipment (such as: mobile phones, tablets, etc.), wireless communication has become the fastest-growing and most widely used communication method. To another place, so as to realize the wireless transmission of data, the main technologies include radio communication, infrared communication, microwave communication and optical communication. A wireless communication network is a communication network composed of wireless communication devices connected to each other based on communication standards and protocols. In the network, the communication terminal communicates by accessing the network and relying on the network. According to the way of accessing the network, it can be divided into two types: self-organizing network and centralized network with a central control point.

4. Discussion

4.1. Design and Performance Analysis of Web Information Intelligent Extraction Template Generator

In order to study the intelligent extraction method of Web information for agricultural product quality and safety system, this paper studies a web information extraction system that can automatically extract agricultural product quality and safety system based on a template. The template generator is a key part of the system, and its principle is mainly based on the template generation algorithm designed in the article "Automatic Extraction of Web Information Based on Templates" published by the author. The algorithm is mainly divided into two parts. First, a rule set is generated to identify the segmentation marks on the web document structure tree DOM, and then the obtained segmentation marks are configured into a template. Configure this information accordingly in the template. There are two ways to configure. The first is a fully automatic machine configuration algorithm. In this method, the program automatically configures the separation markers identified by the rule generator into the template automatically without manual intervention. This method obviously reduces labor, but the extracted information is relatively rough. Only the mixed information of title, time, user name, content, etc. can be extracted, and they cannot be separated. The intelligent extraction efficiency of this fully automatic machine configuration algorithm in the web information security system of agricultural products is more efficient than traditional extraction data, as shown in Table 2.

Table 2. Extraction efficiency of automatic machine configuration algorithm for information

Task load	Traditional extract information	Configuration algorithm
The user information	0.32	0.45
Time information	0.41	0.53
Header information	0.12	0.25
Content information	0.23	0.32

The data in Table 2 reflect the superiority of the fully automatic machine configuration algorithm for intelligent extraction of Web information. It can be seen that using this fully automatic machine configuration algorithm, the speed of information extraction, the accuracy of information extraction, and the recall rate are much larger than traditional methods. Improved, work efficiency increased by about 25% than before, but this method can only be used to extract simpler information.

The second algorithm of Web Information Intelligent Extraction Template Generator is a semi-manual screening algorithm. The algorithm first uses the automatic separator to get the separation mark between comments and adds it to the comment separation item on the template. Then take one of the comment's DOM sub trees, and then use the separator auto-recognizer to obtain the separation marks between the internal information of the comments, and manually specify which data item each separation mark is in, because the machine cannot identify the separation mark. Semantics. If you need to extract detailed information, you can use this method. The accuracy of the extracted information can reach 100%. You can also configure some filtering information options in the template. Of course, the template can also be configured manually. This semi-manual screening algorithm is more efficient in extracting Web information from agricultural product quality and safety systems than traditional extraction data, as shown in Figure 1 below.

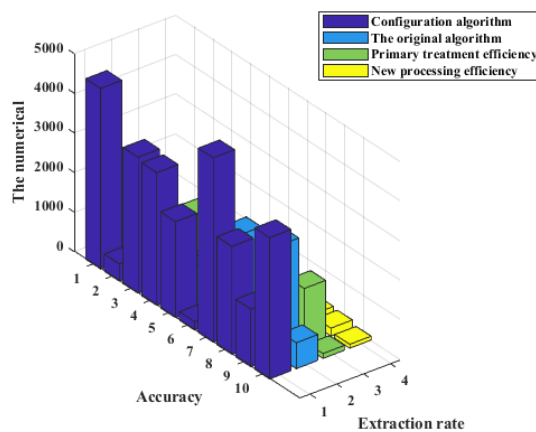


Figure 1. The efficiency of semi-automatic filtering algorithm to extract Web information

From the data in Figure 1, it can be seen that the semi-manual screening algorithm is intelligent for the extraction of Web information from agricultural product quality and safety systems. The information extraction speed is average, and it can basically achieve timely extraction and processing. Under the influence of this semi-manual screening algorithm, the speed and accuracy of information extraction have been improved, and the basic improvement is about 15%.

This type of Web information intelligent extraction template generator designed in this paper uses a combination of two algorithms. According to the actual needs of the extracted information, it can configure the division marks between content, title, user information, posting time, and so on. The specific method is to first select the minimum information surplus sub tree where this part of information is located according to the prompt, and then run the rule set generator again to generate the segmentation marks of these information. Of course, the specific meaning of these marks or where they should be configured needs According to the prompt of the program, it is completed by human participation, and finally the corresponding configuration information is generated by the program and added to the template. This comprehensive algorithm is more efficient in extracting Web information from agricultural product quality and safety systems than traditional extraction data, as shown in Figure 2.

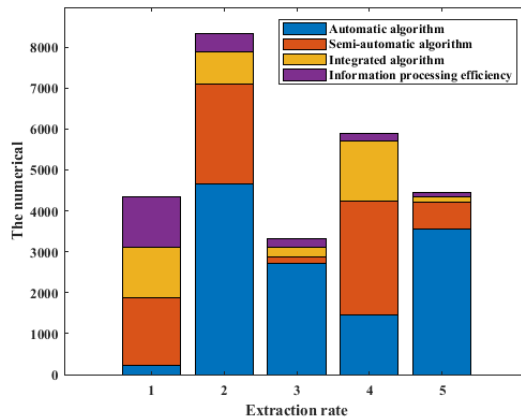


Figure 2. Efficiency of Web information intelligent extraction template generator

As can be seen from the data in Figure 2, the superiority of the Web Information Intelligent Extraction Template Generator is high in automation and high in work reliability. It can basically realize the automatic extraction and processing of the web information of the agricultural product quality and safety system, and greatly improve the system's working efficiency. Compared with the normal method, the extraction speed is increased by 25%, and the accuracy rate is increased by 12%.

4.2. Design and Performance Analysis of Web Information Intelligent Extraction Information Extractor

Web information intelligent extraction information extractor is a core content of a web information extraction system that can automatically extract agricultural product quality and safety system based on a template. Before extracting the relevant information of a page, the information extractor first reads the relevant template of this part, and then extracts the information of the clue list page according to the clue list template; extracts the information of the text page according to the template of the text page, followed by the information During the extraction process, the validity of topic clues is also judged, that is, if a topic clue is an announcement or an advertisement, it is invalid and needs to be filtered out. Finally, the extracted information is stored in the form of Topic Thread objects. The different tools used in this storage method also affect the efficiency of Web information extraction. The specific types are shown in Table 3:

Table 3. Performance analysis of information extractor

Tool	Automation level	Support for complex objects	The data type
Minerva	Artificial	Coding	SD
TSIMMIS	Artificial	Coding	SD
Web OQL	Artificial	Coding	SD
W4F	Semi-automatic	Coding	SD
XWRAP	Automatic	Support	SD
Road Runner	Automatic	Support	ST

As can be seen from Table 3, the storage capacity and automation level of the information extractor designed in this paper are better at present, which can basically meet the storage space requirements of Web information extraction. In addition, the basic flow and key technologies of the information extractor in this article include the following. The first is to extract information based on the validity of the content title (or topic). The validity of the topic in the system IWIES has three meanings: First, the quality certificate: The page of the template download contains not only things that are valuable for research, but also advertisements, announcements, copyrights, operating manuals, instructions and other clutter Information; second, the system's secondary information extraction, a large part of the web page information has been extracted, so it does not need to be re-processed; the third, the theme caused by reprinting, excerpts, etc. is repeated, that is, different sites have many Identical or similar information. This method of extracting

Web information based on the validity of the information content title (or topic) is more efficient than traditional data extraction, as shown in Figure 3.

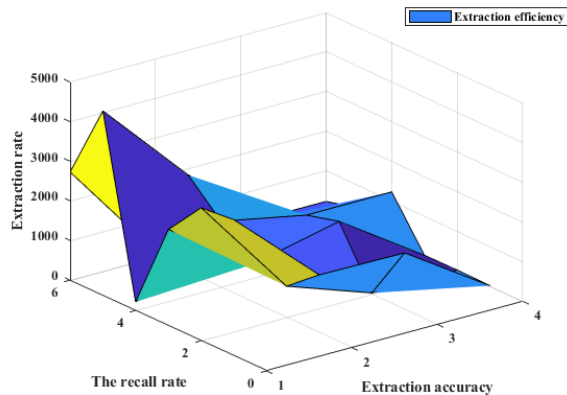


Figure 3. Information extraction efficiency of topic validity judgment method

As can be seen from the data in Figure 3, the Web Information Intelligent Extraction Information Extractor extracts the superiority of Web information efficiency based on the validity of the information content title (or topic), and can fully automatically extract Web information from the agricultural product quality and safety system. And processing, and greatly improve the system's working efficiency. Compared with the normal method, the extraction speed increased by 23%, the accuracy rate increased by 10%, and the recall rate increased by 26.

Secondly, the Web information intelligent extraction information extractor designed in this paper can effectively process according to the level of the inspected institution. In addition, the general inspection results of the inspected institutions are comprehensively ranked, and the agricultural product production and processing institutions with good integrity are given priority and fast processing. Long and careful inspection of the quality and safety web information of inspected agricultural products with a structure of poor credit. Hierarchical processing can greatly increase the efficiency of information extraction and processing. After statistical analysis, the results are shown in Figure 4:

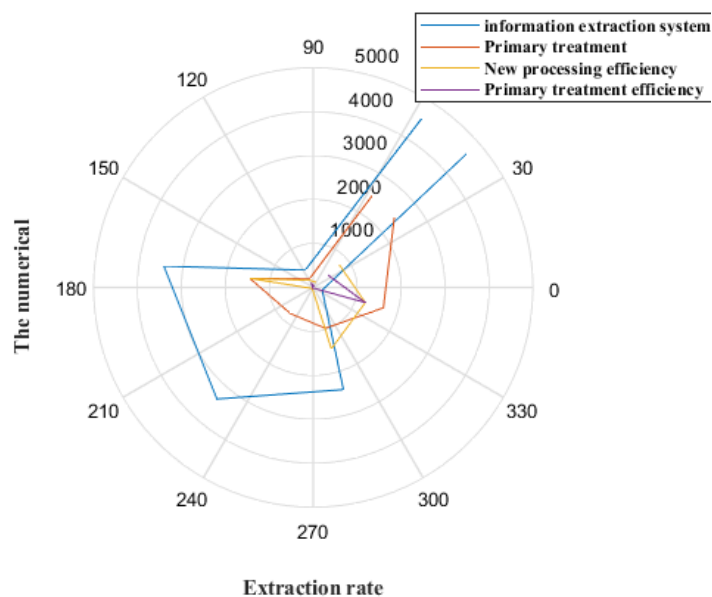


Figure 4. Information processing efficiency of information extractor

It can be seen from Figure 4 that after experimental testing, the expected effect of the intelligent extraction of Web information extractor basically meets the standards, which also reflects the superiority of the Web information extraction system for automatic extraction of agricultural product quality and safety systems based on the template proposed in this paper. After adopting this processing method, comprehensive analysis can be obtained: its extraction speed is increased by 25%, the accuracy rate is increased by 12%, and the recall rate is increased by 30%.

5. Conclusions

(1) This article analyzes the common problems existing in the quality and safety of rural products at present, discusses these problems without solving them, and proposes corresponding solutions. The development and impact of rural product quality and safety systems are briefly introduced, and the impact of Web information extraction on system work efficiency is studied. The advantages and disadvantages of various current Web extraction methods are analyzed. The feasibility and superiority of the information extractor that can automatically extract the core part of the web information extraction system of the agricultural product quality and safety system based on the template were discussed and verified.

(2) Analyze the design and performance of a template generator that can automatically extract the key parts of the web information extraction system of agricultural product quality and safety system based on the template studied in this thesis. The corresponding working principle and theoretical guidance are proposed, which confirms this the feasibility and superiority of the template generator for processing some basic information of the agricultural product quality and safety system Web can effectively increase the efficiency of information extraction. Web information intelligent extraction template generator has the advantages of high automation level and high reliability. It can basically realize the automatic extraction and processing of the web information of the agricultural product quality and safety system, and greatly improve the system's working efficiency.

(3) The Web information intelligent extraction information extractor designed in this paper can effectively process according to the level of the inspected institution. In addition, the general inspection results of the inspected institutions are comprehensively ranked, and the agricultural product production and processing institutions with good integrity are given priority and fast processing. Long and careful inspection of the quality and safety web information of inspected agricultural products with a structure of poor credit. Hierarchical processing can greatly increase the efficiency of information extraction and processing. It is verified by experiments that this scheme is a dynamically expandable information extraction system, which can independently implement dynamic configuration templates according to different needs. In addition, the web information extraction speed of the agricultural product quality and safety system is increased by 25%, the accuracy rate is increased by 12%, and the recall rate is increased by 30%.

References

- [1]Chen Jinbo, Cao Xiangliang, Fu Han-Chi. (2018). "Agricultural Product Monitoring System Supported by Cloud Computing", *Cluster Computing*, 5(3), pp.1-10.
- [2]ZHANG Xiaoyun, LI Zhemin, XIAO Hongli. (2018). "Analysis of Security Level for Improving Quality Safety of Agricultural Products in China", *Journal of Agricultural Science & Technology*, 12(5), pp.12-13.
- [3]Mari Maeda-Yamamoto, Toshio Ohtani. (2018). "Development of Functional Agricultural Products Utilizing the New Health Claim Labeling System in Japan", *Bioscience Biotechnology & Biochemistry*, 82(10), pp.1-10.
- [4]Zheng C Y, Li X G, Zhang X S. (2018). "Problems in Constructing the Modern Circulation System of Agricultural Products in China and Countermeasures", *Cluster Computing*, 8(1), pp.3-5.
- [5]Guanghong ZHOU. (2018). "Highlights in Agri-Product Quality and Safety", *Frontiers of Agricultural Science & Engineering*, 7(11), pp.6-7.
- [6]WANG Qiang. (2018). "Agricultural Products Processing Characteristics and Quality Evaluation", *Journal of Integrative Agriculture*, 17(5), pp.55-59.

- [7]Han J, Lin-Gang L U, Qin Y C. (2018). "Investigation and Construction of Food and Agricultural Products' Quality and Safety Risk Assessment System", *Bioscience Biotechnology & Biochemistry*, 2(10), pp.23-24.
- [8]ZHAO Hai-xiang, DONG Zhao-feng, ZHANG Jian. (2017). "Current Situation of Agricultural Products Quality and Safety Inspection System in Zhen'an County", *Heilongjiang Agricultural Sciences*, 15(7), pp.17-18.
- [9]Pan Shouhui, Wang Kaiyi, Wang Zhibin. (2017). "Emergency Tracking Model of Agricultural Product Quality and Safety on Web Based on Incremental Clustering", *Journal of Intelligence*, 17(4), pp.45-48.
- [10]Mercedes Arguello Casteleiro, Dmitry Tsarkov, Bijan Parsia. (2017). "Using Semantic Web Technologies to Underpin the SNOMED CT Query Language", *Journal of Intelligence*, 8(4), pp.16-18.
- [11]Zhan, Mu Qing, Lu, Rong Hua. (2016). "Design and Implementation of Web Extraction System of Ceramic Products Information in the Business Website", *Advanced Materials Research*, 6(14), pp.25-26.
- [12]Lie H, Nayak R, Wyeth G. (2017). "Spatial Information Recognition in Web Documents Using a Semi-supervised Machine Learning Method", *Journal of Intelligence*, 3(6), pp.36-37.