Automatic summary generating technology of vegetable traceability for information sharing

Zhang Zhenxuan¹ and Peng Minjing¹,²
¹School of Economics and Management, Wuyi University, Jiangmen, Guangdong 529020, China
²Engineering Technology Center of E-commerce Augmented Reality of Guangdong Province, Jiangmen, Guangdong 529020, China

Abstract. In order to solve problems of excessive data entries and consequent high costs for data collection in vegetable traceability for farmers in traceability applications, the automatic summary generating technology of vegetable traceability for information sharing was proposed. The proposed technology is an effective way for farmers to share real-time vegetable planting information in social networking platforms to enhance their brands and obtain more customers. In this research, the influencing factors in the vegetable traceability for customers were analyzed to establish the sub-indicators and target indicators and propose a computing model based on the collected parameter values of the planted vegetables and standard legal systems on food safety. The proposed standard parameter model involves five steps: accessing database, establishing target indicators, establishing sub-indicators, establishing standard reference model and computing scores of indicators. On the basis of establishing and optimizing the standards of food safety and traceability system, this proposed technology could be accepted by more and more farmers and customers.

1 Introduction

With the development of social networks and public concerns on food safety, customers has put forward higher requirements on vegetable planting for higher quality. The requirements make consumers eager to obtain and understand the related information in the process of vegetable planting, which requires a system for vegetable traceability to present and explain the planting process to customers to enhance consumers’ confidence in the quality of the planted vegetables.

Due to the social responsibility of food safety, governments at all levels are responsible for promoting the application of of vegetable traceability. However, farmers do not want to invest in vegetable traceability, because it will increase costs, but can not enhance their incomes.

In this research, the automatic summary generating technology of vegetable traceability will be proposed to stimulate farmers to actively apply traceability technology by sharing vegetable planting information in social network platforms to enhance their brands. In the vegetable traceability system, the summary for vegetable planting is automatically generated to solve the problem of excessive work loads for data collection and data input.

The proposed system consists three parts: (1) The data part. There are two types of data: the fact data and the derived data. Fact data are collected and shared by farmers. Fact data include the type of the plant, the soil of the plant, the beginning and ending time of the planting, etc. The derived data are collected from the published data of Internet. The derived data include the accumulated temperature, light periods and local soil environmental data, etc. (2) The reference models. These models are established based on the standards from departments of agriculture and research institutions of...
agriculture[11]. (3) The summary. The summary is generated based on the collected data and the reference models by inputting the fact data and the derived data into the reference models to obtain scores and reviews[1,2].

2 Data collection

The key points of the data collection for the traceability of vegetables include the collected data and the collecting methods[8].

2.1 Fact data

The fact data are the most important basis for generating summary. The fact data include the type of planted vegetable, the planting location, the starting time and the ending time of the planting. They are the principal part for judging the planting status and the quality of vegetables.

Types of vegetables determine the requirements for the planting environments, growth duration, and required planting methods. By correctly conduct the planting activities, farmer will be expected to achieve vegetables with the standard health.

Planting location determines the growth environments of the vegetables. The job of data collection is helpful in evaluating whether the soils, climates, the sunshine, the humidity, and the irrigation meet the required planting standards. And customers can evaluate the health levels and the nutrition levels of the planted vegetables to make their purchasing decisions.

Knowing the starting times of planting processes is helpful for us to be aware whether the planted vegetables are growing in the suitable seasons, which may greatly affect consumers’ purchasing decisions, because that vegetables bud and grow differently with different values of temperature and humidity in different seasons, and consequently the corresponding vegetable products have different levels of quality.

The ending time of planting indicates harvesting time and harvesting methods. Consumers use it to judge whether the vegetable is planted in a standard way, whether the vegetable has the required completely growing cycle, and whether it has reached the maturity conditions so that it isn’t prematurely harvested. The three factors are helpful in preserving the nutrition of the vegetable[9].

2.2 Derived data

The most important parts of the derived data are the accumulated temperature, the photoperiod and heavy metal pollution of the soil.

Accumulated temperature is the sum of the daily average temperature over 10 degrees Celsius during the period of a year. It is an indicator of the relationship between temperature and the growth rate of vegetables. For vegetables in some planting environments meeting the basic conditions, the temperature and the growth rate of the vegetables is positively correlated. For the different types of vegetables, they have different growth stages, their birth initiation temperatures (the lowest temperature to start growth and development) are also different. The temperature plays a vital catalytic role in the growth and development of biological organisms when the daily average temperature is higher than the birth initiation temperature. Accumulation temperature on the vegetable development and growth stage is critical, and determines the health situation of the plants in the development and growth stage, and provides the vegetable with the ability to achieve the desired nutritional value.

Photometric phenomenon is the biological response of vegetables to the pattern of light and dark circadian cycle. Most of the annual plant growth depends on the length of sunshine per day. The seeds need enough sunshine or continuous light to germinate.

The degree of damage to vegetables by heavy metal pollution is mainly determined by heavy metals in soil. The amount of heavy metals absorbed by plants is linearly related to the heavy metal in the soil. The research on plant heavy metal pollution mainly focused on plant yield index and its influence on heavy metal absorption. Heavy metal contaminated soil can directly or indirectly lead to changes in plant quality of agricultural products. Heavy metal pollution has critical effects on the yield and quality of plants. The collected data of heavy metal pollution can provide reference for preventing heavy metal pollution on human health through vegetables[10].
2.3 Method of data collection
The collection of vegetable fact data depends on farmers’ sharing behavior on vegetable planting activities. Farmers proactively collect and share every processes of the agricultural activity. We collect fact data and build data sources from shared information by farmers.

2.3.1 Fact data. Collected data include: the type of planting vegetables, planting soil conditions, the starting time of planting, the ending time of planting.

In the process of data collection, QR code (Quick Response Code) is used to input information. The input process involves two steps: (1) farmers scan a QR code, which standard for an input pages in the proposed system, to input data by their smart phone; (2) the system saves the input data into the database through the Internet.

The drawbacks of the traditional approach are prone to the original record is not clear, which leads to the input process error, and finally lead to information mistakes. The proposed method of data input can effectively avoid the emergence of the problems of the traditional approach.

The system is operated based on the Android operating system of smart phones, the SQLite database and the application server[6,7].

Using the proposed method of data collection, the data input could be conveniently completed in the fields by farmers. In this way, the costs of data input are greatly reduced.

The form of collected data is also enriched from the simple text to the graphics, the video, etc. These forms of collected data greatly enhanced the credibility of traceability information.

2.3.2 Derived data. Derived data mainly consist of the accumulated temperature, the photoperiod and the level of heavy metal pollution. The data source comes from the published information of the websites of the relevant departments of National Land Resources Bureau and the National Weather Bureau.

2.4 Computing model of summary generating
Generated summaries are visually helpful for customers to understand the product quality of the planted vegetables. In order to allow customers to intuitively understand the quality of vegetable, a summary report needs to be generated based on every indicators influencing human health according to the the weight values.

At first, some domain experts are invited to rate each indicators, and then AHP (Analytic Hierarchy Process) is used to get the corresponding weight of each indicators. If the traceability information of the vegetables was complete, all the relevant parameters of the above indicators could be obtained[12].

The value $y_i$ of the parameter $i$ is given in two forms: the bilateral function is in equation (1) and the unilateral function is in equation (2).

$$
y_i = \begin{cases} 
W_i, & \text{if } e_l \leq x_i \leq e_u \\
-1 \times W_i \times \left( x_i - e_l \right), & \text{if } x_i < e_l \text{ or } x_i > e_u \\
-1 \times W_i \times \frac{1}{1 + e^{-\left|x_i - e_u\right|}}, & \text{if } x_i > e_u
\end{cases}
$$

(1)

$$
y_i = \begin{cases} 
W_i, & \text{if } x_i \leq e_u \\
-1 \times W_i \times \frac{1}{1 + e^{-\left|x_i - e_u\right|}}, & \text{if } x_i > e_u
\end{cases}
$$

(2)

In the equations, $x_i$ is the indicator value; $W_i$ is the corresponding weight coefficient of parameter $i$; $e_u$ and $e_l$ are the upper limit and the lower limit of parameter $i$.

The value of $y_i$ represents the degree of deviation of parameter $i$ from the standard range. The greater is the degree of deviation, the closer is the value to 1; The smaller is the degree of deviation, the closer is the value to 0.
Based on the proposed computing model, the health degree of each vegetable can be calculated and the summary can be generated automatically\(^4\), which can help consumers to understand the whole growth cycle of the to-be-purchased vegetable.

3. Establishing process of the standard parameter model
Using the method proposed in section 2, the indexes of vegetable growth were compared with the standard reference model to calculate the vegetable health. The establishment of the standard parameter model can be divided into the following five parts as shown in figure 1.

**Figure 1:** Establishing process of the standard parameter model

1) **Accessing database**
We need access database in the server to obtain the required data model of the standard parameters from relevant agricultural authorities\(^3\).

2) **Establishing target indicators**
The target indicators of the standard parameters are to be established based on the database accessed in step 1). At first, the types of planted vegetables were classified, then the target indicators of vegetables were classified. The indicators involve the following items: the accumulated temperature, light, planting quality, the water required, the period required for growth, etc. So the establishment of each target indicator is necessary for the health evaluation of the planted vegetables.

3) **Establishing sub-indicators**
The values of weights of indicators in the standard parameter model are given using AHP. The weights are assigned to each sub-indicators for vegetables. For example, the corresponding proportion of the standard of temperature accumulation for vegetable growth is needed to compute the corresponding part of health scores of the target indicators.

4) **Establishing the standard reference model**
The standard reference model will eventually output an overall evaluation score for the vegetables. The output score describes the health and quality status of the target vegetable that the customers track\(^5\).

5) **Computing score of the target indicators**
The standard parametric model is divided into the health part and the unhealthy part. The traced vegetables are evaluated by consumers based on the score. For example, a customer needs to query the
traceability information of a cabbage information, and the queried score is 88, which indicates the cabbage is in a healthy range.

The establishment of the reference model of vegetables is multi-level and omni-directional. It is an indispensable and important element in establishing and optimizing the traceability system. The purpose of the application of the standard reference model is to effectively trace back the quality, efficiency and development of each vegetable in the planting and production process. Therefore, we must attach importance to the standard parameter model from a strategic perspective. We could effectively promote forward the system progress well and quickly only if we continue to improve the standard parameter model level. In this way, we could effectively improve the quality of vegetables, and enhance the competitiveness of agricultural products in the market.

4 Conclusions
Fact data and derived data are collected for vegetables. They are integrated to produce scores and summaries describing the vegetables. The scores of sub-indicators are computed according to the standard reference model from authorities based on the fact data and the derived data. The target scores are calculated based weights and scores of sub-indicators, and then the score was used to evaluate the health status of the vegetables. Finally, summaries are automatically generated based on target scores.

The proposed automatic summary generating technology of vegetable traceability for information sharing is a complex engineering system. On the basis of establishing and optimizing the standards of food safety and traceability system, this proposed technology could be accepted by more and more farmers and customers.

Acknowledgements
This research is supported by National Science Foundation of China (Grant No. 71203162), Science and Technology Planning Project of Guangdong Province, China (Grant No. 2014B040404072), Natural Science Foundation of Guangdong Province, China (Grant No. 2015A030313642), and Innovation Project of Wuyi University (Grant No. 2014KTSCX128 and 2015KTSCX144).

References
[1] Nenkova A, Maskey S, Liu Y, et al. “Automatic summarization”, Meeting of the Association for Computational Linguistics, (2011).
[2] Brugmann S, Bouayadagha N, Burga A, et al. “Towards content-oriented patent document processing: Intelligent patent analysis and summarization”, World Patent Information, 2, pp. 30-42 (2015).
[3] Sauper C, Barzilay R. “Automatically generating wikipedia articles: structure-aware approach”, Proceedings of the IEEE, (2009).
[4] Martinezromo J, Araujo L, Fernandez A D, et al. “SemGraph: extracting keyphrases following a novel semantic graph-based approach”, Journal of the Association for Information Science and Technology 67.1, pp. 71-82 (2009).
[5] Burke M G, Knobe K, Newton R, et al. “Concurrent collections programming model”, Parallel Computing, 22, pp. 364-371 (2011).
[6] Sridhara G, Hill E, Muppaneni D, et al. “Towards automatically generating summary comments for Java methods”, Automated Software Engineering, 132, pp. 43-52 (2012).
[7] Qiao S, Wei Z, Yang Y, et al. “Research on vegetable supply chain traceability model based on two-dimensional barcode”, International Symposium on Computational Intelligence and Design, pp. 317-320 (2013).
[8] Hu J, Zhang X, Moga L M, et al. “Modeling and implementation of the vegetable supply chain traceability system”, Food Control, 30, pp. 341-353 (2013).
[9] Aung M M, Chang Y S. “Traceability in a food supply chain: safety and quality perspectives”, Food Control, 114, pp. 172-184 (2014).
[10] Dabbene F, Gay P, Tortia C, et al. “Traceability issues in food supply chain management: review”, Biosystems Engineering, 28, pp. 65-80 (2014).
[11] Zhang J, Bhatt T. “A guidance document on the best practices in food traceability”, Comprehensive Reviews in Food Science and Food Safety 13.5, 4, pp. 1074-1103 (2014).
[12] Thorne S. “Data analysis in qualitative research”, Evidence-Based Nursing, 3, pp. 68-70 (2000).