Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2021 Page No. 131-141
Article History: Received: 05 January 2021, Revised: 02 February 2021, Accepted: 10 March 2021, Publication: 30 April 2021

Cultivation of College Students' Employment and Entrepreneurship Ability of Agriculture and Forestry Engineering Major Under the Background of Internet Plus

Ji Changxin[1*]
[1]Henan Vocational College of Agriculture, 451450 Zhengzhou, China
*Corresponding Author.

Abstract:

In view of the requirements of enterprises for the Talents Innovation of agriculture and forestry engineering major in the employment of students, this paper studies the training of College Students' employment and entrepreneurial ability and the docking platform of University Enterprise Talents Internet plus background. Firstly, this paper proposes a new keyword extraction algorithm based on graph model for college students' innovation ability information document to tap students' innovation potential. The algorithm uses the keywords extracted from the traditional textrank to construct the vertex, and calculates the edge weight based on the time and the number of hits. Finally, the key capability elements are extracted by random walk iteration. This article uses Spring Boot+Mybatis+ Vue.js The whole stack technology has realized a platform of talent docking system of school and enterprise. This paper uses the above algorithm to realize the module of position keyword extraction and user preference path mining. Finally, the system is tested in function and performance. The results show that web data mining algorithm improves the innovation ability of students and the efficiency of talent docking between school and enterprise.

Keywords*: Agriculture and forestry engineering major, innovative demand, Internet plus, employment and entrepreneurial ability, docking between school and enterprise talents.*

I. INTRODUCTION

School enrollment has always been a problem faced by college students when they graduate.

Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2021 Page No. 131-141
Article History: Received: 05 January 2021, Revised: 02 February 2021, Accepted: 10 March 2021, Publication: 30 April 2021

Most of them find jobs through school enrollment. However, the market survey results show that the first employment satisfaction rate of college students is low, 62% is far from their imagination, and only 10% have found the job they really want [1-2].

This shows that, with the optimization and upgrading of the enterprise industry, there is a gap between the current higher theory training mode and the actual operation ability demand of the enterprise, which leads to students not knowing the real demand of the enterprise for talents [3]. For enterprises, on the one hand, there are spring recruitment and autumn recruitment twice a year, and each recruitment fair may cross cities or even provinces, both direct and indirect costs are not small expenses; on the other hand, because students' ability does not match the needs of enterprises, they often have to spend more on pre job training. Moreover, after the recruitment of talents, it is often difficult to communicate and retain talents. And once there is a job hopping situation, the enterprise will spend a lot of cost to recruit again.

In view of the above problems faced by students and enterprises, this system uses enterprise innovation culture as the media to design and build a system platform for talents docking between universities and enterprises, so as to promote mutual understanding between students and enterprises. In addition, in the past few decades, the Internet has been booming. Up to now, people can learn almost any information they want to know through the Internet, which brings great convenience [4-5]. Web data has a huge amount of data, but its structure is scattered, which is significantly different from the data objects of traditional data mining. In order to find regular and valuable information from the above-mentioned incalculable web data, we must first collect the data, which may be a very easy task for traditional data mining, but it is very difficult in Web data mining.

To sum up, it is of great research value and application significance to build a platform for students and enterprises to exchange and understand each other, and integrate Web Data Mining sub platform into it, so as to better provide users with better service and use experience.

II. KEYWORD EXTRACTION ALGORITHM OF POST INFORMATION BASED ON HYPERGRAPH

2.1 Keyword extraction process

The algorithm flow of post information keywords with the same theme based on hypergraph proposed in this paper is shown in Figure 1. Firstly, the algorithm constructs a weighted undirected hypergraph. When constructing a hypergraph, the weights of vertices and hyperedges need to be calculated. The super edge is composed of post information documents,

Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2021 Page No. 131-141
Article History: Received: 05 January 2021, Revised: 02 February 2021, Accepted: 10 March 2021, Publication: 30 April 2021

and the weight of the super edge is calculated by considering the time factor and the click volume factor TextRank can't get the wrong result in the field of keyword extraction, so this paper uses TextRank to extract keywords from preprocessed data. And take the top 20 words of TextRank value as the vertices of hyperedges; After that, the weights of super edges are calculated, then random walks are performed, and finally the keywords are iteratively generated [6]. The process flow is shown in the following figure:

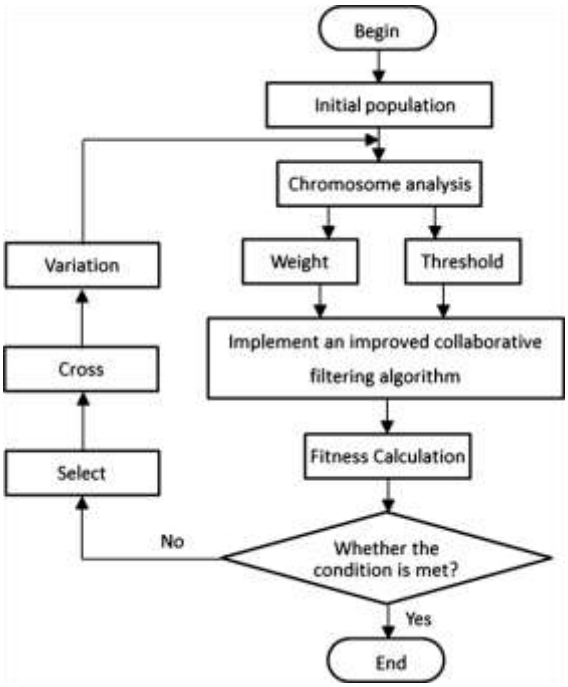

Fig 1: Algorithm flow of post information keywords based on hypergraph

2.2 Data preprocessing

This algorithm uses the data from the Internet recruitment website of the National University data driven innovation research competition in the open research data platform of Peking University, which contains the recruitment information published by 58 citywide and Zhilian recruitment websites in July 2018 [7-8]. Each data includes enterprise name, position name, position information, publishing time, geographical location, industry, welfare, salary, reading amount and other information.

Data preprocessing is the first and most important part of data mining. Here, the original experimental data is processed according to the requirements of the algorithm, and the useless and disturbing data of the algorithm to be executed are eliminated. In the keyword extraction

Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2021 Page No. 131-141
Article History: Received: 05 January 2021, Revised: 02 February 2021, Accepted: 10 March 2021, Publication: 30 April 2021

algorithm described in this chapter, the main work of preprocessing is to do a simple cleaning of the data (mainly to remove noise data), and then use the word segmentation tool for part of speech tagging. The tagged word set will extract the noun or noun phrase to construct the attribute word set, and then preliminarily screen the attribute word set by screening single words and constructing stop words.

(1) Data cleaning

To properly process the original data, the processing goal is to achieve data consistency by deleting redundant data, eliminating information irrelevant to the mining target, and correcting wrong information. For this paper, the purpose of data cleaning is to remove the irrelevant data, incomplete missing data or incorrect data, such as the irrelevant data of post information publishing platform and welfare, to prevent it from interfering with the extraction of post information keywords.

(2) Part-of-speech tagging

Keyword candidates are generally nouns and noun phrases. To identify nouns and noun phrases in documents, we need to use some third-party open source word segmentation tools to mark them. The word segmentation tool used in this chapter is ANSJ word segmentation, which is implemented by Java of ICTCLAS, Chinese Academy of Sciences, and basically rewrites all data structures and algorithms. The dictionary is provided by the open source version of ictclas, and is partially optimized manually. Use this word segmentation tool to find nouns and noun phrases.

2.3Preliminary screening of candidate word set

In the word set obtained by the ansj word segmentation tool, many words can be judged from the structure that they can not be the keywords to be extracted. Therefore, the work in this section can screen out some candidate attributes from the structure of the candidate attribute set, so as to reduce the calculation of theTextRank algorithm.

1) Word screening

According to life experience and previous studies, combined with the characteristics of job information data, it is almost impossible for a single word noun to become a job information attribute. The reason is that there are fewer single word nouns and fewer words that can be used as post information.

Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2021 Page No. 131-141
Article History: Received: 05 January 2021, Revised: 02 February 2021, Accepted: 10 March 2021, Publication: 30 April 2021

2) Stop word screening

Stop words refer to the general words without semantic expression, which are not helpful for keyword extraction of post information and should be eliminated to avoid interference to the following algorithm calculation.

2.4TextRank algorithm

TextRank is a common keyword extraction algorithm based on graph, which holds that if a word appears after multiple words, it is relatively important; If another word appears after a word with a high TextRank value, the TextRank value of this word will increase accordingly. The formula for extracting keywords using TextRank is defined as [9-10]:

$$S(V_i) = (1-d) + d * \sum_{V_j \in In(V_i)} \frac{w_{ji}}{\sum_{v_k \in Out(V_j)} w_{jk}} S(V_j) \quad (1)$$

In this paper, TextRank is used to extract keywords, and the 20 keywords or keyword phrases with the highest weight are selected as the vertices of the post information document hypergraph, and the corresponding TextRank values of keywords are used as the weights of the vertices. An edge of a hypergraph can be generated by these 20 words, that is, a post information document is represented by the 20 words with the highest TextRank value.

III. IMPROVED APRIORI ALL BASED USER PREFERENCE PATH MINING ALGORITHM

3.1Log data format description

The data studied in this chapter is the server log of a news website, and some of the log records are shown in the following figure:

Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2021 Page No. 131-141
Article History: Received: 05 January 2021, Revised: 02 February 2021, Accepted: 10 March 2021, Publication: 30 April 2021

Fig 2: Logging fragment

As can be seen from the figure, each log records a user's request in great detail. The log entry and its function are shown in the following table:

TABLE I. Web log entry description

RECORD ITEM	IDENTIFICATION	SIGNIFICANCE
Request date	date	The date when the client initiated the request, such as December 15, 2013
Request time	time	The specific time requested by the client is accurate to seconds, such as 08: 05: 02
Customer IP	c-ip	IP where the client initiates the request, such as 211.66.184.35
Service IP	s-ip	IP of the server served by the client
Service port	s-port	Port number of service provided by client
Request method	cs-method	HTTP method types, such as Get, Put, Delete, etc.
Requested URL	cs-uri-stem	Specific URL requested by client, excluding domain name
URL parameters	cs-uri-query	Data attached by the client when sending the request
Returned status code	sc-status	HTTP status codes returned to customer requests, such as 404, etc.

3.2 User frequent path mining process based on improved apriorall

The specific mining steps of the improved Apriori all user preference path mining algorithm

Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2021 Page No. 131-141
Article History: Received: 05 January 2021, Revised: 02 February 2021, Accepted: 10 March 2021, Publication: 30 April 2021

proposed in this paper are as follows:

(1) The transaction obtained by data preprocessing is stored in the transaction database;

(2) Support statistics of each URL in the above transaction database;

(3) The frequent 1-itemsets are selected according to the preset minimum support threshold.

(4) Using the optimization algorithm proposed in this paper, instead of scanning the whole data, only the sequences containing 1-itemsets are scanned to generate candidate 2-itemsets, which only contain valid itemsets.

(5) Follow the steps (3) and (4) to continue to find the candidate k-itemsets. When the candidate k-itemsets do not exist, end the whole algorithm process.

(6) After the algorithm stops, all frequent itemsets in the web log can be obtained.

The corresponding flow chart is shown in Figure 3.

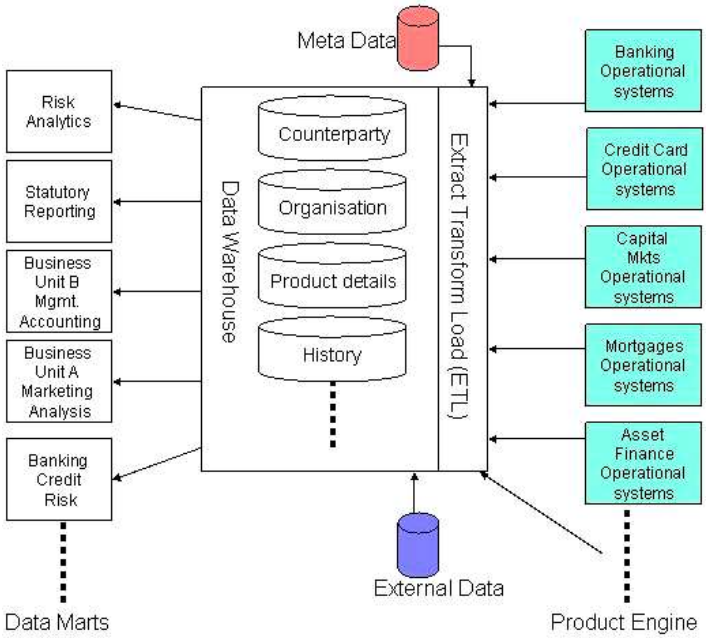

Fig 3: Algorithm flow

Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2021 Page No. 131-141
Article History: Received: 05 January 2021, Revised: 02 February 2021, Accepted: 10 March 2021, Publication: 30 April 2021

IV. SYSTEM DESIGN AND IMPLEMENTATION

4.1 Overall system design

The system adopts browser/server architecture and uses SpringBoot+Mybatis_Vue.js full stack technology introduced in Section 2.3 to complete the development. The front-end uses Vue.js framework, the back-end uses SpringBoot framework, the database is Mysql, and the front-end and the back-end interact through HTTP tool Axios. The system architecture is shown in Figure 4:

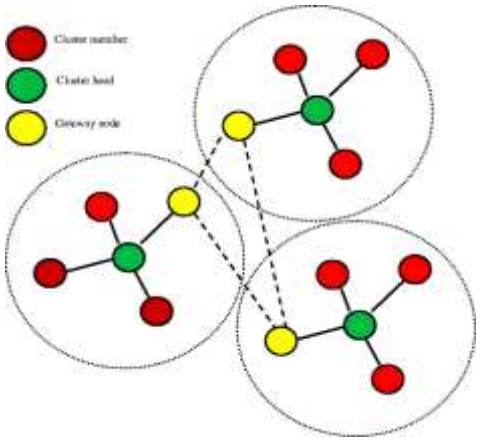

Fig 4: Overall structure diagram of school-enterprise talent docking platform

4.2 Overview of system architecture design

In addition to various types of users, there are also enterprise position information and web server logs. The following describes the role of business processing, position information keyword extraction and web log mining from the perspective of system architecture.

(1) Business processing: the system realizes and processes the functions of system administrators, students, enterprises and other users to ensure the normal use of these users.

(2) Keyword extraction of position information: using the position information filled in by the enterprise, the keywords of each similar position information are extracted, and the keyword database is generated. When the user views a position, the server sends the data to the front-end page. The front-end page displays the word cloud through the Vue wordcloud plug-in. Each item of data in the plug-in is keywords and word frequency, which can just meet the system requirements.

Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2021 Page No. 131-141
Article History: Received: 05 January 2021, Revised: 02 February 2021, Accepted: 10 March 2021, Publication: 30 April 2021

(3) Web Log Mining: mining user preference path from original web server after data preprocessing, and forming preference data set. System management users can find out the most frequently used functions by viewing these preference path sets, and then optimize the network function structure when upgrading the system, and put these frequent functions where users can click most easily, so as to continuously optimize the user experience.

The overall architecture of the system is shown in Figure 5.

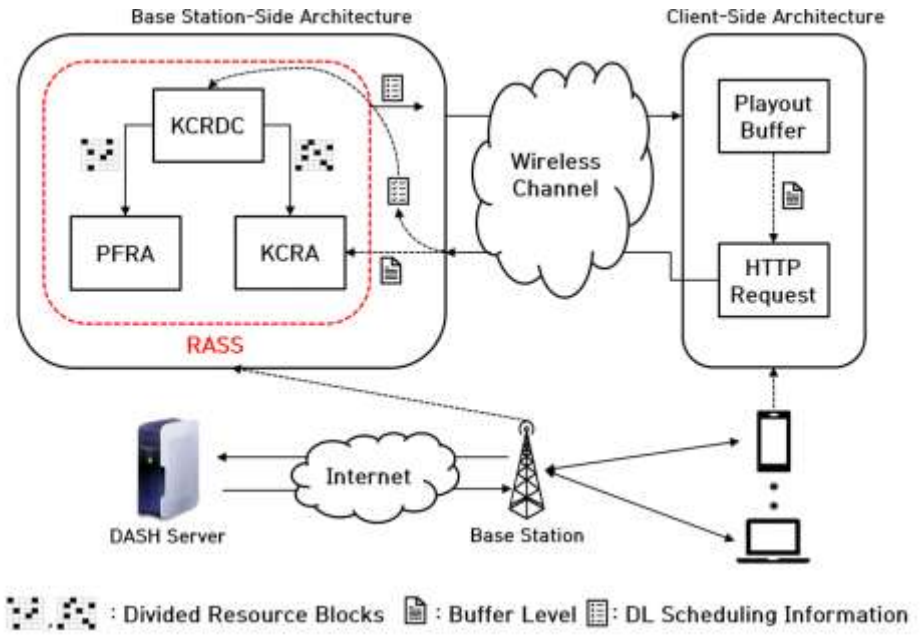

Fig 5: Overall system architecture

4.3Implementation of Web Log Mining sub module

After the above data preprocessing module processes the original data, the improved algorithm realizes the sequential pattern mining of Web log. The module provides the setting of some parameters, such as minimum support. In order to realize the algorithm as simply as possible, this module is divided into four stages: sorting, mining frequent itemsets, sequence and maximizing sequence

(1)Sorting stage. The preprocessed data will be stored in the database. In this step, the tasks are mainly arranged according to the user id (corresponding to the user_id field) and the transaction id (corresponding to the t_id field) according to certain rules, and then the

Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2021 Page No. 131-141
Article History: Received: 05 January 2021, Revised: 02 February 2021, Accepted: 10 March 2021, Publication: 30 April 2021

corresponding data sets are loaded into memory from the database according to these user_id and t_id.

(2) Sorting state. After preprocessing, the data will be stored in the database. The task of this step is mainly based on the user ID (corresponding to user ID)_ ID field) and transaction ID (corresponding to t)_ The ID field is arranged according to certain rules, and then according to these rules_ ID and t_ ID loads the corresponding data set from the database into memory. (2) Candidate set mining stage. As mentioned earlier, before the algorithm is executed, the user will be asked to input a parameter with minimum support (of course, there will be a default minimum support). The task of this step is to search the frequent itemsets satisfying the minimum support in the dataset through layer by layer search after loading the dataset into memory in step (1). Then, using the improved idea proposed in this paper, a new candidate set with only valid itemsets is generated.

(3) Sequence phase. This step is similar to the classical Apriori all algorithm, and its task is to generate new frequent sequence itemsets after the above self join and prior pruning.

(4) Maximize the sequence phase. Step (2) and step (3) above can mine all the sequence item sets that meet the conditions in the dataset after continuous iteration. The task of this step is to find out the sequence item set with the largest length, which is the preference path.

V. CONCLUSION

In this paper, through the actual research, starting from the current pain points of students and enterprises, and taking the enterprise innovation culture as the media, we design and build a talent docking system platform for universities and enterprises. In order to promote the communication between students and enterprises, a system platform is designed to realize the talent docking between universities and enterprises through the media of enterprise innovation culture. But no matter in Web data mining or system, there are still some areas that can be further optimized: when mining frequent sequence sets of Web logs, we can consider clustering users, and take into account the time zone, time and frequency of access analyzed from web logs. It should be able to improve the accuracy of user preference path mining to a certain extent. For the part of the system, although the high concurrency, high availability and other performance characteristics are considered at the beginning of the design, there may still be some shortcomings, which need to do more stress testing; if necessary, it is better to use micro service as the back-end. In addition, the system administrator background function module is too complex, can classify administrators, different administrators for different system management.

Forest Chemicals Revew
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2021 Page No. 131-141
Article History: Received: 05 January 2021, Revised: 02 February 2021, Accepted: 10 March 2021, Publication: 30 April 2021

REFERENCES

[1] Zhang Hui. Internal Economic Mechanism of Industrial Cluster Competitiveness. China Soft Science, 2003 (01): 70-74
[2] Liu Hengjiang, Chen Jixiang. Review of Foreign Industrial Cluster Policy Research. Foreign Economy and Management, 2004, 26 (011): 36-43
[3] Xu Kangning. Industrial Clusters and Competitiveness in Open Economy. China's Industrial Economy, 2001 (11): 22-27
[4] Chen Jiagui, Wang Qin. Sustainable Development of China's Industrial Clusters and Public Policy Choices. China's Industrial Economy, 2005, 9: 5-10
[5] Cai Ning, Wu Jiebing. Network Innovation Capability and Collective Learning Mechanism of Industrial Clusters. Scientific Research Management, 2005 (04): 24-30 + 23
[6] Shen Zhengping, Liu Haijun, Jiang Tao. Research on Industrial Clusters and Regional Economic Development. China Soft Science, 2004, 2: 120-124
[7] Zhao Shuling, Cao Kang. Research on the Relationship Between Industrial Clusters and Urbanization. Henan Social Sciences, 2005, 013 (002): 136-138
[8] Wei Shouhua, Wang Jici, Zhao Yaqin. Industrial Cluster: a New Theory of Regional Economic Development. Economic Jingwei, 2002, 2: 18-21
[9] Jiang Luquan, Wu Ruiming, Liu Hengjiang. Evaluation Analysis and Index System Design of Industrial Cluster Competitiveness. Economic Geography, 2006, 26 (1): 37-40
[10] Wei Shouhua, Liu Guanghai, Shao Dongtao. Characteristics and Strategies of Indirect Financing of Smes in Industrial Clusters. Financial Research, 2002, 28 (009): 53-60