

Principal Component Analysis Method-Based Research on Agricultural Science and Technology Website Evaluation

Jian Ma^{1,2}(✉)

¹ Agricultural Information Institute of Chinese Academy
of Agricultural Sciences, Beijing 100081, China
majian@caas.cn

² Key Laboratory of Agricultural Information Service Technology (2006–2010),
Ministry of Agriculture, Beijing 100081, People's Republic of China

Abstract. Agricultural science and technology website is a very important supporter of driving agricultural information and servicing agriculture. An evaluation method is proposed on agricultural science and technology website based on objective data and artificial ratings, using principal component analysis method. Finally the author used the model to evaluate 18 agricultural science and technology websites, and proposed some suggestions on development of agricultural science and technology websites based on the evaluation result which would act as reference to agricultural science and technology website construction.

Keywords: Agricultural science and technology website · Principal component analysis method · Website evaluation

Agricultural science and technology website is an important window of agricultural science and technology departments and relevant units on the Internet to show their own image, also a concentrated expression of service ability and service level and service features, At the same time also as a important carrier of serviceing “three rural” and promoting agricultural informationization. From the current point of view, although the agricultural science and technology website is rich in resources, but the quality is uneven, how to better improve the website construction is a problem urgent need to solve. A Scientific evaluation method, can make the website administrator to complete understanding of the operation of the site, consummate the existing problems, improve the quality of the website. On the basis of the research of the peer, the principal component analysis method is used, exploring a new method of comprehensive evaluating websites.

1 Overview of Website Evaluation Method

At present, there are no unified standards and methods for the evaluation of the website. The author consulted a large number of literature data, summarized the following two kinds methods of the evaluation of the website:

(1) web link analysis

At present, Link analysis and web impact factor measure related to it are widely used to evaluate the websites. Link analysis through the number of sites are linked to reflect the quality of the site. The evaluation is on the basis that a web site is linked another website is approval and use of this website, and the content of two website is related; the more number of external links of a web site, Explain Its influence is greater. The network impact factor measure is based on the link analysis, reflect the influence of the web site by the size of web impact factor.

Although this method is applied to a wide range of applications, but it also has shortcomings: it only start from the link analysis of the website, not comprehensive evaluation of a website, the authority is to be verified. At the same time, the data used to analysis mostly get from Alta Vista or Google and other search engines, which makes the results depends too much on the search engine, but due to the drawbacks of itself, search engine may not included all external links and internal links of a website.

(2) analytic hierarchy process

Analytic hierarchy process divide the decision problem according to total target, sub target, evaluation criteria until the specific input sequential scheme is decomposed into different levels of the hierarchy and layer by layer analysis, ultimately, get the importance weights that the lowest level factor for the highest level factor.

Analytic hierarchy process can be used to evaluate the site, but the relative importance of each factor in the same level must be evaluated in the construction of evaluation matrix. This will have error inevitably due to the subjective behavior.

Based on the above research, In this paper, we give the evaluation index and calculation method of agricultural science and technology website, and use the principal component analysis method to analyze the evaluation index. Finally, the model is used to obtain the comprehensive ranking of 18 agricultural science and technology websites, and the relevant suggestions are given according to the evaluation results.

2 Principal Component Analysis Method for the Comprehensive Evaluation of the Principles and Methods

Principal component analysis is also called the principal weight analysis, which is an important method to study how to transform the multi index problem into a less comprehensive index. Because there is a certain degree of correlation between multiple variables, people naturally want to extract information from these indicators as quickly as possible through linear combination. Principal component analysis can change the problem of high dimensional space into a low dimensional space to deal with, make the problem become more simple, intuitive, and the comprehensive index of these less interaction and provide most of the information of the original index.

In practical application, the specific steps of principal component analysis are:

- (1) standardization of raw data
- (2) set up the correlation coefficient matrix of variable
- (3) obtain the eigenvalues and eigenvectors of the correlation matrix.
- (4) the number of principal components is determined by the cumulative variance contribution rate, and the principal components are extracted.
- (5) weighted synthesis of principal components, get comprehensive evaluation

3 Evaluation Index and Method of Agricultural Science and Technology Website

The evaluation of agricultural science and technology website mainly investigation from the content of the website, website design and user operation three major aspects. Each of the major aspects of the index selection consider of agricultural science and technology website features. The evaluation model is shown in Fig. 1:

The evaluation method of the index is too dependent on the subjective score, so the evaluation of the importance of the indicators in the evaluation of the importance of a

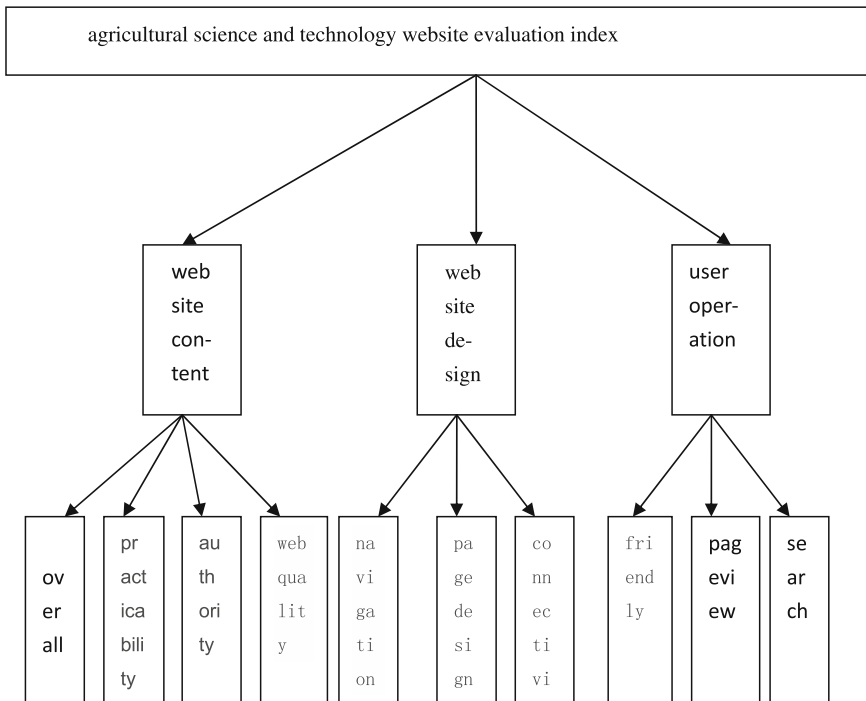


Fig. 1. Evaluation index model of agricultural science and technology website

strong subjectivity. At the same time, simply rely on the objective data can not be a comprehensive reflection of agricultural science and technology website. So we make full use of the advantages of existing systems and professional personnel, the index can be described with the objective data as an objective data, can only rely on artificial score index scoring more than re synthesis method is implemented. The index data is more scientific. Specific practices are as follows:

(1) website content

Web content is good or bad is a valuable key of a agricultural science and technology website. For the content of the site we were from the following four aspects to Investigate:

Comprehensiveness. Comprehensiveness is the breadth of agricultural science and technology website contains content. Web site can be used to quantitatively expressed by the number of web pages. The number of sites webpages can be estimated approximately using the search engine included page numbers.

Practical application. Whether the choice of the content of the site is in line with the “three rural” needs, whether it is suitable for the specific user base, which is suitable for the “three rural” information. This indicator cannot be described by objective data, and it is determined by expert scoring method.

Authority. Authority is the impact of agricultural science and technology sites and the extent of the popularity of the site. This index can be described by using the website of agricultural science and technology anti link number, which the number of links from other sites to this site. The more the number of a website’s anti link, the greater the influence, the greater the authority.

The quality of Webpage. Web quality evaluation can be from the user point of view. Bounce Rate Bounce Rate Bounce Rate Bounce Rate is an important index to measure the quality of web pages. The Bounce rate is the percentage of the total number of visits from a particular portal to visit a site, which only access to one page on the number of visits to the total number of visits. When the site’s bounce rate is high, the quality of the page is very poor, do not attract users.

(2) website design

Good website design should have reasonable structure, the page is simple and beautiful, easy to use. For the website design we investigate from the following three aspects:

Navigation function. Navigation function for the user to use the entire site is essential. Design good navigation function can make the user more convenient, more quick browse information. This indicator is determined by expert scoring method.

Page design. Mainly to examine the page of friendship, include that, the structure is clear, the layout is reasonable, Logo is beautiful, etc. This indicator is determined by expert scoring method.

Connectivity. Effective connectivity mainly investigate the link to the page, if the webpage have broken link or dead link. This indicator can be used to described with broken link rate, broken link rate is the number of all the broken links of the website divided by the number of all links of the website.

(3) user operation

User operation is to measure the website good or bad from the users. For the user operation we investigate from the following three aspects:

- ① Friendly degree. Here is not to examine the degree of friendship from web design, but examine the entire web server response time, and page download time. Obviously, the response time of the server and the web page download times shorter, the higher the site the user-friendliness. Therefore, there are two indicators measure friendliness: server response time, page download time.
- ② User visits. User visits Is the number of user visits the site. We can use the average daily IP traffic number to describe user visits, visitors every page view can be described by Daily average page views.
- ③ Search function. Search function for users to find resources is very important, well-designed site should have a comprehensive in-station search function, which should have advanced search and fuzzy queries. The index score is determined by experts. Detailed indicators and evaluation of the site's content as shown in Table 1:

Table 1. Evaluation System of agricultural science and technology website

Level indicators	Secondary indicators	Evaluation content
Website content	Comprehensiveness	Include the number of pages
	Practical application	Experts rate, standard: meet the needs of agricultural information
	Authority	Number of anti-link
	The quality of Webpage	Bounce rate
Website design	Navigation function	Experts rate, standard: ease of navigation
	Page design	Experts rate, standard:page friendly degree
	Connectivity	Broken link rate
User operation	Friendly degree	Speed (download time, response time)
	User visits	average daily IP traffic, average daily Page view
	search function	Experts rate, standard: the station search function

4 Examples of Application and Analysis

4.1 Evaluation Object

The evaluation object of this paper is 18 agricultural science and technology information website, which has 7 national agricultural science and technology website, 11 provincial agricultural science and technology website, as shown in Table 2:

Table 2. 18 evaluation objects

Number	SiteName	Site url
1	China Agricultural Science and Technology Information Network	http://www.cast.net.cn/
2	Hunan Agricultural Science and Technology Information Port	http://www.hnagri.com/
3	Guizhou Agricultural Science and Technology Information Network	http://www.gzaas.com.cn/
4	Chinese Agricultural Extension Network	http://www.farmers.org.cn/
5	Shandong Agricultural Science and Technology Information Network	http://www.saas.ac.cn/saas.asp
6	Shanxi Agricultural Science and Technology Information Network	http://www.sxagri.ac.cn/
7	National Agricultural Science Data Sharing Center	http://www.agridata.cn/
8	Golden Agriculture Network	http://www.agri.com.cn/
9	Hangzhou Agricultural Science and Technology Information Network	http://www.hznky.com/
10	Liaoning Luyuan Agricultural Science and Technology Information Network	http://www.last.gov.cn/
11	Chinese Academy of Agricultural Sciences Network	http://www.caas.cn/
12	Anhui Academy of Agricultural Sciences Network	http://www.ahas.org.cn/
13	Zhongshan, Guangdong Agricultural Science and Technology Park	http://www.zsnk.com/
14	Hainan Agricultural Science and Technology 110 Network	http://www.hnnj110.com/
15	Nong Bo Technology	http://www.aweb.com.cn/
16	Beijing Agricultural Information Network	http://www.agri.ac.cn/
17	Tianjin Agricultural Information Network	http://www.tjagri.gov.cn/
18	Nine Hundred Million Network	http://www.new9e.com/

4.2 Data Sources

Data sources of this paper are the following four:

- (1) network technology resource monitoring, analysis, evaluation system.

The system long-term monitoring 18 agricultural science and technology website in Table 2, the system provided data included the number of pages, the number of anti-link, download time and the response time four indicators.

- (2) artificial: This paper invited 5 of the author's colleagues, 5 ordinary users to form a group of experts. 5 of the author's colleagues engage in agricultural science and Technology Information Research for many years, while the other 5 are ordinary users that often used agricultural science and technology website.

Respectively, usability, navigation, page layout and search functions independently scoring four indicators, score interval [0,1], 1 on behalf of full marks, 0 on behalf of 0. For each item of data to remove one of the highest and lowest scores, then averaging the remaining data is the index score.

- (3) Alexa: Alexa is a site that specialized publishing website ranking. Alexa every day collect more than 1000 GB of the information, not only give billions of web site links, but also rank for each of the sites. It can be said, Alexa is currently has the largest number of URL, released ranking information the most detailed site.

Alexa provides the data for: average daily IP access, the average daily amount of page view browsing and bounce rate.

- (4) Chinaz (Chinese webmaster station): is a specializes in providing information for Chinese site, technology, resources and services of the website, website existing millions of users.

Chinaz provides data for the number of broken links of the site and the total number of links to the site.

4.3 Data Extraction

Extracting data from the four data sources mentioned in Sect. 4.2, after finishing get the relevant index data of 18 agricultural science and technology site as shown in Table 3:

Table 3. Index data of 18 agricultural science and technology website

Number	Include page number	Practicability	Anti link	Bounce rate	Navigation function	Page design	Broken link rate	Download time (s)	Response time (s)	Daily IP access	Dailypage view browse	Search function
1	937	0.85	2540	33.30 %	0.9	0.92	0.146	1.18	0.02	360	2160	0.83
2	2140	0.73	1470	81 %	0.7	0.51	0.026	0.619	0.44	780	1170	0.82
3	2520	0.9	1420	78 %	0.88	0.9	0.102	0.494	0.27	300	300	0.65
4	20700	0.73	2480	60 %	0.9	0.85	0.037	0.445	0.22	300	300	0.87
5	4990	0.8	2680	37.50 %	0.85	0.8	0.4	0.864	0.01	780	4680	0.76
6	2470	0.85	1830	77.80 %	0.78	0.85	0.048	4.909	0.23	420	420	0.25
7	1060	0.87	1190	12.50 %	0.79	0.74	0.021	0.623	0.06	1500	180000	0.98
8	16400	0.92	723000	50.90 %	0.81	0.92	0.09	1.034	0.04	9600	37440	0.84
9	1040	0.78	349	50 %	0.85	0.74	0.225	0.784	2.77	180	230	0.83
10	4360	0.72	691	50 %	0.86	0.8	0.032	6.599	0.07	250	270	0.2
11	6800	0.88	20100	34.10 %	0.97	0.9	0.076	0.541	0.01	6000	23400	0.8
12	477	0.72	353	67 %	0.85	0.85	0.017	0.848	2.87	240	280	0.7
13	116	0.68	47	84 %	0.6	0.64	0.07	1.338	3.89	180	200	0.82
14	728	0.54	503	60 %	0.78	0.82	0.164	3.611	0.1	120	120	0.75
15	7580	0.71	79500	35.30 %	0.85	0.9	0.1	2.127	0.04	64200	46680	0.8
16	7080	0.68	18500	57.90 %	0.88	0.83	0.04	2.539	0.05	660	990	0.75
17	873	0.57	2360	54 %	0.95	0.8	0.069	12.244	0.08	160	220	0.85
18	2760	0.95	1560	18 %	0.8	0.75	0.09	0.829	0.17	180	210	0.8

Note: the number column is one corresponding to the number in Table 2.

4.4 Comprehensive Ranking of Agricultural Science and Technology Website Based on Principal Component Analysis

In this paper, using SPSS13.0 execute the principal component analysis, and the specific steps are:

- (1) standardization of raw data

In principle component analysis method, the standard method is Normal standardization, and for the practical application, according to the difference of the index, divided the bigger the better and the smaller the better. So we take the following standard method:

With m evaluation object, n evaluation index. All data constitute a $m \times n$ order matrix $X = (X_1, X_2, \dots, X_n)$. Where X_i is a $m \times 1$ dimensional vector. Let $\max(X_i)$ is the maximum value of X_i , $\min(X_i)$ is the minimum value of X_i . $\max(X_i) = (\max(X_i), \max(X_i), \dots, \max(X_i))^T$ is a $m \times 1$ dimension vector, $\min(X_i) = (\min(X_i), \min(X_i), \dots, \min(X_i))^T$ is the $m \times 1$ dimension vector, For the bigger the better type index, the formula is:

$$X_i^* = \frac{X_i - \min(X_i)}{\max(X_i) - \min(X_i)} \quad (4.1)$$

For the smaller the better type index, the formula is:

$$X_i^* = \frac{\max(X_i) - X_i}{\max(X_i) - \min(X_i)} \quad (4.2)$$

In this paper, $n = 12$, $m = 18$. The bigger the better type included web number, practical, the number of anti link, navigation, page design, average daily IP access, the

Table 4. Characteristic root and variance contribution rate table

Total Variance Explained						
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.008	25.069	25.069	3.008	25.069	25.069
2	2.076	17.300	42.370	2.076	17.300	42.370
3	1.525	12.710	55.080	1.525	12.710	55.080
4	1.287	10.726	65.806	1.287	10.726	65.806
5	1.125	9.371	75.177	1.125	9.371	75.177
6	.836	6.966	82.143			
7	.725	6.042	88.185			
8	.586	4.882	93.067			
9	.306	2.549	95.616			
10	.272	2.268	97.883			
11	.159	1.326	99.210			
12	.095	.790	100.000			

Extraction Method: Principal Component Analysis.

average daily amount of page view browsing and search functions eight indexes. The smaller the better type included bounce rate, broken link rate, download time and response time four.

(2) Use SPSS 13.0 for factor analysis

Get the standardized data input SPSS data editing window, the 12 indicators were named X1 ~ X12. Select Analyze->Data Reduction->Factor menu item in the SPSS windows, Tune out factor analysis main interface and move the variable X1 ~ X12 into the variables box, click the OK button, execute factor analysis process. Get the characteristic roots and variance contribution rate table shown in Fig. 4 and factor loading matrix as shown in Table 5:

Table 5. Factor load matrix

Component Matrix ^a					
	Component				
	1	2	3	4	5
X1	.582	-.074	.604	.083	-.175
X2	.526	.383	.146	-.363	.564
X3	.484	.140	.637	.130	-.079
X4	.690	.202	-.581	.001	.047
X5	.587	-.585	-.250	-.185	-.044
X6	.684	-.471	.128	-.174	-.037
X7	-.119	-.054	.262	.704	.451
X8	.223	.735	.237	-.376	.019
X9	.634	-.380	-.173	.129	.204
X10	.400	.011	.024	.411	-.471
X11	.424	.511	-.349	.489	.227
X12	.252	.592	-.197	.078	-.495

Extraction Method: Principal Component Analysis.

a. 5 components extracted.

In Table 4 in total column is each factor corresponding characteristic root, In this case extract five common factors; % of Variance column is the variance contribution of each factor; Cumulative % column is the cumulative variance contribution rate of each factor, As can be seen, the first five factors can explain 75.177 % of the variance.

(3) using factor analysis result execute principal component analysis

Get Table 5's data input SPSS Data Editor window, Column named a1 ~ a5. Then click on the menu item Transform->Compute, Recall Compute variable dialog, Enter $z1 = a1 / \text{SQRT}(3.008)$ in the dialog box, Click the OK button, you can get a

Table 6. Feature vector matrix

	z1	z2	z3	z4	z5
x1	0.336	-0.051	0.489	0.073	-0.165
x2	0.303	0.266	0.118	-0.32	0.532
x3	0.279	0.097	0.516	0.115	-0.074
x4	0.398	0.14	-0.47	0.001	0.044
x5	0.338	-0.406	-0.202	-0.163	-0.041
x6	0.394	-0.327	0.104	-0.153	-0.035
x7	-0.069	-0.037	0.212	0.621	0.425
x8	0.129	0.51	0.192	-0.331	0.018
x9	0.366	-0.264	-0.14	0.114	0.192
x10	0.231	0.008	0.019	0.362	-0.444
x11	0.244	0.355	-0.283	0.431	0.214
x12	0.145	0.411	-0.16	0.069	-0.467

first feature vector. Similarly, it can be calculated for all the feature vectors, which results are shown in Table 6:

By multiplying the matrix of the original data with the eigenvector matrix, can obtained the 5 principal components Y1–Y5. Then the 5 main components weighted comprehensive, you can get the comprehensive score of the agricultural science and technology website, Specific data shown in Table 7.

The formula for calculating the comprehensive score is

$$Y = \frac{\lambda_1}{\sum_{i=1}^5 \lambda_i} Y_1 + \frac{\lambda_1}{\sum_{i=1}^5 \lambda_i} Y_2 + \cdots + \frac{\lambda_1}{\sum_{i=1}^5 \lambda_i} Y_5 \quad (4.3)$$

Y is a composite score, Y_i is the main component i, λ_i is the characteristic root of the main components i.

4.5 Analysis and Suggestion

We compare the ranking of this paper with the Alexa ranking, and the results are shown in Table 8:

From Table 8 we can get the following results:

- (1) We can see, in this paper the first top seven sites is consistent with Alexa ranking, but the first and the second ranking is not consistent. This is because in Alexa ranking, its information flow rank is predominant factor, the impact of other parameters is very small. At the same time also can be seen, when the site information flow reach a certain degree, the greater the information flow, the better the site of the index to maintain, the higher the overall ranking.

Table 7. Agricultural Science and Technology site principal component scores and order

	Y1	Y2	Y3	Y4	Y5	Score	Order
China Agricultural Science and Technology Information Network	1.958	0.16	-0.088	-0.093	0.539	0.729	4
Hunan Agricultural Science and Technology Information Port	1.441	-0.02	0.08	0.154	0.353	0.556	9
Guizhou Agricultural Science and Technology Information Network	1.221	-0.487	-0.374	0.424	0.154	0.312	18
Chinese Agricultural Extension Network	1.659	0.504	-0.2	-0.141	0.694	0.702	6
Shandong Agricultural Science and Technology Information Network	1.47	0.091	0.201	-0.239	0.644	0.592	8
Shanxi Agricultural Science and Technology Information Network	1.641	0.274	-0.252	-0.543	0.149	0.509	11
National Agricultural Science Data Sharing Center	1.746	0.097	0.432	0.092	0.234	0.721	5
Golden Agriculture Network	2.229	0.415	0.811	0.131	0.397	1.045	1
Hangzhou Agricultural Science and Technology Information Network	1.117	-0.016	-0.14	0.108	0.096	0.373	17
Liaoning Luyuan Agricultural Science and Technology Information Network	0.786	0.575	0.172	0.257	0.449	0.517	10
Chinese Academy of Agricultural Sciences Network	1.977	0.174	-0.06	0.493	-0.087	0.749	3
Anhui Academy of Agricultural Sciences Network	1.264	-0.394	0.018	0.204	0.772	0.459	14
Zhongshan, Guangdong Agricultural Science and Technology Park	0.403	0.753	0.306	0.126	0.181	0.401	16
Hainan Agricultural Science and Technology 110 Network	1.148	0.496	-0.096	-0.327	0.167	0.455	15
Nong Bo Technology	1.866	0.931	-0.579	0.486	0.789	0.901	2
Beijing Agricultural Information Network	1.754	0.168	-0.233	-0.245	0.472	0.608	7
Tianjin Agricultural Information Network	1.166	-0.2	0.265	0.039	0.889	0.504	12
Nine Hundred Million Network	1.034	0.253	0.156	0.0006	0.377	0.477	13

- (2) The websites ranked in 8–18, this article ranking and Alexa have a larger discrepancy, so we can see, Alexa for the websites comprehensive ranking is not high not has a very high reference value.
- (3) Overall, the ranking of the national agricultural science and technology websites is higher than the local agricultural science and technology websites. But there are exceptions, such as nine hundred million network, so the nine hundred million network need to be further improved. Single from the national agricultural science

Table 8. Comparison of the ranking of this paper with Alexa

Website name	Article rank	Alexa rank
Golden Agriculture Network	1	2
Nong Bo Technology	2	1
Chinese Academy of Agricultural Sciences Network	3	3
China Agricultural Science and Technology Information Network	4	4
National Agricultural Science Data Sharing Center	5	5
Chinese Agricultural Extension Network	6	6
Beijing Agricultural Information Network	7	7
Shandong Agricultural Science and Technology Information Network	8	9
Hunan Agricultural Science and Technology Information Port	9	8
Liaoning Luyuan Agricultural Science and Technology Information Network	10	11
Shanxi Agricultural Science and Technology Information Network	11	10
Tianjin Agricultural Information Network	12	14
Nine Hundred Million Network	13	15
Anhui Academy of Agricultural Sciences Network	14	13
Hainan Agricultural Science and Technology 110 Network	15	12
Zhongshan, Guangdong Agricultural Science and Technology Park	16	16
Hangzhou Agricultural Science and Technology Information Network	17	18
Guizhou Agricultural Science and Technology Information Network	18	17

and technology websites: Golden agriculture network, nong bo network and Chinese Academy of Agricultural Sciences network, the comprehensive ranking is higher, the utilization of its website is also higher. From the local agricultural science and technology websites: Beijing, Shandong, Hunan stay ahead, It explained that the Ranking of Agricultural Science and Technology websites are relevant with the local information level and agriculture level. Guangdong, Zhejiang although the level of economic development is higher, the level of agricultural information needs to be further improved.

In this paper, through construction agricultural science and technology websites evaluation model, Using principal component analysis to analysis, For 18 agricultural science and technology site ranking, And based on ranking results comparison with Alexa ranking, giving agricultural science and technology websites analysis and evaluation. Through this type of evaluation, can play an important promote role to agricultural science and technology websites healthy development and agricultural Information Development.

References

1. Sha, Y.Z., Ou, Y.X.: Evaluation of China Provincial websites influence_website link analysis and web impact factor measurement. Information and Documentation Work, Beijing (2004)
2. Li, C., Yang, Y., Ge, Y.: Compare of principal component analysis and analytic hierarchy process in quantitative evaluation of comprehensive index. J. Nanjing Univ. Finances Econ., Nanjing (2005)
3. Du, J., Li, D., Li, H.: Evaluation System of Agricultural Information Websites in China. Jiangxi Agricultural Sciences, Jiangxi (2010)
4. Liu, Y., Li, H.: Evaluation Method and Evaluation Index System of Agricultural Website. Agricultural Network Information, Beijing (2010)