

Analysis and Evaluation the Websites of Agridata Base on Link Analysis

Wang Jian, Wu Dingfeng

► **To cite this version:**

Wang Jian, Wu Dingfeng. Analysis and Evaluation the Websites of Agridata Base on Link Analysis. 7th International Conference on Computer and Computing Technologies in Agriculture (CCTA), Sep 2013, Beijing, China. pp.145-151, 10.1007/978-3-642-54344-9_18 . hal-01220907

HAL Id: hal-01220907

<https://hal.inria.fr/hal-01220907>

Submitted on 27 Oct 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Analysis and Evaluation the Websites of Agridata base on Link Analysis

WANGJIAN^{1,a}, WuDingfeng^{1,b}

¹ Agricultural information institute of CAAS, Beijing 100081, China;

^a wangjian02@caas.cn, ^b wudingfeng@caas.cn;

Abstract. Through data collection tools, the Pagerank, inlink and the total number of retrieved pages of Agridata are computed. By considering website traffic as the flag of network influence levels, the paper analyze correlations between website traffic and Pagerank, inlink and the number of retrieved pages. Finally, the feasibility to regard the web link analysis as a factor to measure network influence of agricultural websites is discussed.

Keyword. Inlink, Pagerank, Link analysis, Agridata

1 background of the research

With the information technology sweeping the globe, agricultural information has been attached great importance to government around the world and the level of agricultural information has become one of the key factors that have passed the agricultural sector competitiveness. As the main symbol of the agricultural basis of the information and level of development, agricultural sites not only are a window of agricultural information, but also become the main channel of agricultural research service for the agriculture-rural areas-farmers. Actually, agricultural sites have also become one of major route for traditional agriculture to modern agriculture

In recent years, Chinese government attaches great importance to the construction of agricultural websites, which makes agricultural information website in China has made rapid development. According to the Ministry of Agriculture Information Center in China, the number of China's agricultural information site is increased from 2200 to 29738. In addition, among the first eight months of 2009, agricultural information sites' number is increasing 8183, which the growth rate is 38 percent, far

higher than the average growth rate of Internet sites. However, although the current agricultural information website in China developing rapidly, it is worth noting that there are still many defects, the most prominent of which is the site of construction quality and service levels are not high. Actually, many agricultural sites are focus on building and light management, which results in low traffic of sites and constrains the long-term development of sites. Therefore, the establishment of a scientific and rational agricultural site evaluation system is important for promoting the agricultural site standardize operations and long-term development, which is also one of the focuses of the research areas of agricultural information.

The agricultural scientific data center (Agridata) is one of the experimental data centers supported by National Facilities and Information (NFII) of Ministry of Science and Technology. Based on agricultural scientific data sharing Standard, Agridata Integrated 12 types of agricultural scientific data which included crop science, animal science & veterinary medicine, agricultural resource and environment, grassland science, food science and standards, etc. these data resources could support the agricultural technology innovation and management decision greatly. The website group of Agridata has eight sub-sites, which is shown in table 1. Up to 2009, Agridata Formed a stable user group including more than 150 group users and 8,000 individual registered users. The total visit count of Agridata was 1.8 million. It showed that Agridata was an important information source of network users in rural areas.

Table 1. sites in the website group of Agridata

Name of websites	URL
The main Center	http://www.agridata.cn
Crop sub-center	http://crop.agridata.cn/
Animal sub-center	http://animal.agridata.cn/
Tropical crops sub-centers	http://trop.agridata.cn/
Fisheries and aquatic sub-centers	http://fishery.agridata.cn/
Grass sub-center	http://grassland.agridata.cn/
Zoning sub-center	http://region.agridata.cn/
Agricultural scientific and technological sub-center	http://stb.agridata.cn/

In this passage, a set of evaluation system is proposed for agricultural information website based on the method of link analysis, which make the website group of Agridata as study object. On the other hand, the way of the log analysis is used for

verifying the science and rationality of the evaluation system. Thus it will provide a reference for evaluation system of agricultural sites.

2 Indicators of link analysis

From the general sense, Link relationship among sites in Internet is very similar to citation relations between traditional literatures. This means that links in the sites like citations in publications can be used as quantitative indicators to measure the relative quality of the site. This method using link analysis can improve existing technologies of information science for evaluating websites. Therefore using link different among the number, distribution, aging and functions can be more accurate estimate websites' characters, the amount of information and information organization. Eventually a more scientific and reliable site evaluation results will be obtained.

In general, indicators of website evaluation include reverse links, outbound links, link, link density, the network impact factor, and so on. In these indicators, "reverse link" is also called inlink which is means that when there is a link in page A to page B, the page A can be view as reverse link for Page B. in this way, during the process of website evaluation, we can make the following assumptions on the link analysis: page A supports page B if page A has reverse link for page B. Thus, the reverse link is an important symbol for the influence and information carried by a website. The greater of the number of reverse links for a website, the higher attention and utilization degree for the resources of the site's information are made, which can prove that the website has the better the quality.

PageRank that is developed from principle of link analysis is an indicator for measuring the importance of websites used by Google. The level of PageRank covers from 0 to 10. The higher PR value of a website, the more attention for the site is made. From the essence of algorithm, PageRank is based on the principle of the democratic nature of the network and reverse links which means that the value of a page is represented by link structure. In other word, a link is viewed as a vote, when page A has a link to page B, we can conclude that the page A voted for page B.

Of course, PageRank is not solely consider the number of voting, but also to analyze the voting website, which means that a vote for a page made by an important website will be strengthen its significance. Since PageRank is a quantifiable number, the

quality of the theme site can be evaluated by calculating the average PageRank for all the pages of the website.

When using the method of link analysis evaluating agricultural sites such as Agridata, targets of the method must be based on the characteristics and main problem of agricultural sites. In this passage, we collect google search results, the reverse link and PageRank values for 8 websites of Agridata to evaluate the quality of Agridata website group. Through the result gotten above and consequent from network access log analysis, the relationship between the site score and site traffic patterns and trends will be found for further research.

3Evaluation the Websites of Agridata base on Link Analysis

3.1 problem proposing

The purpose of this passage is to find the trend or pattern of the relationship between PageRank, reverse link and site visits. On the other hand, whether the method using PageRank and reverse link can assess the influence of agricultural site and then analyzing the influencing factors traffic and the overall evaluation of websites is also the aim of this passage.

3.2 Research Methods

When using link analysis methods for site evaluation, data can be collected by Internet measurement and data mining. In general sense, these two methods for data collection may be used with some tools which include gathering, processing and statistics during the process of link analysis.

In the actual process, the Google search engine and Google toolbar will be used for data collection tools which gather the number of Google engine search results, value of PageRank and the number of reverse link for 8 sub-centers' homepage of Agridata. The specific collection strategy is as follows: for example, when collection data for the main center of Agridata, we get information such as the number of Goolge search results, the time consuming of search and so on through the result page as inputting "www.agridata.cn", the URL of the main center of Agridata in Google search engine. At the same time, when inputting "inlink www.agridata.cn", we can get the value of

reverse link for the main center of Agridata. Furthermore, we can obtain the value of PageRank for main center using the function of Google Toolbar while opening the page which URL is “www.agridata.cn” by the browser with Google Toolbar. On the other hand, as the search engine uses regular collection and updating for maintenance, statistics in different stages of Google will be changed which leads to the value of PageRank and reverse links can be less than actual values. This condition will affect results of site evaluation and data analysis. So while data acquainting, we collect the appropriate data repeatedly in the period of time (three days) and choose maximum values as samples for site evaluation criteria and data analysis

On the other side, we can use the website access log as reference for site evaluation system. These log files will be obtain from the sever of website and their format is W3C(W3C Extended Log File Format). At first, the data of access log is stored in a text format and then is converted to other format such as SAV format (the format file for SPASS) for the convenience of data processing.

During the process of the site evaluation, in this passage, the rate of a website is according to PageRank value of the website, which means that if a site’s PageRank value is higher than the other, then we regards that its rate is higher. When some sites have the same PageRank, their rates is according to the number of reverse link or search results in Google. The site with higher number of reverse link or search results in Google has higher score.

In the area of data analysis, we apply the software of Weblog Expert to process web log files and use Eviews software as data analysis tools for finding the relationship between traffic and score of a web.

3.3 Result of Data analysis

According to the above method, the experiment was collected all kind of data of Agridata website group from 7 to 9 March 2012. With these data, we scored for 8 sites of Agridata website group using the methods described above. Results of this experiment are shown in table 2.

Table 2 network data and site evaluation

Rank	Name of website	Page Rank	Reverse link	Number of search results
1	The main Center	5	6220	59,700
2	Crop sub-center	5	139	34,400
3	Zoning sub-center	5	93	41,000
4	Animal sub-center	4	1190	11,900
5	Fisheries and aquatic sub-centers	4	165	29,600
6	Tropical crops sub-centers	3	295	20,500
7	Grass sub-center	3	85	15,000
8	Agricultural scientific and technological sub-center	3	35	13,200

Using the software of WebLog and Eviews to analyze the relationship between score results and site traffic, we can find that the site's ranking has reflected the level of site traffic in a certain extent. In other word, it has a high linear correlation between the site's ranking and site traffic through regression analysis with Eviews. As shown in figure 1, Location of the point is mostly close to the line, which means that the score for a web using the method based on reverse link mentioned in this passage reflects the level of the site traffic and thereby embodies popularity of the site.

With bivariate correlation to analysis the relation the sites ranking which is gotten by the method based on Reverse link and PageRank and site traffic, we can find that pearson correlation coefficient R^2 between these two values is 0.683, correlation close to 0.7, while the explanatory power of the score value on site traffic to 0.596 which shows that there are large correlation between the site score and site traffic. Thus, we can conclude that site score can largely reflected the level of site traffic and has as a high statistical significance. It can be used as indicators to measure site traffic or influence.

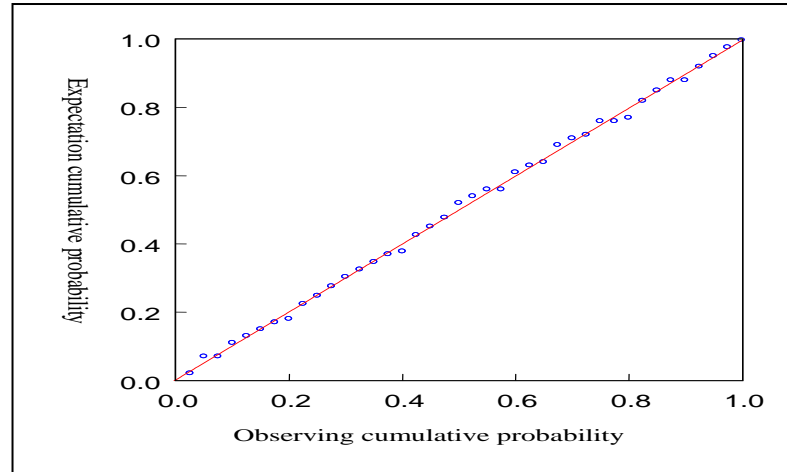


Fig1 Distribution map of normal regression

4 Consideration and enlightenment for promoting the site of Agridata construction

The evaluation method for Agricultural sites based on link analysis is able to correctly reflect the site's traffic, and thus reflects the influence level of the agricultural site. According the evaluation result for Agridata website group and log analysis combining with development requirements of Agridata website, we think that there are several aspects for promoting the site of Agridata construction.

(1)The member of Agridata website group should strengthen the work of maintaining its content to improve the quality of content. When they lined to each other, the link method should be standardize and link points could be clear, which might increase website reverse the number of links, improve the PageRank value and enhance impact factor of websites in a certain extent. At the same time the influence of Agridata is promoted.

(2)Agricultural data resources Agridata provided should be standardized. It means that Agridata website group should provide professional, personalized information services to meet the the data needs of agricultural scientists producers and manager. On the other side, the web group of Agridata can integrate the system of business, optimize column content and rational plan and organize data resources of Agridata to enhance sites functions. Through this way, website application system of resources

standardization, section vivid, sever personalized is built for upgrading the overall image of Agridata.

(3)Agridata must allocate the professional category within group members and optimize the organizational structure of the individual site according professional characteristics. Through these methods, the reciprocity between the sites will be enhanced and internal link frequency of sites in the website group is promoted. So the core member of the group which has a top-ranking will play the leading role to increase network influence of Agridata.

5 Conclusion

Link analysis can be detailed and comprehensive analysis of resource utilization, special characteristics, influence and the degree of openness of the agricultural site. With this method coordinating with manual score, we can get more comprehensive results which the objectivity and effectiveness will be increased. These results can be used to assess influence and research productivity of the site. On the other hand, the link analysis used in the agricultural site evaluation is still in the initial stage and this method has not fixed pattern and means. Since the influence and resource utilization of the website depends not only on the site's traffic, but also by many other factors, thus the study for the evaluation method for agricultural websites based on link analysis is only an exploratory stage. This means that method is only an auxiliary evaluation methods and its result of the method is an indicator that partly reflects the site influence. We will carry further research and exploration for building comprehensive and deep system for website evaluation.

References

1. Jing Du, Daoliang Li, Hongwen Li. The study on the system of China's agricultural information website evaluation. ,Acta Agriculturae Jiangxi,2010,23(3):191-193.
2. Tao Zhou, Xiangdong Chen. The evaluation method for website based on link analysis. Journal of Guangxi College of Education, 2008,96(4):138-141.
3. Xinying An, Na Lv. Research on University Academic LevelBased onWeb Link Analysis. Journal of Medical Informatics,2009,30(10):38-41.
4. Wang Yi,WangSuo Zhu. A Synthetic Approach to E-government Website Evaluation

Based on Web Log. Information Science,2007,25(10):1495-1499.

5. Zhang Yingjie, Leng Fuhai. S & T Websites Influence Evaluation Based on Social Network Analysis. Library and Information Service,2011,(12):56-60.
6. Huanghe Fang,Sun Jianju. Study on Website Evaluation based on Link Analysis:A Case Study of the Four Portal. Journal of Intelligence,2011,30(1):74-76.
7. Lili Song. Study and Application of Hyperlink Analysis Method in Website Evaluation. Journal of Library and Information Sciences in Agriculture,2009,21(6):15-18.
8. Juping Qiu. Study on HyperLink Analysis and Website Evaluation.WuHan: Wuhan University,2004.
9. Li Jiang. Analysis on the Defects of Link Indicators Applied to Web Site Evaluation. Library Journal,2008,27(5):40-44.
10. Maryellen,MottAllen. Six degrees of affiliation: the small world of search tools.Online,2000, 24(5): 49-51.