

Construction of Knowledge Map on Cloud Platform of Agricultural Technology Extension

Wang Liang, Li Xiufeng, Wang Wensheng, Sun Zhiguo

Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing 100081 China

email: singlelight@foxmail.com

Keywords: Agricultural Technology Extension; Cloud Platform; Knowledge Map; Construction

Abstract. The Cloud Platform of Agricultural Technology Extension (CPATE) is a comprehensive knowledge platform serving for the agricultural technicians all over China. Along with the development of the project and the rapid increase of data quantity, the platform urgently needs an effective knowledge management tool to help the technicians in finding the most needed knowledge in a short time. This paper introduces the development and characteristics of knowledge map and puts up an idea about applying knowledge map to the field of agricultural science, and designed a scheme of constructing knowledge map in CPATE, including design of knowledge map architecture, extraction of location and content attributes for static knowledge nodes, social tagging system for dynamic knowledge nodes, design of knowledge network association algorithm. A strategy of showing the maps was illustrated in detail.

Introduction

Knowledge map was proposed the first time by information scientist Bertram Claude Brookes in his classic "Information Science Foundation". The concept is still not a uniform definition. But from many related papers, we can find that knowledge map follows the following characteristics: 1) Knowledge map is a guide to the sources of knowledge. It is not some specific knowledge, but the knowledge about the sources of knowledge[1]. 2) Knowledge map takes revealing the relationship between knowledge as its core content, and creates new knowledge in this way. 3) Knowledge map will be displayed to the users in an intuitive and graphical form. 4) Its ultimate goal is to achieve knowledge sharing and knowledge discovery[2].

Along with the development of computer technology and popularization of the concept of knowledge management, knowledge map as an effective tool for knowledge management, has been applied from the field of library to many other fields, such as Products Design, Process Design, Human Resource Management, and Education[3]. However, in the field of agricultural science and technology, especially in the field of agricultural technology extension, in which an efficient knowledge management tool is needed, the Knowledge Map is not implemented.

The Cloud Platform of Agricultural Technology Extension(CPATE), abbreviation of Cloud Platform of the National Grassroots Agricultural Technology Extension Informatization Based on Modern Information Technology, is a knowledge service platform about Agricultural Technology Extension developed by Agriculture Information Institute of CAAS. The platform builds a network environment where grassroots agricultural technicians and other related social forces can get opportunities of initiative participation, self learning and self creation[4]. Besides, it makes full use of modern mobile communication technology that makes agricultural technicians be able to help farmers in solving problems encountered in the agricultural production and management without the worry of time and location.

The application of knowledge map will promote a more effective use of the knowledge in CPATE, and it will provide agricultural technicians a tool to acquire knowledge, organize and present knowledge, and reveal the connection between the knowledges and then realize knowledge sharing and reuse. Based on the requirements of the platform, we define the main function of the knowledge map as the two following ones: 1) The navigation map will lead agricultural technicians to the useful information when they do not know clearly what they need. 2) The recommendation

map works when agricultural technicians have a relatively clear knowledge demand and find some knowledge nodes through retrieves or some other ways. it provides users knowledge recommendation service for their deeper or wider levels of learning. Given the characteristics of knowledge map and the cloud platform, this paper will realize a combination, apply knowledge map into CPATE.

Architecture Design of Knowledge Map

The architecture of the knowledge map contains six layers: data resource layer, knowledge extraction layer, dynamic nodes layer, knowledge association layer, presentation layer and interaction layer.

Data resource layer

Data resource layer provides the basic data for the construction of the map, which mainly comes from the knowledge resources in CPATE. These information resources contain agricultural basic database, agricultural resource database, agricultural dynamic knowledge, interactive information, agricultural case database and the database about agricultural knowledge rules. And these databases involve agricultural knowledges, general knowledges, interactive information, user information etc. Besides, they exist in different formats such as text, log, image or video.

Knowledge extraction layer

Knowledge extraction layer, overlaying the data resource layer, extracts and purifys knowledges of CPATE to get the static knowledge nodes with an unified format. The content of these nodes contains characteristic information and the informaiton of structural relationship with other nodes.

Dynamic nodes layer

Dynamic nodes layer is constructed on the basis of static nodes. It appends the three tuple relation of nodes, tags and users as dynamic attributes to the static nodes by social tagging system. These tags will be constantly updated, and the dynamic knowledge nodes that updated with usage will be constructed.

Knowledge association layer

The knowledge association layer establishes the connection between knowledge nodes by the degree of correlation between nodes calculated with association algorithm according to the characteristic information and tags of knowledge nodes. Then the logic knowledge network will come into being.

Presentation layer

Presentation layer devides logic knowledge network into several segments according to user's personal information and the functional requirements of the platform. Then this layer provides different types or forms of global or local knowledge maps for different needs of users.

Interaction layer

Interaction layer, composed of interactive modules and knowledge learning modules of CPATE, is the medium of presentation layer and user layer. It provides display space for knowledge maps. and operating for users.

The Annotation of Knowledge Nodes

The extraction of static knowledge nodes

At the beginning of the construction of CPATE, a clear knowledge classification system was confirmed according to the knowledge system of agricultural technology extension. The knowledge nodes in the same category have a stronger correlativity. So the category information should be an important property of a node, and the platform reflects the property as a location information. Take the knowledge item “The key point of management of wheat field in spring” as an example, its location in the platform is “Knowledge platform - Highlight of agricultural technology extension - Technology instruction - Wheat”, so this informaiton is the node’s location property.

In addition to location property, the static node should also include the content attributes. Non text resources have no text content except their title. In order to manage heterogeneous resources

uniformly in the system, the content attributes are composed of keywords coming from the extraction of segmentation of the title of the knowledge resource or provided by its author. By integrating all the independent keywords in CPATE, we get the keywords set as $\{W_1, W_2, \dots, W_p\}$. Then the content attributes of each knowledge node can be expressed by a vector of p dimensions as $R_{wk}=(w_{k1}, w_{k2}, \dots, w_{kp})$. w_{kx} is a Boolean value, it means whether the keyword X appears in the content attributes of the node K or not.

The social tagging of dynamic knowledge nodes

In order to ensure that the attributes of knowledge nodes will be updated with the development of agricultural technology and that the knowledge map maintains a dynamic expansion with the increase number of knowledge resources in the platform, dynamic attributes should be added to the knowledge nodes. We can achieve the target by using the method of social tagging. Each knowledge node will be marked by the users, which can not only ensure the knowledge nodes to be dynamic, but also can be used to enhance the semantics of the knowledge nodes[5].

Social tagging, a kind of popular classification, is one of the most prominent features of the Web 2.0 Platform. In this kind of classification, tags come from users, not from experts. Compared to keywords, tags contain more semantic information[6]. We can get more diversified perspectives by combining the information of nodes, tags and users: Users will mark resources with tags; tags are the medium of users and resources; resources and users are connected with a tag network.

No authoritative and normative text classification will be formed in a social tagging system, which makes the tags inaccurate, abusive and lack of hierarchy. Because of the uneven quality of tags, we need to filter out of the tags with relatively low quality by constructing quantification model of tag quality[7]. Integrating all the filtered and independent tags, we get the tags set as $\{T_1, T_2, \dots, T_n\}$. Then dynamic attributes of each knowledge node can be expressed by a vector of n dimensions as $R_{tk}=(t_{k1}, t_{k2}, \dots, t_{kn})$. w_{kx} is a number value, it represents the frequency the keyword X appears in the dynamic attributes of the node K .

Linking the Network of Knowledge Nodes

The knowledge map is constructed according to the connection of attributes between nodes. The attributes of dynamic knowledge nodes include resource identification, location information representing the structural relationship with other nodes, keywords vector coming from the platform resources and tags vector marked by the users.

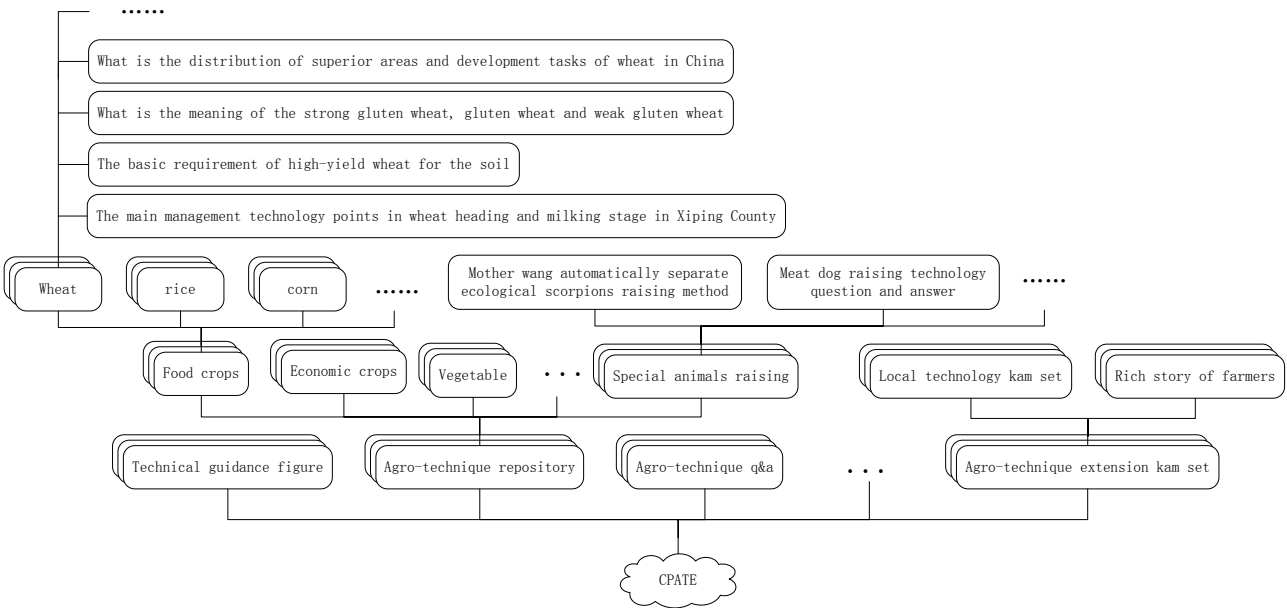


Fig.1. Tree about the distribution of knowledge nodes

Computing location correlation coefficient of nodes

Accordindg to the location information of the nodes, we can build a tree about the distribution of knowledge nodes, as is shown in Figure 1. The Cloud Platform of Agricultural Technology

Extension is the root node of the tree. Branch nodes represent function modules at corresponding level in the platform. All the knowledge resource nodes are represented by the leaf nodes of the tree.

We calculated the distance “d” between two leaf nodes by adding the distances that the two nodes to the same ancestor node with highest level. We can obtain the location correlation coefficient between the two nodes by taking a reciprocal of the distance as $Cor_{loc}=1/d$. For example, the knowledge node “Key point of management technology of wheat in Xiping county during heading and filling stage” and the knowledge node “What are the basic requirements for the soil of Wheat” have a distance of 2, so the location correlation coefficient should be $Cor_{loc}=0.5$; Another example, the knowledge node “Key point of management technology of wheat in Xiping county during heading and filling stage” and the knowledge node “Mother wang automatically separate ecological scorpions raising method” have a distance of 5, so the location correlation coefficient is $Cor_{loc}=0.2$. Obviously, the location correlation coefficient of a knowledge node and its brother nodes is 0.5, and the location correlation coefficient with its uncle nodes is 0.33.

Computing content correlation coefficient of nodes

If two nodes have the same knowledge keywords or tags, we can presume that the two nodes are correlative. In order to quantify the relationship, we use vector space model[8] to calculate the close degree of content correlation between the resources.

We can build keyword matrix and tag matrix as follows by extracting keyword attributes and tag attributes of nodes.

$$R_w = (R_{w1}, R_{w2}, \dots, R_{wm})^T = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1p} \\ w_{21} & w_{22} & \dots & w_{2p} \\ \dots & \dots & \dots & \dots \\ w_{m1} & w_{m2} & \dots & w_{mp} \end{pmatrix}, \quad R_t = (R_{t1}, R_{t2}, \dots, R_{tm})^T = \begin{pmatrix} t_{11} & t_{12} & \dots & t_{1n} \\ t_{21} & t_{22} & \dots & t_{2n} \\ \dots & \dots & \dots & \dots \\ t_{m1} & t_{m2} & \dots & t_{mn} \end{pmatrix}$$

And we can get the resource matrix of the knowledge nodes as $R = (R_w, R_t)$,

$$w_{ij} = \begin{cases} 1, \text{Keyword } W_j \text{ appears in the knowledge node } R_i \text{ at least once} \\ 0, \text{Keyword } W_j \text{ doesn't appear in the knowledge node } R_i \end{cases} \quad 0 < i < m+1, 0 < j < p+1$$

$$t_{ij} = \begin{cases} a, \text{Tag } T_j \text{ appears in the knowledge node } R_i \text{ for } a \text{ times} \\ 0, \text{Tag } T_j \text{ doesn't appear in the knowledge node } R_i \end{cases} \quad 0 < i < m+1, 0 < j < n+1$$

The correlation of knowledge nodes can be calculated by Dice Coefficient of vector space model, as is shown in the formula (1), (2).

$$Cor_w(R_i, R_j) = \frac{2 \sum_{k=1}^p w_{ki} w_{kj}}{\sum_{k=1}^p w_{ki}^2 + \sum_{k=1}^p w_{kj}^2} \quad (1)$$

$$Cor_t(R_i, R_j) = \frac{2 \sum_{k=1}^n t_{ki} t_{kj}}{\sum_{k=1}^n t_{ki}^2 + \sum_{k=1}^n t_{kj}^2} \quad (2)$$

We can get the content correlation coefficient Cor_w calculated from keywords and Cor_t calculated from tags. The bigger the value of Cor is, the closer of the relationship between nodes is.

Building the logic knowledge network

We can calculate the integrated correlation coefficient $Cor(R_i, R_j)$ according to the above three correlation coefficients, as is shown in the following formula:

$$Cor(R_i, R_j) = \alpha Cor_{pos}(R_i, R_j) + \beta Cor_w(R_i, R_j) + \gamma Cor_t(R_i, R_j), \quad \alpha + \beta + \gamma = 1 \quad (3)$$

This coefficient is designed for the construction of an indiscriminate knowledge map, the global resources navigation map. Based on user characteristics, we improve this coefficient to fit it for the resources recommendation map that will meet the demand of personalized service. We calculate the degree of a user concerning about a resource with user weight tags including the tags the user use to mark resources and the tags the user being tagged as a resource. The degree of a user concerning about a resource is represented by μ , μ_{UR} means the times that user weight tags of user U appears in the resource R.

The user oriented correlation coefficient $Cor_U(R_i, R_j)$ can be calculated as the following formula:

$$Cor_U(R_i, R_j) = (1 + 0.1\mu_{Ui}) (1 + 0.1\mu_{Uj}) Cor(R_i, R_j) \quad (4)$$

Set a limit value σ . When the inequation $Cor(R_i, R_j) > \sigma$ or $Cor_U(R_i, R_j) > \sigma$ is true, we presume that there is a strong connection between two resources and link the two resources. We can get a dynamic knowledge network in logic layer by linking all the resources with strong connection.

Presenting the Knowledge Map

In the real world, visualized objects are more cognizable for people. The quantity of information observer can obtain directly from graphs is about 10 times larger than words[9]. Therefore, after the dynamic knowledge network in logic layer is constructed, the way it will be presented to the user is the key point to the construction work of knowledge map. During the construction of knowledge map on CPATE, to enhance the practicality and effect of visualization of the map, attract the user's attention and arouse their strong interest to use the map, we followed the visual psychology principles, select appropriate ways to display the knowledge map according to the types and the themes of knowledge resources, and the characteristics of users:

Firstly, the way of display should conform to the theme and content of the map. In this design, the navigation map will be displayed as an extensible tree map. We take industries and crops as the top categories, and the map can be extended on user's click following the sequence as industries-crops- (fields-) keywords- resources- associated resources- secondary associated resources. The recommendation map will take a form as web map with a central node. The content of the map is a division of the global network. And this map will be displayed in the current resource page according to the characteristics of the user and the resource.

Secondly, the amount of nodes and the radiation levels of the map should be limited, according to the space of the current resource page and the characteristics of agricultural knowledge, to avoid information overload. The display form of global resource navigation map is an extensible tree, which makes it not only contain all the nodes but the display space can be controlled; Resources recommendation map using web map with two radiation levels is enough to display the relationship of nodes clearly.

Thirdly, to attract the reader's attention, stimulate their reading interest and get a good information transfer effect, the overall design of the map must be artistic and aesthetic. At the same time, a reasonable color matching plays an important role to improve the sense of wholeness, highlight the theme and rich the expression.

Conclusion

Now the platform have got a great support from the ministry of agriculture and it will be extended to 5-10 provinces providing service for almost 100 thousand agricultural technicians. Then it will instruct one million agricultural demonstration households and 10 million famers, which will collect a large amount of knowledge for the platform. The knowledge map will play an irreplaceable role in helping agricultural technicians finding their most needed knowledge.

This paper put up a set of scientific, reasonable, feasible design and implementation scheme for the knowledge map on CPATE. This knowledge map will improve the service performance of the cloud platform, enhance the management of agricultural knowledge in the platform, establish the association between the knowledge nodes, and display knowledge resources visually, to guide the users to find the most needed knowledge efficiently, and provide the same theme of knowledge

resources systematically.

With the development of computer technology and network technology, a sea of information has appeared following an increasing number of Internet users. How to make a good use of these resources and how to help users find their really wanted knowledge in these huge amount of information is the burning question for all kinds of knowledge platforms. Knowledge map, as an effective tool for knowledge management, provides a good approach to the efficient use of knowledge in knowledge platforms, and the design and implement scheme illustrated in this paper can provide a reference for them.

Acknowledgement

In this paper, the research was sponsored by Application of Rural Information Service Cloud Storage and Cloud Computing Technology Foundation which is National Science and Technology Support Program (Project No. 2013BAD15B02) and the Scientific and Technological Innovation Project of Chinese Academy of Agricultural Sciences.

References

- [1] Vail III, Edmond F. Knowledge Mapping: Getting Started with Knowledge Management[J]. Information Systems Management,1999,16(4):16-23.
- [2] Li L. Knowledge Map-The Effective Tool for Knowledge Management[J]. Information Studies:Theory & Application,2005,28(3):233-237.
- [3] Zhao J,Xu S T. Key Technologies and Typical Applications of Knowledge Maps[J]. Information Studies:Theory & Application,2012,35(12):101-105.
- [4] Wang W S. Break through the Bottleneck of Innovation and Extention of Agricultural Scientific Research with Modern Information Technology [J]. China Rural Science & Technology, 2012, (7):27-28.
- [5] Ball E C. Annotation an effective device for student feedback: A critical review of the literature. Nurse Education in Practice. 2010,10(3): 138-143.
- [6] Wei J L,Zhu Q H. Advances in Personalized Information Recommendation Based on Social Tagging[J]. Journal of the China Society for Scientific and Technical Information, 2010, 29(4):625-633.
- [7] Li J,Zhang H,Wu H X,et al.. Text Classification Model Framework Based on Social Annotation Quality[J]. Journal of Computer Applications,2012,32(5):1335-1339.
- [8] R T Snodgrass. Computing Temporal Aggregates[C]. Taipei, Taiwan: Proc. of the 1995 Intl. Conf. on Data Engineering, 1995. 222-231.
- [9] He Jie. Advertising & Visual Communication [M]. Beijing:China Light Industry Press,2004.