A Research on the Task Expression in Pomology Information Retrieval

Dingfeng Wu, Jian Wang[(✉)], Guomin Zhou, and Hua Zhao

Agricultural Information Institute of CAAS, Beijing 100081, China
`wangjian02@caas.cn`

Abstract. By analyzing the progresses of pomology information retrieval which are driven by tasks, the author put forward an idea of expressing the tasks in pomology information retrieval. An interview test was used to validate the idea and based on analyzing the data of the test, a method of expressing the tasks in pomology information retrieval, in which the difference of searchers' acknowledgement is considered, was put forward.

Keywords: Pomology information retrieval · Task · Task expression

1 Introduction

With the rapid rise of agricultural websites and databases, the research of agricultural information retrieval is becoming more and more important to agricultural industry and agricultural scientific research. The study on information retrieval is now undergoing a great change from simply focusing on the statistical model and the mathematical algorithm to considering about the situation factors, such as user needs, job tasks, social environment and so on [1]. Within those circumstance variables, the job task is very important [2]. The researchers announced that the users' information retrieval behavior originated from their information needs, and the information needs caused by the job tasks, so the job tasks became the driving factors of the information retrieval behavior [3]. Apperceiving the users' job tasks can help the IR system to understand the users' intention and improve the precision of information retrieval system. To achieve this, the users' abstract tasks must be described in a specific way.

2 The Idea of the Task Expression in Pomology Information Retrieval

By observing and analyzing the pomology information retrieval process, we found that in pomology information retrieval, a user have to master some pomology knowledge which is necessary to complete his job task [2]. By taking that pomology knowledge, a user upgraded his pomology knowledge structure to which we called target pomology knowledge structure of the task. When there is a gap between the target pomology knowledge structure and the inherent pomology knowledge structure, the information retrieval behavior will be triggered to make up the gap. We believe that the pomology

knowledge gap often consists of some sub pomology knowledge gaps and that's why a series of interrelated information retrieval behavior often be observed together. Because every information retrieval behavior corresponds to a retrieval request query, so in pomology information retrieval, tasks can be expressed in to a series of retrieval queries. Figure 1 shows a typical pomology information retrieval process driven by job task. The user input two related retrieval request to make up the knowledge gap and upgraded the inherent pomology knowledge structure to the target pomology knowledge structure. In this process, task eventually expressed as a series of information retrieval queries.
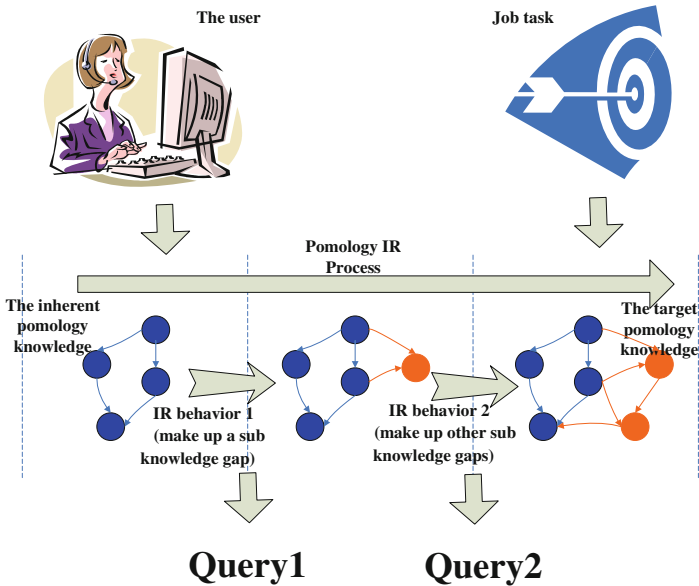

Fig. 1. A progress of pomology information retrieval which is driven by a task

3   Materials and Methods

In order to verify the assumption mentioned above, we interviewed the users of the pomology information retrieval system. In the interview process a large number of simulated job tasks were arranged and at information retrieval process were recorded and analyzed.

The participants including pomology scientific researchers, agricultural experts, scholars and orchard workers, all of whom had the experience of using pomology information retrieval system. The experiments took place at two places, one was the reality experiment site which was at the 101 room of office building of Pomology institute of CAAS, the other one was the Internet remote site, which used the instant messaging platforms, screen recording software, microphones and other video and audio signal capture tools to ensure the real-time communication between the participants and the experiment assistants and record the details of test results and test information behavior.

The setting of the simulation tasks is one of the key problems to be solved of the experiment design. Simulated tasks need to satisfy two conditions: for the first, It must be closely related to the pomology production and scientific research, and it is better to be derived from the practice of production and scientific research in order to be accepted by the participants. For the second, Must have some complexity, in order to identify the change of the ideas of the participants in the IR process, so the simulated tasks should not be single information search tasks but be tasks with practical background and realistic objectives. The source of the work task is provided by the participants themselves, that's because that a lot of typical and impression tasks were accumulated by the participants in their past pomology production and scientific research. The benefit of using the experienced tasks is that the task can be accepted by the participants easily and the retrieval behavior will be simple and intuitive to analyze since the participants are familiar with the tasks. Before the experiment, serval tasks should be prepared in case of the participants can provide few tasks.

In order to extract the intention of every information retrieval behavior smoothly, during the experiment, the participants were asked and guided by the experiment assistants to help them to expose their own aim of every search behavior. The search intention can be exposed in the two part of the experiment. The one is when the retrieval query was typed in, the query directly reflects the information retrieval intention. It should be noticed that sometimes it is difficult to find out a suitable query which can reflect the intention of participants accurately, so a lot of queries are ambiguous, incomplete, or too general. In response to the above situation, the experiment assistants have to ask questions to clarify the intent of the participants, sometimes even help the participants to get queries which are more suitable. The other one is when the participants do the relevance judgment after they get the result list. Usually, the result list is a superset of information and knowledge which the participants need, so the participants can be lead to express the judgement results and evidences which can further exposure the search intentions when the relevance judgments are being carried out.

4   Data and Analysis

There were eleven participants in this experiment who completed twenty-six virtual tasks. During the experiment, seventy-nine information retrieval behaviors are observed.

The retrieval intention can be obtained when analyzing the queries and the relevance judgment behaviors. The users intention can be divided into the following three classes: get the Knowledge points which directly related to tasks (intention 1), get the facts related to the tasks (intention 2) and get the Information that has nothing to do with the tasks (intention 3). If the results of the analysis of the query and the analysis of the relevance judgment behaviors are the same, it is concluded that the conclusion is drawn, but if they are different with each other, the intention obtained from the analysis of the relevance judgment behaviors will be accepted firstly. If there are multiple intentions, It is suspected that multiple information retrieval intents may be merged into one information retrieval behavior and that behavior should be departed into many sub-behaviors to be analyzed. On the basis of analyzing every information retrieval behavior, the participants' intentions can be counted out as shown in Table 1.

Table 1. The statistical results of information retrieval intention

Tasks	Total	Intention 1	Intention 2	Intention 3	Intention 1 ratio (%)	Intention 2 ratio (%)	Intention 3 ratio (%)
Task1	3	3	0	0	100	0	0
Task2	4	3	1	0	75	25	0
Task3	2	2	0	0	100	0	0
Task4	4	4	0	0	100	0	0
Task5	4	2	1	1	50	25	25
Task6	4	3	0	1	75	0	25
Task7	1	1	0	0	100	0	0
Task8	2	2	0	0	100	0	0
Task9	4	3	1	0	75	25	0
Task10	2	2	0	0	100	0	0
Task11	2	2	0	0	100	0	0
Task12	2	2	0	0	100	0	0
Task13	2	1	1	0	50	50	0
Task14	3	3	0	0	100	0	0
Task15	7	3	4	0	42.86	57.14	0
Task16	2	2	0	0	100	0	0
Task17	3	2	1	0	66.67	33.33	0
Task18	2	2	0	0	100	0	0
Task19	4	1	1	2	25	25	50
Task20	2	1	1	0	50	50	0
Task21	2	2	0	0	100	0	0
Task22	4	4	0	0	100	0	0
Task23	4	3	1	0	75	25	0
Task24	4	4	0	0	100	0	0
Task25	3	2	1	0	66.67	33.33	0
Task26	3	3	0	0	100	0	0
Grand total	**79**	**62**	**13**	**4**	**78.48**	**16.46**	**5.06**

Experimental data shows that during the pomology information retrieval process, 80 % of the retrieval behavior is focus on the Knowledge points which directly related to tasks, 16 % of the retrieval behavior is around the facts related to the tasks which can suppose the users to deduce out the knowledge points they need, the remaining 5 % are unrelated to the tasks. Considering about the reason of the information retrieval behavior, we found that job tasks are the most important driving factor which leads to several times more information retrieval behavior than the other factors. Only the task 15 is an exception, that's because in task 15, the participants were asked to find out the treatment method of the apple bagging black spot disease which is rare and lack of research, so the participants have to look for the potentially useful information by searching other similar diseases.

Twenty-six simulation tasks in the experiment include five categories: pomology disease information, pest information, variety information, breeding information and pomology cultivation techniques [4]. We found that different participants will look for different knowledge points when facing the same tasks and this phenomenon might because that their inherent knowledge structures are different. For example, when facing a task which asks the participants to find out the treatment method of a pomology disease, the plant protection professional participants are more inclined to search methods and new drugs, but the other professional participants usually look for the Mechanism and symptoms of the disease firstly. A possible explanation of this phenomenon is that the plant protection professional participants only had few knowledge gaps to make up, so they only need to look for the information which most urgent need to confirm. Relatively speaking, the other professional participants who are not so familiar with pomology diseases have more gaps to fill, so they need to get the basic information first. For the same reason, we noticed that when facing some tasks which are in the same category, a participant might use different Search strategy.

5   Conclusion

Through the above analysis, it is proved that a job task can be expressed as a series of information retrieval intention and further expressed as a collection of queries. If we do not consider the relevant knowledge structure, the job task can be expressed as a collection of queries for all the knowledge points needed for the task. A task can be represented by the following formula:

$$Task \rightarrow \{Query_1, Query_2, \ldots, Query_n\}$$

If we consider about the participant's level of familiarity of the task, the weights should be added to the elements in the collection, if $F_1, F_2, \ldots, F_n$ represents the level of familiarity with the knowledge points and $P_1, P_2, \ldots, P_n$ represents the weights of queries in the collection, the task can be expressed as the following formula:

$$Task \rightarrow \{P_1Query_1, P_2Query_2, \ldots, P_nQuery_n\}$$
$$P_k = 1/F_k \ k \in [1, n]$$

If the user is particularly familiar with the knowledge point, F_k trend approaches infinity, P_K is approaching zero, the expression of the task will not contain $Query_k$. On the contrary, if the user is not familiar with the knowledge point, the weight of $Query_k$ will increase.

References

1. Ingwersen, P., Jarvelin, K.: The Turn Integration of Information Seeking and Retrieval in Context. Springer, Dordrecht (2005)
2. Kelly, D., Cool, C.: The effects of topic familiarity on information search behavior. In: Hersh, W., Marchionini, G. (eds.) Proceedings of the Second ACM/IEEE Joint Conference on Digital Libraries (JCDL 2002), pp. 74–75. ACM, New York (2002)
3. Ingwersen, P.: Information Retrieval Interaction. Taylor Graham, London (1992)
4. 王中英, 果树学概论(北方本), 农业出版社 (1994)