HAL
open science

Research on the Method of Simulating Knowledge Structure of the Information Searchers - Illustrated by the Case of Pomology Information Retrieval

Ding-Feng Wu, Guo-Min Zhou, Jian Wang

To cite this version:

Ding-Feng Wu, Guo-Min Zhou, Jian Wang. Research on the Method of Simulating Knowledge Structure of the Information Searchers - Illustrated by the Case of Pomology Information Retrieval. 7th International Conference on Computer and Computing Technologies in Agriculture (CCTA), Sep 2013, Beijing, China. pp.450-457, 10.1007/978-3-642-54344-9_51 . hal-01220966

HAL Id: hal-01220966

https://inria.hal.science/hal-01220966v1

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.

Distributed under a Creative Commons Attribution 4.0 International License

Research on the Method of Simulating Knowledge Structure of the Information Searchers

— Illustrated by the Case of Pomology Information Retrieval

Ding-feng Wu, Guo-min Zhou, Jian Wang, Jian Wang,

Agricultural information institute of CAAS, Beijing 100081, China

zhougm@mail.caas.net.cn

Abstract: This paper reveals that the users of the pomology information retrieval system possess similar pomology knowledge structures after conducting questionnaire survey and interviews. Why such high consistency appears is explained by the comparison and analysis of the non-pomology-information-retrieval-system users. In the end，Based on the above conclusion, the method of simulating the searcher's professional knowledge structure by using the theory of pomology ontology is released in order to provide a theoretical exploration for the future information retrieval system, which will be more efficient and intelligent.

Key words: pomology; information retrieval; users' knowledge structures; ontology

1. Introduction

Nowadays, the way of researching information retrieval has already evolved from simply focusing on the statistical model and the mathematics algorithm to the stage that retrieval situation, user demand, task and social context are all concerned [1]. The knowledge structure of the searchers is very important among various contextual variables [2]. However, there is no final conclusion on how to actualize or utilize the knowledge structures of the users of information retrieval system [3]. The main reason

remains in that there are huge differences between the knowledge structures of the searchers. It is likely that different searchers have largely diverged knowledge structures so that it is hard for the information retrieval system to get its search ability improved by taking advantage of this contextual variable [4].

Ontology is one of the hot issues of information science study. Ontology is defined as "Shared conceptualized explicit defined terminology" in information science and it is described as a set of domain concepts and the relations between them [5]. As an expression way of shared knowledge, ontology makes the knowledge understandable by computers. At the same time the ontology has the advantage of scalable architecture, high scalability and strong logical reasoning ability [6]. Domain ontology is a set of concepts and relations in a specific field [7]. So far, the pomology ontology is the most effective way of actualizing shared pomology knowledge [8].

Because of the independence and self-contained nature of the pomology knowledge structure and the pomology information searchers are often the researchers in or employees around the pomology field, there may be some rules in the knowledge structures of the users of pomology information retrieval system. With the demonstration and utilization of this regularity we may become successful in actualizing and utilizing the contextual variable of the searcher's knowledge structures, which can improve the efficiency of the search engines

Bases its study on analyzing and exploring the theory and applications in the field, and carries out the experiment by questionnaires and interviews, this paper explored the information of the participants' pomology knowledge structure, and reveals that the pomology information searchers' professional knowledge structures are highly similar with each other. Based on the above conclusion, the method of expressing the searcher's professional knowledge structure by taking advantage of domain ontology is mentioned in order to provide a theoretical exploration for the future information retrieval system, which will be more efficient and intelligent.

2. Materials and methods

The experiment mainly consists of questionnaires and individual interviews which have been conducted on two types of participants. One refers to the participants who conduct pomology information retrieval frequently, known as Test Group 1. The other refers to the ones who rarely conduct pomology information retrieval, known as Test Group 2. Online visual communication and face-to-face interviews are methods

of testing the professional knowledge and skills of participants. By questioning, the pomology knowledge was extracted from the participant's words. By supplemented questionnaire surveys, the participant's pomology knowledge structure was reflected out.

Theoretically speaking, the investigation on the domain knowledge structures, which are often tree structures or net structures, can be carried out by traversing all the concepts. However, sample survey of some relatively important parts is adopted in the practice resulted from the obstacle that the concepts in pomology are too many to go through. The principles mentioned below should be followed when designing questions in the investigation.

Firstly, top concepts should all be selected in order to examine how firm the participant has grasped the knowledge in pomology. And at least one sub-concept should also be selected under each top concept for the same purpose.

Secondly, those concepts with the more relevant concepts in the concept map should be given higher priority as they are likely of more significance most of the time.

Thirdly, the concepts that appear frequently in the production practice should be taken since the investigation is aimed at the users of pomology information retrieval system.

Lastly, general concepts which belong to the nearest 2 or 3 layers to the root concepts should be focused on due to the habit that general concepts are usually used as key words when people conduct an information retrieval.

The principles of selecting concepts are showed in Fig 2-1. Top concepts like Concept 1, 2 and 3 should be selected in Fig 1 according to the selecting principles mentioned above. And Concept 4 as a common concept should also be selected. Concept 9 and 14 with relatively more relevant concepts should be selected accordingly. One of Concept 6, 7 and 8 should be selected in order to follow the first principle. Besides, the completeness of the investigation could be enhanced by the addition of some of the concepts between Concept 22 and Concept 62.

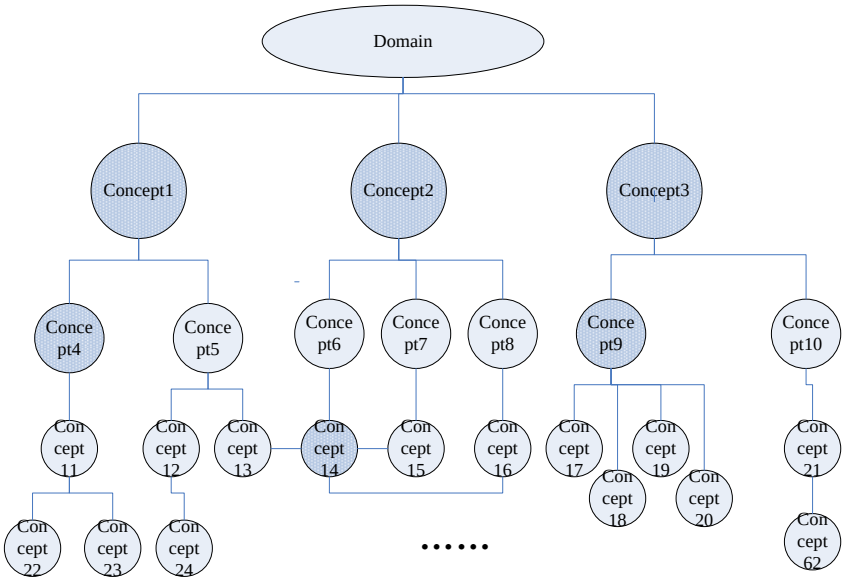

Fig 2-1 The selection of important concept in knowledge structure

Based on the principles above mentioned, 43 of the 486 pomology concepts are selected and incorporated into the 25 questions in the questionnaires [9]. All questions are True or False items for the sake of objective analysis. Besides True or False options, a third option of No Idea is also counted in so as to avoid the coincidence that the participants who know nothing about the relevant knowledge choose a right answer by wild guess. A right answer will score while a wrong answer or a No Idea option won't. The questions in the questionnaire survey are listed in Table 2-1.

Table.2-1 The questions in the questionnaire survey

Concepts	Questions
pomiculture, fruit breeding, fruit tree protection, germ plasm resource of fruit trees, postharvest management	Pomology knowledge is divided into 5 parts known as pomiculture, fruit breeding, plant protection, germ plasm resource, and postharvest management.
fruit breeding	Fruit breeding is a kind of pomiculture knowledge.
fruit tree disease	Knowledge of fruit tree disease is part of pomiculture knowledge
identification of breed	Identification of breed refers to the work of deciding which breed a fruit tree belongs to.

breed	Knowledge of fruit tree breed belongs to fruit breeding knowledge.
stock	Stock is mentioned as a concept in fruit management.
transplant seedlings, full bearing age	Transplant seedlings are conducted during the full bearing age of the fruit trees.
apple preservation	Watering the apples before harvesting helps prevent the apples from going bad.
disease control of fruit trees, pest control, prevention and reduction of natural calamities	Fruit tree protection consists of disease control of fruit trees, pest control, and prevention and reduction of natural calamities like cold injury.
fruit management	Fruit management belongs to the concepts of fruit tree protection.
management of flowers and fruits	Management of flowers and fruits is pomiculture knowledge.
biological control	Biological control is an important topic in pomiculture research.
chemical control, integrated control	The disease and pest control of fruit trees can be divided into chemical control, integrated control, physical control, pesticide control and biological control.
pest control	Pest control is usually carried out in winter.
pruning of fruit trees	The pruning of fruit trees is done intensively in spring.
trimming and thinning of branches, pinching, twisting of tips	Methods of pruning of fruit trees involve trimming and thinning of branches, pinching, and twisting of tips.
fruit coloring	Fuji apples should be exposed to sufficient sunlight so that the fruits will maintain red.
apple canker	Apple canker is a commonly seen disease in apples.
apple anthracnose, apple ring rot	Apple anthracnose is also called apple ring rot because of the ring spots on the fruits.
Carposina niponensis Walsingham	Carposina niponensis Walsingham is not an apple tree pest for it only diets on peaches.
Red Fuji, Fuji	Red Fuji apple is one kind of Fuji apples.
cold resistance of fruit trees, apple industry district	Red Fuji apples are suitable to grow in northeast China in large areas due to their strong cold resistance.
management of rich water, base	Fertilization of fruit trees consists of base fertilizer,

fertilizer, topdressing, composting	topdressing, and composting.
growth period of fruit trees, phonological period of fruit trees	Growth period of fruit trees is also known as phonological period of fruit trees.
grafting, disease resistance of fruit trees, cold resistance of fruit trees	Grafting can create new varieties of nursery stocks but it has no evident influence on the improvement of disease and cold resistance of fruit trees.

3. Results and analysis

The test participants who often conduct pomology information retrieval include science researchers of pomology working in laboratories, science researchers of promoting agricultural technologies in grass-root units and representatives of orchard workers, 16 in total. And there are another 15 participants who rarely conduct pomology information retrieval. The data of the questionnaire survey and the interview is listed in Table 3-1.

Table.3-1 The result of the questionnaire survey and the interview

Participants	Questionnaire scores	integrated analysis of questionnaires and interviews
Test Group 1: Participants who often conduct pomology information retrieval		
Participant 1	96	an experienced science researcher in pomology, working in the fields for years and possessing complete structures of knowledge
Participant 2	100	an experienced science researcher in pomology, working in the fields for years and possessing complete structures of knowledge
Participant 3	100	an experienced science researcher in pomology, working in the fields for years and possessing complete structures of knowledge
Participant 4	92	used to be an excellent science researcher in pomology, working as a manager and possessing complete structures of knowledge
Participant 5	96	used to be an excellent science researcher in pomology, working as a manager and possessing complete structures of knowledge

Participant 6	100	an experienced science researcher in pomology, working in the fields for years and possessing complete structures of knowledge
Participant 7	84	a middle-aged orchard worker, possessing complete structures of knowledge
Participant 8	92	used to be an excellent science researcher in pomology, working as a manager and possessing complete structures of knowledge
Participant 9	100	a doctoral student in pomology, focusing on cultivation and possessing complete structures of knowledge
Participant 10	100	a master student in pomology, focusing on cultivation and possessing complete structures of knowledge
Participant 11	96	a master student in pomology, focusing on breeding and possessing complete structures of knowledge
Participant 12	100	a young orchard worker specialized in pomology, possessing complete structures of knowledge
Participant 13	76	an old and experienced orchard worker with less complete structures of knowledge, or possibly facing difficulty in understanding the terms in the questionnaires
Participant 14	88	a middle-aged orchard worker, possessing complete structures of knowledge
Participant 15	96	a doctoral student in pomology, focusing on postharvest and possessing complete structures of knowledge
Participant 16	92	used to be an excellent science researcher in pomology, working as a manager and possessing complete structures of knowledge
Test Group 2: Participants who rarely conduct pomology information retrieval		
Participant 17	16	being exposed to pomology knowledge in daily work at times and possessing little common sense or any structure of knowledge
Participant 18	64	an expert who comes into contact with pomology in his work often and has acquired much pomology knowledge, but with less complete structures of knowledge
Participant 19	12	having little professional knowledge in pomology, basing his answers on common sense of life

Participant 20	0	having little professional knowledge in pomology, basing his answers on common sense of life
Participant 21	16	having little professional knowledge in pomology, basing his answers on common sense of life
Participant 22	12	having little professional knowledge in pomology, basing his answers on common sense of life
Participant 23	48	having some common sense and a relative as an orchard worker, but with very incomplete structures of knowledge
Participant 24	20	being exposed to pomology knowledge in daily work at times and possessing little common sense or any structure of knowledge
Participant 25	16	having little professional knowledge in pomology, basing his answers on common sense of life
Participant 26	44	having filmed science and educational films in pomology in his early years of life, and accumulated much common sense of life in the years, but with less complete structures of knowledge
Participant 27	0	having little professional knowledge in pomology, basing his answers on common sense of life
Participant 28	24	being exposed to pomology knowledge in daily work at times and possessing little common sense or any structure of knowledge
Participant 29	32	having done research on topic concerning apples and possessing some common sense, but with very incomplete structures of knowledge
Participant 30	28	being exposed to pomology knowledge in daily work at times, and possessing little common sense and less complete structures of knowledge
Participant 31	8	having little professional knowledge in pomology, basing his answers on common sense of life

The analysis of the scores in the questionnaires turns out that the average score for Test Group 1 is 94.25, and that for Test Group 2 is 22.67. The standard deviation, the standard error and the range of the questionnaire scores of Test Group 1 and 2 are showed in Table.3-3.

Table.3-2 The comparison of the questionnaire scores of two test groups

Test Group	standard deviation	standard error	range
Test Group 1	6.85	1.71	24
Test Group 2	17.99	4.65	64

The average score for Test Group 1 is 94.25 while that for Test Group 2 is 22.67, accounting for the fact that most of the participants from Test Group 1 possess relatively more complete structures of knowledge in pomology than those from Test Group 2 do. What is more, the participants from Test Group 2 are in lack of knowledge in pomology, let alone well formed knowledge structures. The standard deviation, the standard error and the range of the questionnaire scores all together indicate that the distribution of scores of the participants from Test Group 1 is centralized, which also means that these participants possess similar structures of knowledge in pomology.

Obviously, the conclusion can also be supported by the integrated analysis of the interviews and questionnaires. Generally speaking, the fact that the participants from Test Group 1 possess higher professional skills and relatively more complete professional knowledge, the knowledge structures of which tend to appear similar, is due to the elements that some of the participants have received years of professional training in pomology, some have abundant practical experience and have kept studying pomology, some have worked for years in the frontline of pomology research, and they are more connected with each other when they are at work, in daily life, studying and doing research, benefiting from numerous opportunities to communicate which are offered or supported by the government, such as Science and Technology Activities to the Country, the promotion of pomology knowledge by the local technicians of the agricultural technological station in the years, and the corporation maintained by the scientific research institutions and the orchard workers in near areas. On the contrary, the similar degree of pomology knowledge of Test Group 2 is small due to that their professional backgrounds, educational backgrounds, interests and life experience are at variance with each other. It is true that some of these participants know a certain amount of pomology knowledge, while some almost know nothing. Therefore, it can be concluded that the participants from Test Group 2 lack professional knowledge as a whole, most of them possess very little common sense in pomology, and some have not formed complete knowledge structures in spite of their reasonable understanding of some of the pomology knowledge.

4. Conclusion

The above experiment has demonstrated that the professional knowledge structures of the users of the pomology search engines are highly similar to each other, which means they tend to converge at the acknowledgeable professional knowledge structures in pomology field. As domain ontology exactly expresses the collection of the structured acknowledgeable knowledge, it is feasible to simulate the knowledge structures of the users of pomology information retrieval system by pomology ontology.

References

[1] Ingwersen, P., & Jarvelin, K.. The Turn Integration of Information Seeking and Retrieval in Context[M].2005.

[2] Kelly, D., & Cool, C.. The effects of topic familiarity on information search behavior. In W. Hersh & G. Marchionini (Eds.), Proceedings of the Second ACM/IEEE Joint Conference on Digital Libraries (JCDL 02).2002. (pp. 74–75). New York: ACM.[3] Ingwersen. P., & Jarvelin. K.. Information retrieval in context. Ingwersen. P.，& Belkin. N. J.(Eds.), Proceedings of the SIGIR 2004 Workshop on Information Retrieval in Context. New York: ACM. 2004.(pp.6~9).

[4] Pennanen, M., & Vakkari, P. Students' cognition and information searching while preparing a research proposal. In H. Bruce, R. Fidel, P. Ingwersen, & P. Vakkari (Eds.), Emerging frameworks and methods. Proceedings of the Fourth International Conference on Conceptions of Library and Information Science (CoLIS4).2002. (pp. 33–48). Seattle, WA: Libraries Unlimited.

[5] Neches R, Fikes R E, Gruber T R, et al. Enabling Technology for Knowledge Sharing. AI Magazine. 1991.8(01). (pp: 36~56).

[6] Lopez V, Pasin M, Motta E. AquaLog: An Ontology-Portable Question Answering System for the Semantic Web. Proceedings of the 2nd European Semantic Web Conference. 2005(pp. 91~126).

[7] Xian-xue MENG .Special Issue – Agriculture Ontology. Journal of Integrative Agriculture. 2012.5(11),