Bulletin of the National Agricultural Research Center 1. 45−56 (2002. 3)
(Bull. Natl. Agric. Res. Cent.)


# Agriculture Case Base

## Decision Support by Harnessing Knowledge and Experience

Akira Otuka[*1], Chikayoshi Kitamura[*2]

## Contents



## Ⅰ Introduction

The sharing and reuse of farming knowledge and experience is very useful and helpful for farmers, especially between those farmers who grow the same crops. By using other's experience, farmers can solve a problem efficiently, or documents describing questions and answers between farmers and extension advisers could give the farmer a good clue as to the solution. This approach to decision-making, experience utilization, is a common way of human problem solving. Actually, farmers have been taking this approach for many years and they will continue to do so. We have, therefore, developed a system, called Agriculture Case Base (ACB), which shares knowledge and experience among farmers and advisers. In this paper, farming knowledge and experience are referred to as agricultural cases. The system retrieves agricultural cases described in textual documents to support user's decisions.

A computer system based on this approach is known as a Case-Based Reasoning (CBR) system. CBR systems were first proposed in the 1980's and much research on CBR and application development, including successful operational systems, was done in the 1990's. Among them, there were some projects

Received: 1 February, 2002
*1 Department of Information Science and Technology, National Agricultural Research Center
*2 Department of Research Planning and Coordination, National Institute of Agrobiological Sciences

in the agricultural domain. Chiriatti and Plant [1] developed a prototype crop fertilization planning system using the CBR techniques of the CHEF system [2]. The Chiriatti's system is a case-based planning system, not a system that finds a solution to a problem. The latter is our target. Hasting et al. [3] developed a system with multiple reasoning paradigms for advising ranchers about the best response to rangeland grasshopper infestation. It uses CBR to estimate forage loss by grasshoppers and it treats cases as collection of attributes with values. In our system, as mentioned before, we handle cases in text style. This is called Textual CBR.

In recent years, there have been a number of Textual CBR works, e.g. FALLQ [4], FAQFINDER [5], SPIRE [6], and etc. All these systems use textual documents as their cases. It is quite natural to address the issue of textual documents from a CBR perspective because many human experiences are stored as texts and used manually in a similar way as in a CBR. We follow the text-based approach.

In the following sections, firstly a definition of agricultural cases and some concrete examples are presented. To determine a favourable structure for ACB, features of the knowledge space that contains all farming knowledge and experience are discussed. Based on the discussion, the system design is presented in the third section. ACB consists of three components and each component is described in detail. In the forth section, various aspects of ACB are discussed, including two types of human decision-making, system performance, and system applicability to the user's community.

## Ⅱ Agricultural Case

In general, a case is defined as *a conceptualised piece of knowledge representing an experience that teaches a lesson fundamental to achieving the goals of the reasoner* [2]. Here, we define an agricultural case (AC) by using this definition with added restriction that the knowledge is related to farming, and the reasoner here is a farmer or extension adviser.

### 1. AC Categories

Examples of ACs include:

- Farmer's experience
- Frequently Asked Questions between farmers and extension advisers
- New farming challenges (both successful and unsuccessful attempts)
- Information on newly developed technology
- Basic farming technologies current farming practices
- Damage information of a meteorological disaster such as the typhoon attack
- Damage information of disease and pest.

Thus, ACs cover a variety of agricultural topics.

### 2. An example of an AC

An example of an agricultural case is shown in this section. The example was drawn from a monthly report called *Local Information*, which is compiled from reports written by local offices of the Ministry of Agriculture, Forestry and Fisheries (MAFF) [7]. The report presents a variety of nation-wide information related to agriculture, forestry and fisheries in local villages.

Pest control with pheromone

Mr. K. Takada (42), who grows rice and apples, has been using pheromones for pest management since last year, under the guidance of the Kasumi Local Extension Center.

This is one of the activities in the center's "Environment Conservative Agriculture Guidance Program". The pheromone used, "Confewza A", is a kind of sex pheromone for apple pests. The pheromone prevents pest reproduction. After 200 stick-shaped pheromones per 10a are put on branches of apple trees, pheromone diffuses throughout the

orchard. It's effects last until harvest.

The center said, "By using this pheromone, apple damage by the pest is reduced. It also reduces the amount of chemicals applied to apple trees, which is favourable for both farmers and consumers who want to buy safe apples. Last year we reduced the number of chemical sprayings from 7 times to only one time. Product condition was good. We are going to promote this effective method intensively this year too."

Contact: 1 Mutsukida, Hanawa,** City, Kasumi Local Extension Center, 0186-23-****

The case describes a successful experience in applying pheromones to protect apple trees. The case has a title and content. The content is divided into four parts: problem (excessive pesticide use), solution (use of the pheromone), result (pesticide reduction) and contact information (address and phone number). Though the case is compactly described, it meets the case definition. The paragraph of problem description specifies which kind of problem the case deals with. The solution and result paragraphs give the approach taken and its result. The case could give a possible solution or a hint to users with similar problem. The contact information in the last section is essential to the case, because users can obtain further information on the case from there. The case description cannot be expected to answer all the questions a user has.

Another source of AC is *Research Information* written by researchers in agricultural research institutes in Japan. Agriculture, Forestry and Fisheries Research Council (AFFRC) provides the research reports at a web site [8]. A case in *Research Information* contains a recent major research result. All cases have a uniform format: title, abstract, background, research result, remarks and contact information. The cases mainly cover well-developed technologies.

## 3. Case Coverage of Knowledge and Experience

Although we believe that the textual CBR approach could assist in solving farming problems, these cases constitute only a small part of all farming knowledge and experience. That is because [1] most cases are still in farmer's brain and adviser's brain and are not written out in text, and [2] new knowledge and experience are being generated continuously. A computer cannot deal with such knowledge and experience. Generally speaking, as the case base stores limited amount of knowledge and experience, its problem-solving ability is also limited. Our approach to overcome this case shortage problem is to present a user with a list of specialists related to the user's problem area. Specialist here means advisers and researchers. When the system does not have a relevant case, it presents a list of relevant specialists. It is rather easy to make a list of agricultural specialists because the number of the specialists is smaller than the number of the agricultural cases. For this reason, we developed an agriculture case base that has two retrieval systems: Case Retrieval System and Specialist Retrieval System.

## 4. Farmers' participation

The cases described in section Ⅱ-2 were made by local offices of MAFF, or by researchers. They, therefore, cover a wide variety of topics and advanced technology. They do not have much farmer-oriented information, or information connected closely to farming fields. From the user's point of view, ACB needs to store cases that are more closely related to actual farming activity. In order to collect farmer-oriented cases, we linked a farmers' mailing list group into the ACB system. This component is known as the Mailing List System. In the mailing list, farmers are actively exchanging farming information. These emails contain field-oriented cases.

# Ⅲ Agriculture Case Base

Figure 1 shows a schematic diagram of ACB. It is a web-based system. A user accesses the system using any commonly available web browser and email software. As described in section Ⅱ, ACB stores three kinds of cases, stored in the Case Retrieval System, Specialist Retrieval System or Mailing List System. Each of these three components is described in the following sections.

## 1. Case Retrieval System

Boolean and exact matching approaches to information retrieval have been widely used. These methods return documents with exactly the same words as in your query. But you might sometimes fail to find what you want because the search words don't match those in the documents that you needed. To make matters worse, a document without query words is not retrieved by the boolean method even if the document is conceptually related to the query.

To overcome this problem, we used Latent Semantic Indexing (LSI), which is a vector-space model of best matching approach [9].

### 1) LSI method

The Case Retrieval System has been implemented using LSI. It uses the statistical structure of terms in the agricultural case.

The LSI method is briefly described. At first, the term-case matrix $X$ is made from the case base. The element $(i, j)$ of $X$ is TFIDF (Term Frequency Inverse Document Frequency) of the term $i$ in the case $j$, where

TFIDF = (term frequency in document)*((total number of documents)/log(document frequency for the term))

The term weighting has some variations, but we used a simple and well-established one. Then the term-case matrix is decomposed by the singular value decomposition into the three matrices:

$$X = T_0 S_0 D_0'$$

Here, $T_0$ stands for a matrix of the left-singular vectors, $D_0$ a matrix of the right-singular vectors and $S_{00}$ a diagonal matrix of singular values. Matrix $X$ is then approximated by keeping only the first 100 to 300 larger singular values and the corresponding columns from $T_0$ and $D_0$ matrices. The operation reduces the matrix dimension.

$$X = T S D'$$

Row vectors of matrices $T$ and $D$ represent term vectors and case vectors, respectively. The vector dimension depends on the dimension of the matrix $X$. The vectors construct a conceptual space of the cases. A query vector is basically calculated as sum of the term vectors that appear in the query.

Generally, *similarity* is defined as an inner product of two normalised vectors. (A product value of the inner product and 1,000 is used as similarity in the following figures.) Saying that two terms are *related to* each other therefore means that the corresponding vectors point out almost in the same direction in the conceptual space. The query result is sorted by the values of similarity between the query and cases.

### 2) System Description

Case Retrieval System is implemented as shown in Figure 1. The system is a web-based system. A query submitted by a user is transferred to the case retrieval servers that calculate the similarity of cases and return a list of cases with higher value of similarity. Figure 2 shows the user interface of the system. Words and phrases in the interface are translated into English and the original interface is in Japanese. The user can submit both keywords and sentences in Japanese natural language and receives the list of the relevant cases. An example of a query result is shown in Figure 3.

### 3) Retrieval Evaluation

To evaluate the retrieval method, we used 2,690 cases from *Extension Information* that was written by the nation-wide extension centers in 1996 [10]. These extension centers are about 600. The centers

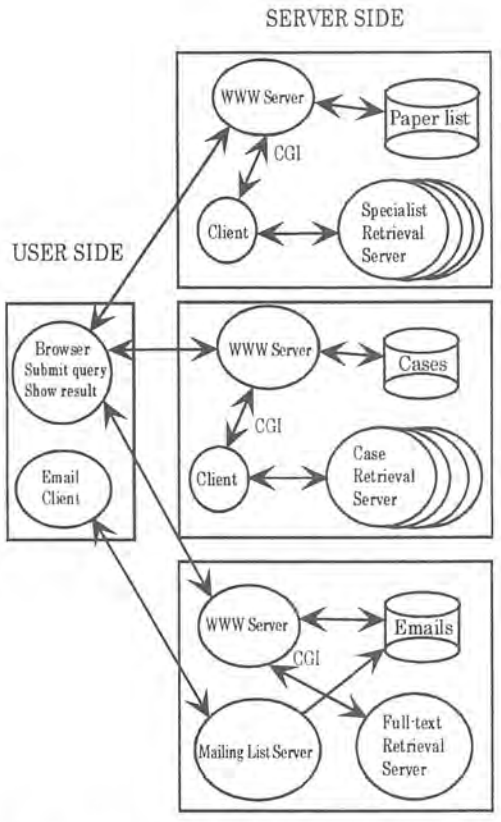


Fig. 1 . Schematic diagram of Agriculture Case Base.

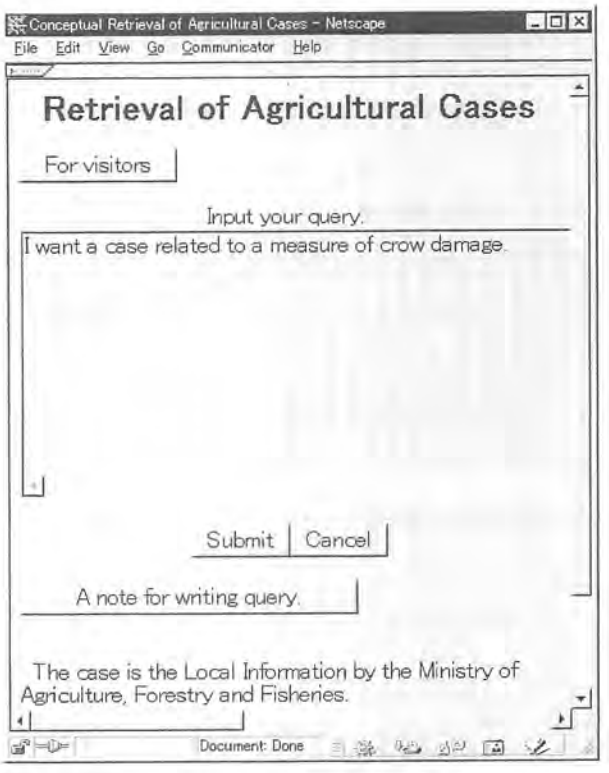


Fig. 2 . User interface of the Case Retrieval System.

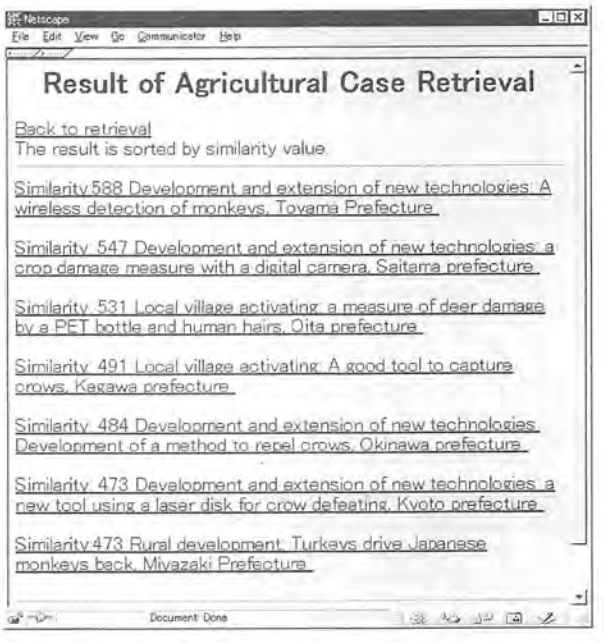


Fig. 3 . An example of the case retrieval result by the query " I want a case related to a measure of crow damage" in Japanese.

report their activities every March. The report format is uniform. The followings are common items:

file name, serial number, 40 kinds of categorisation codes, 28 kinds of information codes, date, title, prefecture, address, reporter, extension center, abstract, descriptions of activity and result

The cases cover every aspect of farming activities.

In the case processing, we firstly had to make a term-case matrix. Only nouns were extracted from the cases to make a term-case matrix. The tool "ChaSen" developed by Nara Institute of Science and Technology was used to analyse Japanese morphological structure of the sentences [11] . "ChaSen" correctly divided sentences into morphemes with the exception of some technical terms. To improve the analysis, we developed Agriculture Term Dictionary, a dictionary of 57,000 agricultural technical terms [12] . Each term has its representation in Kanji characters and phonetic pronunciation. The combined system dictionary contained about 200,000 terms.

The morphological analysis processed the cases and extracted 27,111 nouns. After singular value decomposition of the matrix, the dimension was reduced to 300.

System performance was evaluated by the

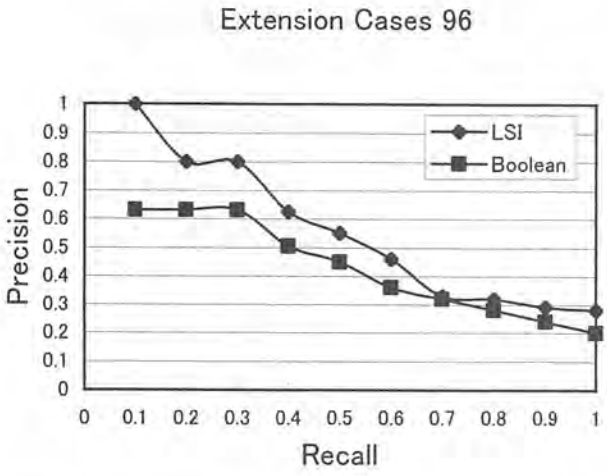


Fig. 4 . An evaluation result of the retrieval method.

Table 1 . An example of the case retrieval. Similarity value here is a product of cosine value and 1,000.

| |
|---|
| Similarity:544 [Title] :Heat control measure of cows |
| Similarity:421 [Title] :Heat control measure of cows |
| Similarity:344[Title] :Effect of a heat measure in dairy farming |
| Similarity:230[Title] :Restructuring of production center in oyster production |
| Similarity:187[Title] :Establishment of cultivation of flower in west-south warm region |
| Similarity:175[Title] :Investigation of feed production, feed supply and quality of milk |
| Similarity:170[Title] :Improvement of management of dairy farming by heat measure |
| Similarity:165[Title] :Measure to summer drought for orange |
| Similarity:162[Title] :Sustainable measure of pig breeder |
| Similarity:160[Title] :Measure to root disease for *Cruciferae* vegetable |

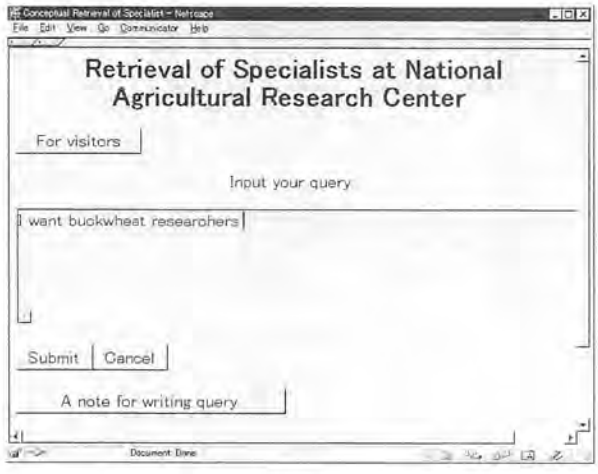


Fig. 5 . User interface of the Specialist Retrieval System.

precision-recall curve, which is commonly used for the evaluation of retrieval systems (Fig. 4). The precision is defined as (number of correct documents retrieved by the system)/(number of retrieved documents). The recall is defined as (number of correct documents retrieved by the system) / (number of all correct documents in the document set). For comparison, an evaluation curve of a simple boolean retrieval system was plotted. The result shows that averaged precision of LSI approach was 20 percent better than the boolean system. The next example shows typical features of the retrieval response. Table 1 shows that some document titles without the query term were retrieved for a Japanese query "防暑" (heat control). The first and second cases have the query term. Although the third and seventh cases have no query term, these cases were related to the query by their contents. This never happens in the boolean system. This feature of the case retrieval improved recall. The feature is discussed more in section Ⅳ.

## 2 . Specialist Retrieval System

The Specialist Retrieval System presents a list of the specialists related to user's problem, which complements the Case Retrieval System as mentioned in section Ⅱ-3. The specialists currently are researchers in National Agriculture Research Center (NARC). As explained in the previous section, a case was characterised by the terms in the case. In the Specialist Retrieval System, the specialist was characterised by the terms in the paper titles that he/she wrote in recent 10 years [13] . The LSI method was used to index the specialists. The system configuration is almost the same as the Case Retrieval System. Figure 5 shows an interface of the system. A retrieval result is a list of researcher's names as shown in Figure 6. When one of the names is clicked, a list of their papers is presented to illustrate the kind of knowledge and experience the specialist has (Fig 7). By using the retrieved information, the user may choose either to read related papers or to ask the specialist about the

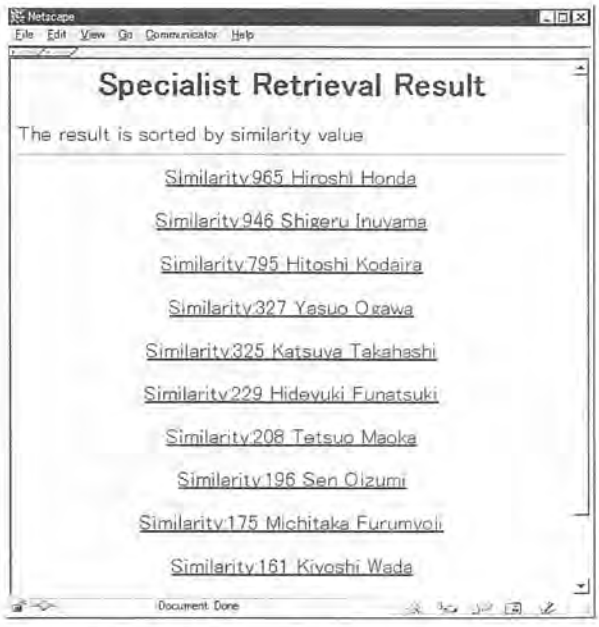


Fig. 6 . An result of the specialist retrieval by the query of "I want buckwheat researchers" in Japanese.

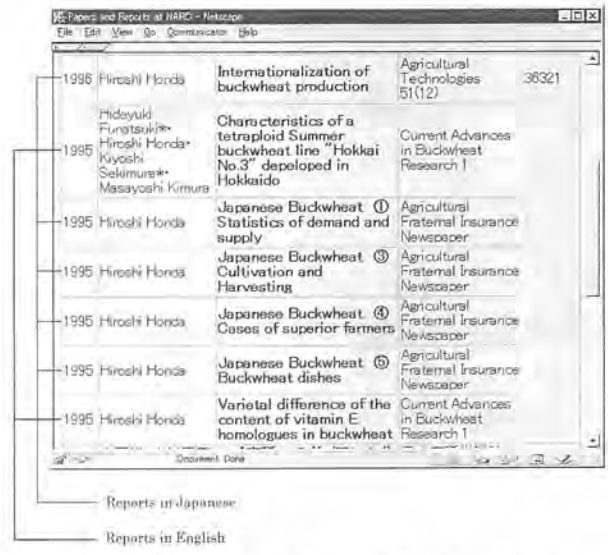


Fig. 7 . A list of papers that is linked from the specialist name of Hiroshi Honda.

problem.

### 3 . Mailing List System

As a result of the prevalence of personal computer among farmers, some mailing list (ML) groups have emerged to discuss specific crops. An example is a mandarin orange ML that collaborated with this research. The members consist of farmers, extension advisers and researchers and company workers mainly in Wakayama prefecture. Emails exchanged in the ML focuses on various things related to the mandarins and Japanese apricots grown in this area. The emails consist mainly of questions, replies to them and administrative communications to members. They are practical, concrete and field-oriented, and are very useful to the members.

The Mailing List System consists of two components: a mailing list server and a full-text retrieval system as shown in Figure 1. The mailing list delivers a new email to all the members. The email is also sent to the full-text retrieval system and is stored automatically. The member can retrieve past emails from a password-protected web site that is accessible only by the members. At present, more than 3,000 emails are stored and reused among the members.

## Ⅳ Discussions

### 1 . Decision making

There are two types of human decision-making. One is the normative approach, and the other is the descriptive approach [14] . The normative approach is based on logic, and it generally is said to be scientific. Decision making with a variety of agricultural crop models is categorised in this type. On the other hand, the descriptive approach is based on everyday experience. Decision-making by heuristics is an example of this approach. The approach of ACB is also descriptive because it is based on farming experience.

It has to be emphasised that they are complementary each other and both are important at the same time, especially in agriculture. Agriculture deals with complex crop ecosystem where many of factors interact each other. A natural characteristic of agriculture, therefore, is that it is complex and hard to predict. In another words, the normative approach alone is limited and insufficient.

## 2. Retrieving related cases without query terms

This section discusses a characteristic of the retrieval method. A simple retrieval system is usually a boolean system that returns document containing a query term. In order to retrieve relevant documents, such a system forces a user to input an appropriate query, which is often difficult when the user is not familiar with the content of the database. In other words, the user has to find a good query for its retrieval requirements. On the other hand, the ACB system has its particular retrieval character. That is, the system returns cases that do not include the query term, because the cases are related to the query. This never happens in the boolean system. This increases the recall value.

Figure 6 shows output resulting of the Specialist Retrieval System from a query for "ソバ" (the Japanese word for buckwheat). Like the boolean system, the three highest ranked specialists had the query term in their paper titles. The 4th and 5th specialists are not related to the query. But the 6th specialist named "Hideyuki Funatsuki" had only "buckwheat" in English in his paper title (therefore the specialist is correct result for the query). At first glance it looks strange. This happens because the specialist vector was compressed in dimension keeping its dimensions with larger singular values while indexing [9]. Hence some specialist vectors became relevant based on the new vector structure. In this case, the first specialist has an English title with a term "buckwheat" among Japanese titles. On the other hand, the 6th specialist has only English title with the term "buckwheat". It is thought that the indirect relationship between buckwheat in Japanese here and English one there makes the specialists relevant each other.

The result indicates that the retrieval method of the system is advantageous to users who are not much familiar with the content of the case base. To some extent, the system decreases the burden on the user by taking into account different expressions with the same meaning while performing a query, because it automatically calculates similarity between these expressions.

## 3. Applicability

The specialists were characterized by the titles of their papers. For indexing, one only needs some characterizing description of a specialist. Therefore the approach can be applied to other types of specialist, such as extension advisers, because they usually make reports on their projects, which probably characterize them well.

Researchers in our institutes were indexed in this paper. In fact, we have other national institutes in Japan. We are preparing to index all the researchers in them.

## 4. Integration of Model Base retrieval and ACB

There is an ongoing Model Base project to develop agricultural crop models that run on Internet [15]. The Model Base stores and provides such models. Once many models become available, then the models have to be retrieved properly by users. But users don't know what kinds of model are available.

This means a model retrieval system is necessary. We developed the model retrieval system, using the same indexing method. A short description of each model written by a model scientist was used for characterisation.

Figure 8 shows a retrieval result. The result contains three kinds of information at the same time: results of model retrieval, case retrieval and specialist

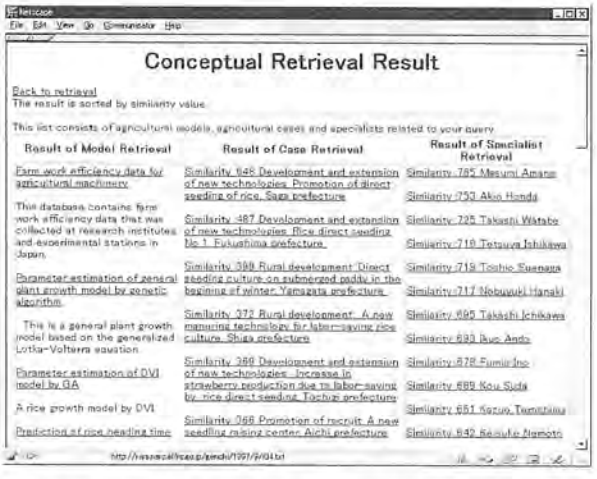


Fig. 8. An integration of Model Base and Agriculture Case Base.

retrieval. Model Base and ACB are integrated by the common retrieval method. By this integration, the user can easily obtain considerable information with only one query. This framework is derived from the fact that the approaches of Model Base and ACB are complementary each other as discussed in the section Ⅳ- 1. This is important from the viewpoint of user convenience.

### 5 . Future works

After an AC is selected by a user as a helpful case and applied to the user's problem, the result of the application becomes a new case. Then the case can be stored back in ACB. This feedback process is called a reproductive process. Through this process, ACB can improve its problem solving ability, storing new cases with different problems, situations, approaches and results. For the ACB system, the next step is to implement the reproductive process and to evaluate its function. Although the mailing list system collects user's feedback, the email is in free format, which makes it difficult to extract cases automatically at present. Emails do not always have all the factors described in section Ⅱ- 2. A mechanism to collect good quality case, particularly field-oriented cases, has to be developed.

In the ACB system, more than 13,000 cases are stored now. But the case number might still be relatively small for users in a specific domain such as soybean growers, vegetable growers and so on. A potato farmer might not find any relevant case when a very specific query was submitted. Specific users might need more specific cases. We still don't know what amount of farming knowledge and experience is needed for operational use. To handle the issue, one approach is to provide the farmer with the relevant specialist's name, which has already been mentioned. The other is to collect cases for a specific domain. Such cases include basic farming technologies, farming experiences, disease and pest information, etc. for a specific crop. We plan to construct the ACB for a specific crop and to evaluate it by measuring the user's response. If a reader is interested in the ACB, please join us to construct a useful case base in your working domain.

## Ⅴ Conclusion

Agriculture Case Base (ACB), a descriptive approach for farming problem solving, was presented. The main points of the paper are summarised as follows:

①Farming knowledge and experience are described in an Agricultural Case (AC). Generally an AC describes new farming challenges, new technologies, FAQs and so on.

②The Case Retrieval System is a web-based system and it retrieves ACs with a query in natural language. Relevant cases are retrieved in order of their similarity.

③The Specialist Retrieval System provides a list of related specialists. Each specialist has knowledge and experience that are relevant to the farmer's problem.

④The Mailing List System shares more farmer-oriented ACs within a farmer-participating mailing list group.

⑤Integration of Model Base and ACB was achieved. A system presents the retrieval results of Model Base and ACB at the same time.

⑥The next steps are to develop a reproductive mechanism and to evaluate the ACB in a specific domain.

## Summary

Because farming is based on a complex system where many factors interact each other, it is rather difficult to construct a crop model that takes all the factors in account in a "scientific" way. This is why farmers have been using their own experiences for their problem-solving so far. The use of heuristics brings the farmer an effective problem solving mechanism. This approach is known as a descriptive approach in the decision making research. We took the descriptive approach and have developed Agriculture Case Base, a web-based system to share and reuse farming knowledge and experience. An agricultural case is a textual document describing farming knowledge and experience. Agriculture Case Base stores and retrieves agricultural cases. The system consists of three components: the Case Retrieval System, the Specialist Retrieval System and the Mailing List System. Characteristics of agricultural cases and Agriculture Case Base are presented in detail in this paper. An integration method of a model-based approach that is a normative decision-making, and Agriculture Case Base is also proposed.



## Acknowledgement

The author would like to thank Dr. Seishi Ninomiya who gave good supports and helpful discussions. The author also expresses great thanks to Dr. Takuji Kiura who helped network programming, which accelerated the research effectively. Dr. Matthew Laurenson read the manuscript and gave valuable comments.


## References


1. Chiriatti, K.C., and R.E. Plant (1996) NPK:A Prototype Case-Based Planning System for Crop Fertilization Decision Support. AI Applications 10 (2), 33-42
2. Kolodner, J. (1993) Case-Based Reasonig, San Mateo, Morgan Kaufmann Publishers, 3-40
3. Hastings J.D., L.K. Branting, and J.L. Lockwood (1996) A multiple-paradigm system for rangeland pest management. Comput. and Electron. Agric. 16, 47-67
4. Lenz, M. and H.D. Burkhard (1997) CBR for Document Retrieval: The FALLQ Project. Case-Based Reasoning Research and Development, Berlin, Springer, 84-93
5. Burke, R.D., K.J. Hammond, V.A. Kulyukin, S.L. Lytinen, S. Tomuro, and S. Schoenberg (1997) Question Answering from Frequently Asked Question Files: Experiences with the FAQ Finder System. Technical Report (TR-97-05), Illinois, University of Chicago, Chicago
6. Daniels, J.J., E.L. Rissland (1997) What You Saw Is What You Want: Using Cases to Seed Information Retrieval. Case-Based Reasoning Research and Development, Berlin, Springer, 325-336
7. Ministry of Agriculture Forestry and Fisheries (2001) Local Information (in Japanese). http://www.toukei.maff.go.jp/genti/
8. , Forestry and Fisheries Research Council. (2001) Research Information (in Japanese). http://www.affrc.go.jp/seika/
9. Deerwester, S., S.T. Dumais, G.W. Furnas, T.K. Landauer, and R.A. Harshman (1990) Indexing by Latent Semantic Analysis. J. Am. Soc. Info. Sci. 41(6), 391-407
10. Otuka, A. and S. Ninomiya (1998) Conceptual Retrieval of Agricultural Cases with Latent Semantic Indexing Approach. Proc. Int. Conf.

Engineering of Decision Support Systems in Bio-Industries, Montpellier, France, February 23-27, 37-38

11. Matsumoto, Y. (2001) ChaSen. http://chasen.aist-nara.ac.jp
12. Otuka, A. and C. Kitamura (1999) Development of Agriculture Term Dictionary for Agriculture Case Base (in Japanese). Proc. Fifth Annual Meeting of The Association for Natural Language Processing, 375-376
13. National Agriculture Research Center (1999) National Agriculture Research Center Annual Report 1998 (in Japanese). National Agriculture Research Center, Tsukuba, Japan
14. Innami, I. (1997) Excellent Decision Making (in Japanese). Chuokoronsha, Tokyo, Japan
15. Hirafuji, M., K. Tanaka, T. Kiura, and A. Otuka (2000) Modelbase System: A Distributed Model Database on The Internet. Proc. Int. Workshop on Asia Pacific Advanced Network and its Applications, Feb. 15-17, Tsukuba, Japan, 57-61

# 農業事例ベース

## ― 経験と知識を用いた意思決定 ―

大塚　彰[*1]・北村實彬[*2]

## 摘　要

営農は多くの要因がお互いに関連し合う複雑なシステムに基づいているので，それら全ての要因を考慮した科学的な作物モデルを作ることは大変困難である．このため，農業者は営農の問題解決のために彼ら自身の経験を用いてきたのである．経験則の利用は農業者に効率的な問題解決手段を提供する．この手法は意思決定の研究分野では記述的方法として知られている．我々はこの記述的アプローチをとり，農業事例ベースを開発した．農業事例ベースとは営農に関する知識や経験を共有し再利用するWEB上のシステムである．農業事例とは営農の知識や経験を記述したテキストの文書である．事例ベースはこの事例を蓄積し検索する．システムは3つの部分からなる．それは事例検索システム，専門家検索システムとメーリングリストシステムである．本論では農業事例と事例ベースの特徴を詳細に紹介する．また事例ベースと規範的意思決定手法であるモデルベースとの統合化手法の提案を行う．

平成13年12月6日受付　平成14年2月1日受理
*1 農業情報研究部営農情報システム研究室
*2 現（独）農業生物資源研究所