**The AGRIS Centre at Kasetsart University:**

**Implementing the new vision**

# THE AGRIS CENTRE AT KASETSART UNIVERSITY:

## Implementing the new vision

THAI NATIONAL AGRIS CENTRE
OFFICE OF THE UNIVERSITY LIBRARY
KASETSART UNIVERSITY
THAILAND

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
REGIONAL OFFICE FOR ASIA AND THE PACIFIC

Bangkok 2005

# Foreword

The international information system for the agricultural sciences and technology (AGRIS), was established more than thirty years ago as an international initiative of the Food and Agriculture Organization of the United Nations (FAO) and its partners to build a common information system for science and technology in agriculture and related subjects, based on a collaborative network of institutions. Recognizing the importance of agricultural knowledge and information, Thailand was quick to become an active partner in the AGRIS network, with the Thai National AGRIS Centre celebrating their twenty-fifth anniversary this year.

During the 21$^{st}$ century, agriculture will increasingly become a science, knowledge and information based sector. Against this backdrop, FAO convened the first Consultation on Agricultural Information Management (COAIM) in June 2000, attended by 161 representatives from 91 countries. COAIM 2000 discussed ways of improving the capacities of decision-makers, professionals and the public at large to access and use information essential for achieving sustainable agricultural development and food security.

The consultation recognized the continuing importance of AGRIS as a key enabler and catalyst to establish a new model of agricultural information management in the 21$^{st}$ century. The "new AGRIS strategy" would follow and capitalize on the growth of technology and decentralized information systems. AGRIS Resource Centres would develop and adopt an agreed upon set of information management standards to allow the effective exchange and retrieval of information and to accrue considerable synergistic benefits and cost efficiencies. Standards would pay special attention to categorizations schemes, thesauri and standard exchange formats. The Resource Centres would document and organize both relevant non-conventional and conventional scientific and technical literature, and create digital repositories accessible via the Internet where possible and reasonable for improved dissemination.

One of the first institutions to take up this new strategy was the Thai National AGRIS Centre. Its vision, development and experiences are described in this valuable report that may serve as both an inspiration and example to others. The Thai National AGRIS Centre has excelled beyond expectation and serves as an outstanding example of commitment to improving the value of agricultural information and improving rural development.

Now more than ever it is important that we focus on the development of tools and contents needed to enable regional and global sharing of agricultural information and knowledge. To succeed in this endeavour, it is essential that we work as partners and continue our collaboration in the service of rural communities and farmers.

FAO congratulates the Thai National AGRIS Centre and Kasetsart University for their contributions to agricultural knowledge and information over the past twenty-five years. The Thai National AGRIS Centre is a cornerstone of the global agricultural information network. We hope they will maintain this leading position as an agricultural information resource, developer of information management tools, and capacity building resource for others in this region over the next twenty-five years.

He Changchui
Assistant Director-General and
FAO Regional Representative for Asia and the Pacific

# Contents

# List of Tables

# List of Figures

# Acronyms and definitions

| | |
|---|---|
| AIBA | Agricultural Information Bank of Asia |
| AGLINET | Agricultural Libraries Network |
| AGRIASIA | Current bibliography of Southeast Asian agricultural literature |
| AGRINDEX | monthly printed bibliography of AGRIS, stopped in 1995 |
| AGRIS | International information system for the agricultural sciences and technology |
| AGRIS AP | AGRIS application profile |
| AGROVOC | a specific controlled agricultural vocabulary |
| APU | AGRIS Processing Unit |
| BT | broader term (AGROVOC) |
| CDS/ISIS | Computerized Documentation System/Integrated Set of Information Systems |
| CIAT | International Center for Tropical Agriculture |
| FAO | Food and Agriculture Organization of the United Nations |
| FAOSTAT | FAO online multilingual statistical databases; http://apps.fao.org/ |
| FORSPA | Forestry Research Support Programme for Asia and the Pacific |
| FTP | file transfer protocol |
| HTML | hypertext markup language |
| ICT | information and communication technology |
| IDRC | International Development Research Centre, Canada |
| IS-A | relation name "Is a" |
| ISO 2709 | ISSN standard used in the library world for bibliographic records |
| KU | Kasetsart University |
| NAiST | Natural Language Processing and Intelligent Information System Technology |
| NLP | natural language processing |
| NT | narrower term (AGROVOC) |
| OCR | optical character recognition |
| SDI | selective dissemination of information |
| SEARCA | Southeast Asian Ministers of Education Organization Regional Center for Graduate Study and Research in Agriculture |
| STREDEO | Multimedia Multilanguage Document Storage, Retrieval and Dissemination for E-Organization |
| XML | extensible markup language |

# 1. Introduction

To facilitate the exchange of information and bring together the world's literature on all aspects of agriculture, the Food and Agriculture Organization of the United Nations (FAO) in 1974 created the international information system for the agricultural sciences and technology (AGRIS). It is a cooperative international information system in which participating countries input references to the literature produced within their boundaries and, in return, draw on the information provided by other participants. More than 200 national, international and intergovernmental centres participate.

An assessment of AGRIS in 2000 noted that the network had been partially successful in achieving its goals but limitations were identified in four areas: 1) difficult access to the original documents, 2) incomplete coverage, 3) independent systems and 4) structural and institutional constraints.

To address these limitations, delegates and participants at the first Consultation on Agricultural Information Management (COAIM) in June 2000 discussed the development of a new strategy for the AGRIS network and its participating centres. A paper was presented to COAIM 2002 on what is now known as the new AGRIS strategy, or "new vision for AGRIS".[1]

That vision looked to turn the information system into a "resource centre". The Thai National AGRIS Centre took the challenge, becoming the first in the network to step up to the challenge. In becoming a resource centre, the Thai AGRIS Centre aims to build a national network, deploying WebAGRIS and adapting tools and standards to suit national needs while contributing to the global information resource. This report describes the work and experience of the Thai National AGRIS Centre in adopting the new AGRIS strategy, with the intention that it may serve as a useful reference to other organizations seeking to do the same.

---

[1] For additional information see http://www.fao.org/agris/

# 2. The role and history of the Thai AGRIS Centre

For more than 24 years, Thailand's affiliate of the AGRIS has been collecting and recording agriculture data and experiences from around the country to feed into the international network. Its database serves as a popular resource for agricultural education, research and development and as a mechanism for exchanging information with other AGRIS centres worldwide. This collaboration allows the international community to benefit from Thai agricultural information while giving the Thai agricultural community access to valuable insights from far and wide.

## Background

Until 1995, all FAO member countries sent information to the AGRIS processing unit (APU) at the International Atomic Energy Agency in Vienna where it was processed and printed for distribution. With the advances in communications technology in the 1990s, the APU began distributing all data online. Other services to member countries include the selective dissemination of information (SDI) publication, the AGRIS CD-ROM and AGRINDEX. Since 1995, the AGRIS secretariat at FAO headquarters in Rome has coordinated all AGRIS activities.

To initiate the Thai AGRIS Centre in 1974, FAO officials approached Bangkok's Kasetsart University, then recognized as the best national agricultural university and which already had a strong library. However, the library lacked staff with professional skills in documentation that was required by AGRIS procedures. While staff members improved their capabilities, FAO worked with the Thai National Documentation Centre of the Scientific and Technological Institute, which had trained specialists, to begin developing the Thai AGRIS Centre and to operate it until the Kasetsart library was ready to take over. For five years, the two facilities closely collaborated until Kasetsart took operating control.

The Thai AGRIS Centre officially opened at Kasetsart in 1980 with a five-year operating budget financed by IDRC. Since 1985, the Thai Government along with library income have supported the centre's operations and the salaries of information scientists.

## Functions of the AGRIS Centre

The main activities of the AGRIS Centre:

1. Collect and manage agricultural and other relevant information that has been produced or published within country.
2. Provide information services on both published and unpublished literature produced in country.
3. Promote agricultural information utilization.
4. Carry out research on agricultural information development.

Specific activities of the Thai AGRIS Centre focus on:

1. Collecting and producing agricultural information. The Thai Agricultural Bibliography (1980-1991) is a major output publication that is distributed nationwide.

2.  Building the Thai agricultural database (1980-present). Data recording is done in English and sent to the central AGRIS database while a Thai version is produced and managed for Thai users. This is available at http://thaiagris.lib.ku.ac.th/.

3.  Coordinating and partnering with AGRIS/FAO. Currently, the centre acts as a focal point for collecting and developing agricultural information in Thailand. Figure 1 shows Thailand's contribution to the international AGRIS network since 1975.

**Figure 1: Thai AGRIS data entry, 1975-2004**
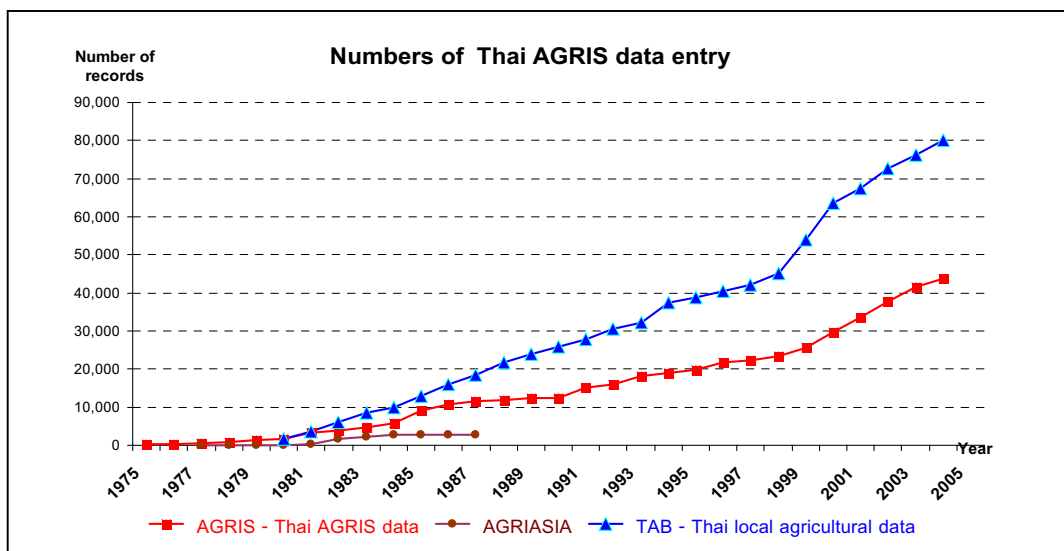


4.  Providing agricultural information from abroad. The centre allows Thai users access to international agricultural information in its central AGRIS database. This is available in the form of the AGRINDEX, AGRIS CD-ROM, AGRIS Online and SDI. Thai users, both general and academic, gain substantial benefit from the fast and convenient accessibility, as Figure 2 suggests.

**Figure 2: Intensity of AGRIS CD-ROM utilization at the AGRIS Centre, 1998-2001**
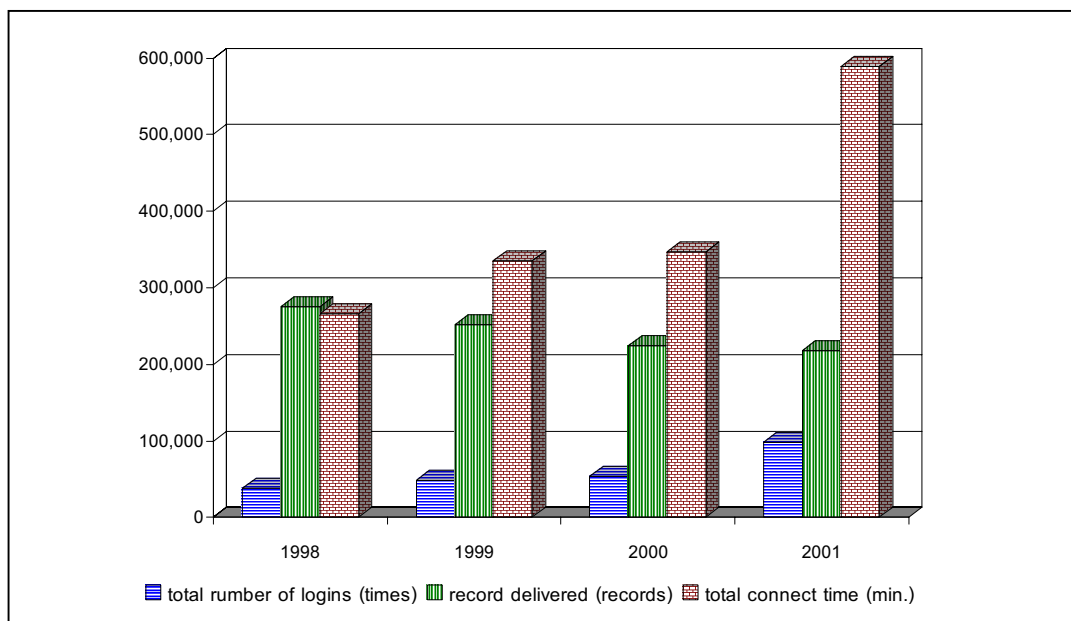
5. Expanding the agricultural information network within the country. However, despite much effort there have not been any satisfactory results due to limited resources, particularly in terms of personnel and finance in most of the cooperating centres. Recent developments are encouraging, such as network discussion forums and policies to enable sharing of resources.

6. Supporting other agricultural information systems in Thailand. The centre provides assistance in the form of consultations and training programmes for users.

7. Conducting training workshops. The centre organizes training sessions for librarians and information scientists from various institutes in Thailand.

## History of Thai AGRIS database development

In 1980, the centre used methodologies recommended by the AGRIS secretariat. Initially, information gathered in Thailand was typed onto an input sheet and delivered by mail in print form to the Agricultural Information Bank of Asia (AIBA) in the Philippines, which was working as a regional centre. That information then was entered into a database and sent to the AGRIS processing unit in Vienna.

The Thai centre acquired its first computer in 1981 with the support from the IDRC of Canada. In accordance with AGRIS requirements, data entry into the Thai AGRIS database was done using Dbase III Plus, in English. As computer technology progressed, the AGRIS system changed. In 1986 data were sent from Thailand to AIBA by diskette and then directly to the AGRIS processing unit, beginning in 1992.

In mid-1994 the Kasetsart University library developed a network system with links to the Internet. However, it was not until 1998 that the AGRIS processing unit started accepting data and information via email and FTP from participating AGRIS centres.

From the beginning it was apparent that most Thai users could not benefit from the AGRIS data because of the language barrier. Thus the centre began processing the information in Thai as well as English. The Kasetsart library collaborated with the university's computer centre to develop a Thai agriculture bibliographic database system. Data recording in Thai language started in 1982, though it included data from the previous two years (1980-1981).

The Thai database was processed every two months and published in the *Thai agricultural bibliography*. This document was distributed to members and agricultural agency libraries around the country up to 1991 when the high cost of printing and delivery forced the publication to cease operation. It has since been replaced with an online database and CD-ROM.

In 1987 the Thai AGRIS database application at the Kasetsart library changed from Dbase III Plus to Micro CDS/ISIS version 1.0, which the AGRIS secretariat distributed. The secretariat also supported training at SEARCA in the Philippines. Currently the centre is using the WebAGRIS version, which uses Internet-based technology. The Thai Agriculture Bibliographic database, recorded at the Kasetsart computer centre since 1982, was moved to the main library and changed to Micro CDS/ISIS. The two databases are thus maintained separately, with English and Thai on CDS/ISIS.

By 1996 it became apparent that managing the two databases in English and in Thai was impractical and costly, and users were not able to utilize them simultaneously.

A study of the AGRIS database revealed that the structure did not support two languages. To resolve the problem, the database structure was modified by duplicating the Thai field into it for internal use only, though still adhering to the AGRIS standard. The fields and format for data recording were adjusted to support information in Thai. Once the structure could support both languages, all Thai data were transferred into what became known as the Thai Agricultural Database. Recording data in two languages in one file began in 1997.

## Development of information services at AGRIS

As previously mentioned, AGRIS makes available various types of agriculture information that is both Thai-based and from other countries through the AGRINDEX, AGRIASIA, the *Thai Agricultural Bibliography* and other resources. In addition, AGLINET, the network of data from all AGRIS centres, provides material to online users through the SDI and full-text information services.

The first retrieval of information in electronic format from the AGRIS processing unit in Vienna began in 1983. This was done online via satellite access to the international database and remote computing service. However, the service then was very costly so the APU provided the offline SDI service. Students and agricultural researchers could request information through the Thai AGRIS Centre and the search results were delivered monthly.

By 1986 the Thai AGRIS Centre was able to link directly to the AGRIS online service in Vienna using modem and satellite connection. This made it possible for users to access the central database, and demand for the service was great. AGRIS provided the service for free, although Thai users had to pay a communication service fee charged by the Communication Authority of Thailand.

Since that time, the online data retrieval service has continually evolved. By 1991 users were able to search for information on the AGRIS CD-ROMs that were made available on a stand-alone computer at the centre. Six years later, the centre provided services within the library network, which had expanded to include the university network (KU CD-NET). The system allowed users to search, download and print data. In 1999 the AGRIS central database became available via the Internet and data now can be accessed through the Internet without any charge at http://www.fao.org/agris/. This information is more up to date than on the CD-ROMs.

The Thai language agricultural information service started in 1992 and users could access it on a stand-alone computer. However, the search program was too complicated and only a few computers were available. The library improved the service in 1995 by installing a local area network. At the same time the Thai AGRIS Centre modified the CIAT search interface to be more user-friendly. This program provided a search and retrieval service and displayed results based on the CDS/ISIS system in Thai. Users promptly gave encouraging feedback and as a result, in 2000, the centre developed a homepage and installed the Web ISIS program to provide the database service through the Internet (at http://thaiagris.lib.ku.ac.th).

An ensuing survey among the users indicated the services were quite valuable because the AGRIS information was accessible for free.

## International cooperation and knowledge management projects

The success of the Thai AGRIS as an agricultural resource and knowledge centre is attributed primarily to three factors: first, the perseverance and devotion of the centre's staff; second, the assistance from AGRIS/FAO; and third, Kasetsart management in providing help and facilities.

In addition, the collaborative inputs from various sources, as the following describes, have added value to the service:

- The Kasetsart library provided the AGRIS Centre with a scanner in 2001 so that printed documents could be input in digital format instead of on microfiche through the agricultural information project. Within that project there is a plan to convert microfiche documents dating from 1980 to 2000 to electronic format. The Kasetsart library also is supporting another staff member for the document scanning.

- AGRIS began in 2001 a five-year project to develop a Thai agriculture thesaurus with funding under the programme of *Development of agricultural information system for knowledge acquisition and dissemination*. Many agencies are involved in this collaboration, including the Kasetsart Research and Development Institute and the NAiST laboratory (a specialty unit of the Natural Language Processing and Intelligent Information System Technology program in the Kasetsart faculty of engineering), along with technical assistance from the AGRIS secretariat.

The thesaurus provides a standard for indexing Thai agriculture information and collects Thai agriculture vocabulary. This process links the Thai agricultural ontology[2] to the development of automatic machine translation.

The implementation of the thesaurus project (see Chapter 4 for details) signified a new direction toward research and development work for the Thai AGRIS and a leap for the integration of information centre activities with the two research institutions, which ultimately will strengthen the centre's capabilities.

- In celebration of its 20th anniversary, the Thai AGRIS Centre organized a conference on information technology for agriculture (6-7 July 2000) with key representatives of agricultural agencies from around the world attending. Discussions focused on cutting-edge technology and agricultural information systems, accompanied with numerous brain-storming sessions. That conference resulted in the development of the Thai Agricultural Information Network.

Technical assistance from the AGRIS secretariat also enabled the Thai AGRIS Centre to develop in 2003 the Thai WebAGRIS multilingual program for data recording and retrieval (see Chapter 8 for details).

---

[2] An *ontology* is a model of organized knowledge in a given domain (e.g. fisheries) consisting of components called concepts, attributes, relations and instances. For more information see http://www.fao.org/aims.

# 3. Networking and national cooperation

It is impossible for any single organization to comprehensively collect all the available agriculture information that exists in a country. It takes a network and cooperation of many organizations and agencies. An important factor that has sustained the AGRIS network is the mutual benefits all member countries enjoy from the central database. By contributing, each has access to a wealth of knowledge and experiences.

Cooperation is possible when organizations are aware of the benefits and added value they receive from working together. Key factors for the success of a network entail a shared vision, clearly defined roles and responsibilities for each contributor, mutual trust and good relationships among members of the network.

The Thai AGRIS Centre began with a mandate to create a network of agricultural agencies within Thailand and to organize training on bibliographic data collection and data input. However, the endeavour has been only partially successful due to the lack of human and financial resources. Improvements have taken place recently due to the cost of information and communications technology (ICT) falling substantially, and because more government agencies are aware of the need to capture and manage information and thus contributing support. As well, the Thai AGRIS Centre has become better equipped, enhancing the support it can give to its network members.

Advancements in information technology facilitate easier networking and allow for greater and more viable collaborations. It is no longer necessary to process data centrally; each member of a network can develop its own database rooted upon a common standard and then data can be linked, allowing for access within the distributed systems.

With the significant developments as well as the considerable reductions in ICT costs during the past decade, more AGRIS centres now can acquire equipment and process their own data in any chosen application (not just the recommended AGRIS tools). In Thailand, one of the advantages is that the processing of AGRIS data is no longer a burden to the national centre but is a shared responsibility of all the Thai network members. The issue now is for the Thai AGRIS network to increase its effort in researching and providing methodologies that support access of information in a distributed environment.

Toward that goal, FAO recently released a new AGRIS methodology that facilitates the exchange of information from distributed systems. The AGRIS application profile (AGRIS AP) is a format that uses well-accepted metadata standards and syntax to permit both inter-domain and intra-domain operability. It is a standard created specifically to enhance description, exchange and subsequent retrieval of agricultural document-like information objects. As a format to produce high-quality metadata, it allows for a platform-independent exchange of information about different types of agriculture resources.

# 4. Localization of knowledge – The Thai AGROVOC project
## http://pikul.lib.ku.ac.th/

AGROVOC is a comprehensive multilingual agriculture thesaurus that was developed with the cooperation of FAO member countries. It is used for indexing data in agricultural information systems and is continually being improved and updated. The first version of AGROVOC was produced in 1982 and distributed to all AGRIS centres.

Vocabulary updating is done by FAO with collaboration from national AGRIS centres. Staff at the centres propose new terms for the database to FAO subject specialists for consideration. The terms selected by the experts are added into AGROVOC. In the past, an AGROVOC supplement was then published and provided to the centres. Now the updated AGROVOC is available online. The proposing of new terms and corrections also can be done through the FAO/AGROVOC web site.

Initially AGROVOC was available in English, French and Spanish but has been expanded to four more languages: Arabic, Portuguese, Czech and Chinese. The Thai AGRIS centre is developing the Thai AGROVOC by using the English AGROVOC thesaurus as a prototype.

The following describes the motivations, work plan and problems in developing the Thai AGROVOC and provides suggestions for its future.

## Limitations of translating AGROVOC to the Thai Language

Data and information produced locally typically are displayed in native language. Thus to be useful they also have to be recorded and indexed in the local language. However, each language is different in alphabet characteristic and structure. Some languages are easy to process by computer while others are very impractical or even unlikely.

For the latter, basing an index on uncontrolled vocabulary and computer processing is extremely difficult. Thai language is such a case. This is because the written Thai language has no space between words; so it is very difficult to use a word-spacing program to separate words into individual and correct words. In addition, many Thai words have the same form but different meaning or different forms with the same meaning.

Because it is difficult to use a word-spacing program for making an uncontrolled vocabulary index and because an index that has been produced from automatic spacing is low in retrieval efficiency, having a controlled vocabulary becomes quite critical. The attributes of Thai language and its limitation for natural language index by computer processing make it inefficient for data retrieval and utilization. For these reasons, the Thai AGRIS staff opted in 2001 to develop a Thai AGROVOC. As mentioned previously, financial support from Kasetsart University's Research and Development Institute and technical support from FAO facilitates the project.

The project implementation has been divided into two phases: The first focused on developing the Thai AGROVOC (and the electronic agriculture thesaurus) and the second relates to adding words in the local language and other maintenance. The electronic agriculture thesaurus is the first Thai database that stores and displays agriculture vocabulary in bi-language – Thai and English.

**Table 1: Objectives of the Thai AGROVOC**

| Objective | Problem to solve |
|---|---|
| 1.  Create standardized terms for indexing in Thai. | • Thai words have the same form for different meanings and the same meaning in different forms; there is no standard word for making an index. |
| 2.  Provide Thai descriptors for inputting data in the AGRIS system. | • Indexing in English is not efficient and erroneous, depending upon English literacy of the person who is making the index. |
| 3.  Increase data retrieving efficiency. | • A Thai word always has various meanings and various forms.<br>• Thai automatic word separator is very difficult and very constraining for searching.<br>• Existing searching does not allow efficient free text searching in Thai. |
| 4.  Link Thai vocabulary with words in other languages in the AGROVOC. | • No local language technical terms and name entities relate with other languages. |
| 5.  Make it more convenient and possible for every level of user to search the index. | • Users who are not keen in either English or agricultural science cannot identify agricultural vocabulary for data searching. |
| 6.  Create Thai agriculture word stock in both English also. | • There has never been an electronic agricultural thesaurus in Thai. |

## Developing the Thai AGROVOC

Developing the Thai AGROVOC has been a kind of research and development process for the centre, with considerable exploration into integrating ICT and knowledge management to enhance the efficiency of information service. The project has involved many resource persons and information sources, as the following outlines:

• 37 experts in 31 fields of study with 12 lecturers and researchers working as translators;

• FAO AGROVOC thesaurus;

• Thai dictionary;

• Special fields dictionary related to agriculture (see the appendix 2 for the listing);

• Text books and document on agriculture; and

• Vocabulary and index in articles and research paper for the past 20 years.

The five-year project evolved in two stages:

• *Stage 1 – Building vocabulary collection (2001-2003)*
    – System analysis
    – Collecting vocabulary from technical dictionary and research papers
    – Translating terms from English to Thai
    – Building Thai agriculture vocabulary database
    – Installing system and providing Thai AGROVOC service through the Internet
    – Promoting public awareness and feedback

- *Stage 2 – Adding new vocabulary and maintenance (2004-2005)*
  - Collecting comments for each term
  - Editing vocabulary
  - Adding local vocabulary into the system
  - Filling up new vocabulary
  - Updating vocabulary collection

**Figure 3: The Thai agricultural thesaurus development process**
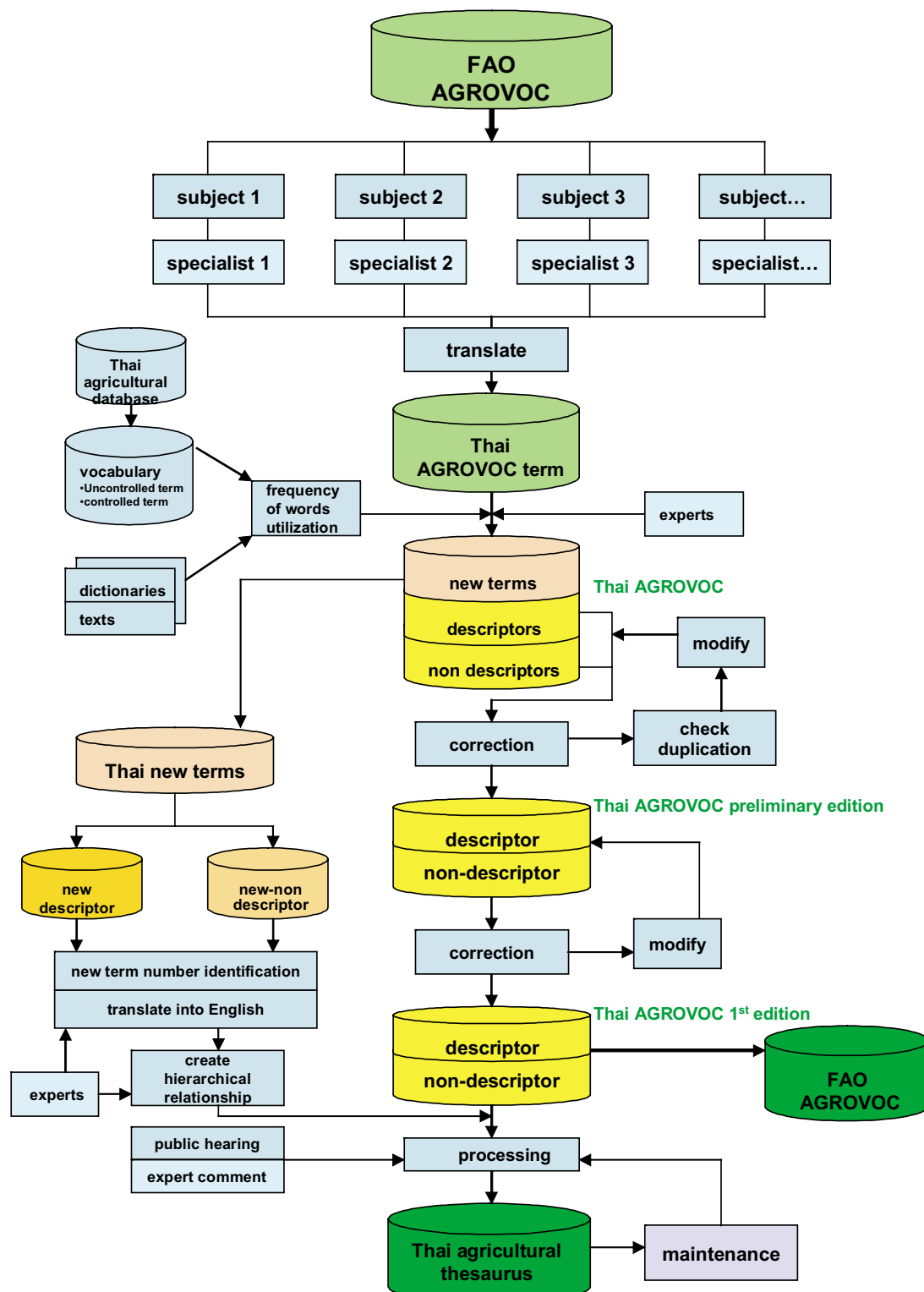
## Detailed achievements of the five-year project

### Stage 1

1.1. Built vocabulary collection using vocabulary database from AGROVOC (Agrovoc.mdb) version 9 Nov 2001 as a prototype. The vocabulary in the database is displayed in three languages (we used English for the prototype, totalling 28,577 terms).

Term number and term type in AGROVOC database (version 9 November 2001)

| English descriptor | Non-descriptors | Deleted terms | Scope note |
|---|---|---|---|
| 16,607 | 10,706 | 164 | 1,264 |

1.2. Categorized vocabulary, descriptors and non-descriptors (except scope note), totalling 27,313 terms into 47 groups in 31 subjects, based on AGRIS/CARIS subject categorization schemes and the field of expert's specialization.

1.3. Created terms list for specialists to translate from English into Thai. Verified them with a technical dictionary.

1.4. Sent the translated vocabularies to respective experts for confirmation, editing and modifying and any translation that the subject specialists could not do. The experts could recommend other descriptors or synonyms also.

1.5. Collected the translated and confirmed vocabulary from the experts.

1.6. Added more vocabulary by selecting terms from titles and index of research papers and articles in the Thai Agricultural Database (dating back 20 years and extracting approximately 92,605 words). Extracting an individual word was done manually because there was no well organized word-extraction tool. This is because the Thai language has no space between words so it is next to impossible for a word-extraction program to extract words without errors. Take for example: "การศึกษาวิธีการปลูกมะม่วงน้ำดอกไม้". It should be extracted to การศึกษา วิธี การปลูก มะม่วงน้ำดอกไม้ but can easily come out as การศึกษา วิธี การปลูก มะม่วง น้ำ ดอกไม้, which does not have the same meaning.

1.7. Vocabulary from the previous step was regarded as a natural-language term. They were ranked and frequency checked. Then they were split into two groups: one was synonymous with AGROVOC and the others were defined as a new term.

1.8. Selected the suitable term as a descriptor. The rest were kept as non-descriptors (synonym). The principle criterion was to use words that are defined and recognized by the Royal Institute. Vocabulary from a technical dictionary and text was a secondary criterion. The third was frequency of use in the literature.

1.9. Entered the Thai vocabulary (descriptor) into the Thai AGROVOC database using the same identification as terms in English by adding a new field for Thai terms. As synonym was recorded separately in another table to be used later as a non-descriptor.

1.10. Checked word redundancy and relationship and corrected errors.

1.11. Submitted the vocabulary to experts for re-verification and re-editing.

1.12. Modified them according to the experts' recommendation and corrected errors.

1.13. Processed a preliminary Thai AGROVOC.

1.14. Installed the system to provide service and get feed back via the Internet.

1.15. Evaluated the system, verified terms and their relationship.

1.16. Edited data and processed Thai AGROVOC first edition.

1.17. Installed a public feedback system with Internet access.

1.18. Launched the service, promoted it through public relations campaigns and disseminated it to interested groups of people for comments.

1.19. Submitted the Thai AGROVOC to FAO.

**Stage 2**

2.1. Collected comments and opinions from the public; summarize and edit data.

2.2. Assembled proposed terms obtained from the public; as in step 1.7, accepted vocabulary were taken either as new words or as synonyms.

2.3. Confirmed those new words with experts, created a relationship with existing words and then translated new words from Thai to English.

2.4. Added terms with relation into Thai AGROVOC database by creating new term identification.

2.5. Recorded vocabularies, checked redundant terms and corrected errors.

2.6. Submitted the Thai vocabulary to the Royal Institute for them to promulgate it as an approved Thai agriculture vocabulary.

Existing words are regularly modified and new words are added to keep AGROVOC up to date. Adding new words is possible when there is a suggestion from users and by developing thesaurus maintenance tools that automatically add words and create a relationship.

**Figure 4: Process of extracting Thai agriculture terms**
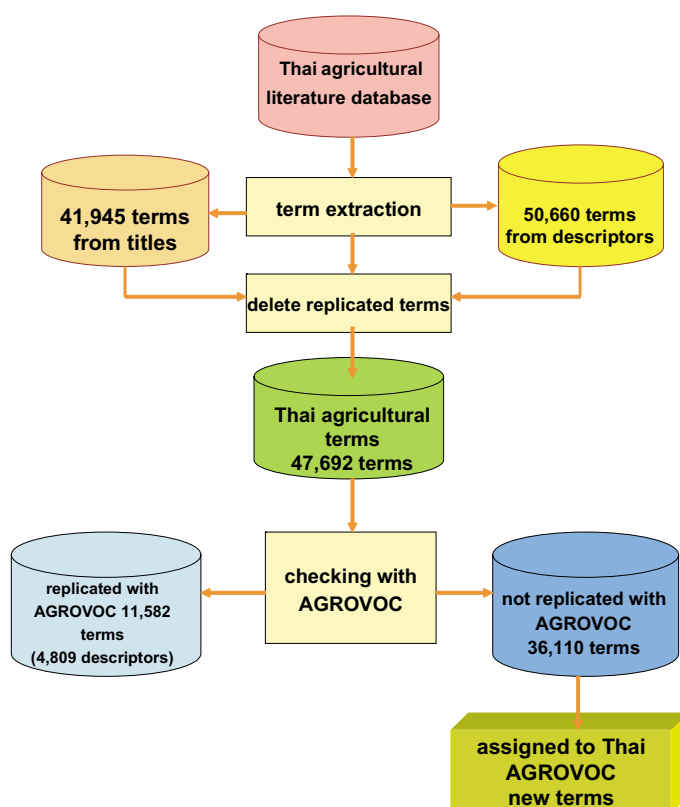
## Critical factors for success of the Thai AGROVOC project

Several elements worked together toward the success of the Thai vocabulary project:

1. Body of knowledge from a variety of sources;

2. Contributions from experts and subject specialists;

3. Intelligence tools from system developers;

4. Collaboration from agencies in the network and their recommendations;

5. Feedbacks from users;

6. Managerial capability of system administrator; and

7. Sufficient budget.


## Constraints of the Thai AGROVOC development project

Several difficulties were encountered in the development of the project:

1. **Incompatibility between local information and vocabulary.** Many descriptors in the AGROVOC are not available in the local scope of knowledge or are incomprehensible to local people. On the other hand, there are also countless local terms, especially the names of plants, animals or other local beings, that do not appear in AGROVOC.

2. **Difficulty of defining new terms in Thai.** Many terms in the AGROVOC have never appeared in Thai so there is a necessity to define them. This endeavour is really burdensome. Defining new terms needs expertise and mutual collaboration from experts in that very specific field. Finally, each term has to be officially accepted and registered as new.

3. **Inconsistent language structure.** In Thai, there are not singular or plural. Therefore the meaning of word with "s" and word without "s" cannot be displayed differently as English terms in AGROVOC. For example, the terms "seed" and "seeds" have different meanings in AGROVOC.

4. **Incompatibility of meaning and synonymous words.** Some words possess more than one meaning in English but mean only one thing in Thai. For example, "corn" and "maize" are separate in English but in Thai they mean the same thing and have only one Thai term. On the other hand, there are Thai terms with different meanings but only one term in English, such as "ควาย" and "กระบือ", which in English refer to "buffaloes".

5. **Very time consuming.** The endeavour has to utilize countless experts in many fields in addition to a lot of personnel for word extraction and data processing.

6. **Maintenance.** Adding words and changing relationships are extremely difficult due to the lack of an efficient tool.

7. **Shortfall.** Some original terms in AGROVOC, as prototype, are not up to date; also, their relationships are incomplete, unclear and inadequate.

## Recommendations

1.  FAO should update the AGROVOC by restructuring, improving and editing the vocabulary.

2.  FAO should design a clear procedure for identifying identifiers and the data structure to facilitate the entering of additional local vocabulary into the system.

3.  It would be very helpful if FAO developed efficient tools for AGROVOC maintenance. This would allow member countries to regularly maintain the vocabulary system on their own.

4.  As such, the development of a national AGROVOC to completion is very critical for local users to effectively access the widest global knowledge base. Efficient tools for maintenance and automatically updating thesaurus terms are necessary and waiting to be developed.

## Thai AGROVOC public hearing

The Thai AGRIS Centre has developed an Internet public feedback system that allows users, specifically experts, academics and agricultural scientists, to voice their opinions or suggestions, confirm the exiting terms and propose new terms (http://pikul.lib.ku.ac.th/). They can verify vocabulary or even suggest new words. This process is expected to continuously improve the quality of the Thai AGROVOC, especially in terms of comprehensiveness, accuracy and relevance to the Thai agricultural body of knowledge.

However, developing the Thai AGROVOC is still an uncompleted attempt. Further integration of technology with expertise in agricultural sciences and information science is needed to enhance the efficiency of AGROVOC for knowledge discovery.

# 5. Localization of tools – Thai WebAGRIS

The AGRIS database is accessible in English, Spanish and French – which makes it unavailable for most Thai users. As well, only the title and descriptors fields in the AGRIS database structure allow data in those three languages. This is not sufficient for data recording and use in Thai.

It is on this basis that the Thai AGRIS Centre decided to develop its bilingual database that could accommodate both the Thai and English languages. In addition to the Thai AGROVOC, the centre also has developed a Thai version of the WebAGRIS program for operating its databases.

## Developing application program for managing bilingual database

In order to create its bilingual database, there was a need to develop database management system that can record, retrieve, index and display results in two languages simultaneously. Its structure had to be consistent with the AGRIS standard (AGRIN 3.0). Therefore fields were added for recording Thai data with the Thai alphabet in the same format as the English fields.

Obviously Thai alphabets differ from Roman alphabets. In the Thai language there are 44 letters, 24 vowels and 4 sound symbols. A Thai sentence is a continuity of words without spaces between them. For example: Food and Agriculture Organization of the United Nations in Thai would appear as องค์การอาหารและเกษตรแห่งสหประชาชาติ.

In order for the database management program to recognize Thai letters and correctly produce an inverted file, the Thai alphabet has to be included in the alphabet code table. This was made possible by the MICRO CDS/ISIS file named ISISAC.TAB and ISISUC.TAB.

## Applying WebAGRIS for managing bilingual database

The Thai AGRIS Centre has been experimenting with WebAGRIS since November 2001 with the objective to use it with Micro CDS/ISIS version 3.08 for managing databases. WebAGRIS is a Web-based program of CDS/ISIS developed by the Institute for Computer and Information Engineering in Poland in collaboration with FAO. WebAGRIS is multilingual and its database structure adheres to AGRIS standards and enables export of data in XML, HTML, ISO 2709 formats. WebAGRIS permits data use and sharing across the Internet and works on a client-server basis.

**Figure 5: WebAGRIS technology**
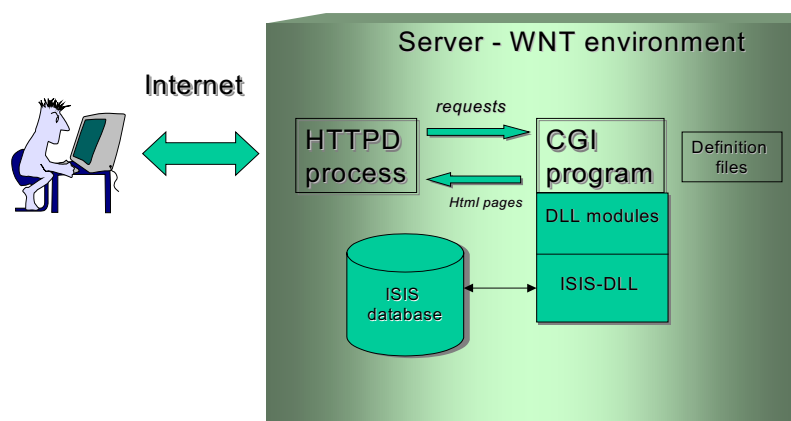


Source: *WWW/ISIS Technical reference manual.* v. 5.1.0/00-12-01. Warsaw, February 2000. http://www.icie.com.pl/

Studies conducted by the Thai AGRIS Centre highlighted the advantages of WebAGRIS for data recording and data exchange for a bilingual database. However, some minor problems noted had to do with inconsistencies in the data structure. Some fields in the WebAGRIS system identified subfields under repeatable fields, which is different from the old structure.

Applying WebAGRIS in data recording and displaying in Thai is done by copying data files to new ones for Thai application. This is possible because the system is purposefully designed to accommodate many languages. The copied data files can then be modified as necessary.

## Thai AGRIS Centre experience with WebAGRIS

WebAGRIS possesses many useful functions for recording and retrieving data that are not available on Dos Micro CDS/ISIS. These attributes make it easier to manage data and to transfer data from the older AGRIN formats using the field selected table developed by AGRIS. The following outline of functions make WebAGRIS a better application for use by the Thai AGRIS Centre:

### Data entry

1. Authority control in some data fields. For instance, in the subject categories where there are fields for author, journal and conference names. The control enables consistency and data quality.

2. A controlled multilingual vocabulary (AGROVOC). Incorporated in the data entry process, this maintains quality and enables indexing in the desired language.

3. Data input rules, which are incorporated making cataloguing easier.

4. Possibility to copy a worksheet from ready data records. This saves time required for data entry.

5. Users can define the default value in the worksheet, a useful procedure when cataloguing many entries from one document.

6. Automatic updating of entries into the respective fields from the authority.

7. In general, entering and editing data is easier and more convenient. There are features that help reduce error from data entry.

### Data searching

1. There are both simple search and advance search with the latter offering the possibility of searching specific field to refine the search.

2. Use of controlled vocabulary and index to search.

3. A search history enables a user to retrace the searching path taken.

4. Help function is in HTML format and linked to other concerned sections. It is also very user friendly.

5. Full text can be displayed by clicking a link to a URL of that e-document.

6. Multi-database searching allows centres that have multiple databases to conduct simultaneous searches.

**Dissemination**

1.  Data dissemination can be far-reaching due to the Internet.

2.  Possibility to export data in XML format, making the AGRIS data available to data providers for other value-added services.

**Database management**

1.  The system can be used in a stand-alone and client-server mode. Users can search data simultaneously through the network. Multiple data entries are possible in the same ISIS database.

2.  The security system is easy to manage. One text file allows the administrator to easily create and manage the user privileges, applying different levels of access.

3.  Data entry and editing is easier than with Dos ISIS because it can be performed online and in real time.

## Performance evaluation of WebAGRIS

In addition to lectures on metadata and agricultural information standards during the first *Workshop on agricultural information management: WebAGRIS system and Thai agricultural information networking* (25-28 August 2003) at the Kasetsart University library, there was a training for WebAGRIS users. Participants were required to provide feedback on the performance. The results were encouraging with the participants being especially happy with the bi-lingual function of the program, its ability to initiate keywords automatically in different languages and its ability to export data in XML format. The following table presents a summary of that performance evaluation of WebAGRIS in Thai version.

**Table 2: Performance evaluation of the Thai WebAGRIS**

| WebAGRIS attributes and its functionality (Thai version) | Percent of response according to different degree of satisfaction | | | |
|---|---|---|---|---|
| | Much | Moderate | Little | No comment |
| Language | | | | |
| 1. Bilingual database (Thai and English) | 100 | 0 | 0 | 0 |
| Searching | | | | |
| 2. Method | 57.7 | 30.8 | 7.7 | 3.8 |
| 3. Displaying result | 76.9 | 15.4 | 0 | 7.7 |
| 4. Help function | 65.5 | 26.9 | 3.8 | 3.8 |
| Data entry | | | | |
| 5. System authentication | 76.9 | 19.3 | 0 | 3.8 |
| 6. Format and method for data recording | 65.5 | 30.8 | 0 | 3.8 |
| 7. Format and method for data editing | 73.1 | 23.1 | 0 | 3.8 |
| Data transfer | | | | |
| 8. Into ISO 2709 format | 53.9 | 38.5 | 3.8 | 3.8 |
| 9. Into text format | 61.6 | 30.8 | 3.8 | 3.8 |
| 10. Into XML format | 61.6 | 30.8 | 3.8 | 3.8 |

## Limitations of WebAGRIS

Despite the favourable response to the Thai version of the WebAGRIS, several constraints have been noted:

1.  Authority to update and delete record is determined at the database level so anyone who has access to the database can also have access to the records that are created by others. This is still problematic because it allows that person to modify the records that are not in their authority.

2.  There are still some errors in the system because of the shortfall of the program; hence sometimes the recorded data are not properly or effectively saved.

3.  Database lock and record lock happen from time to time with unknown reason. And untying the lock can be done only on the server. This makes it impossible for the network agencies to do data recording until the system administrator unlocks the system.

4.  Creating a full inverted file has to be done on the server only while the client can do only the operation of updating inverted file, which is automatic.

5.  Value table and data files for WebAGRIS are plenty and complicated. Modifying the system for a specific purpose is somewhat tricky, though it is not so for network agencies that are working on a standard and completed system.

# 6. Access to electronic documents

The Thai AGRIS Centre receives 99 percent of its information in print and only 1 percent in electronic files. However, it has become evident that most users are interested in acquiring the full text documents of the information held in the bibliographic databases. Print materials are very difficult to locate in a decentralized system such as the AGRIS network and as such many managers are working toward offering their collection online.

Advancements in information technology have broken the barriers of geographical boundaries and disparity of information systems, enabling information services through the Internet without much of the previously experienced limitation. Access to full-text documents online is the ultimate need, beyond information discovery, of every user. Thus availability online has became a priority for every information manager.

Electronic full text documents services in the AGRIS network should be available as open-access repositories. However, this is not possible in many countries due to the copyright issue and due to the fact that many of the documents are not in electronic format or online.
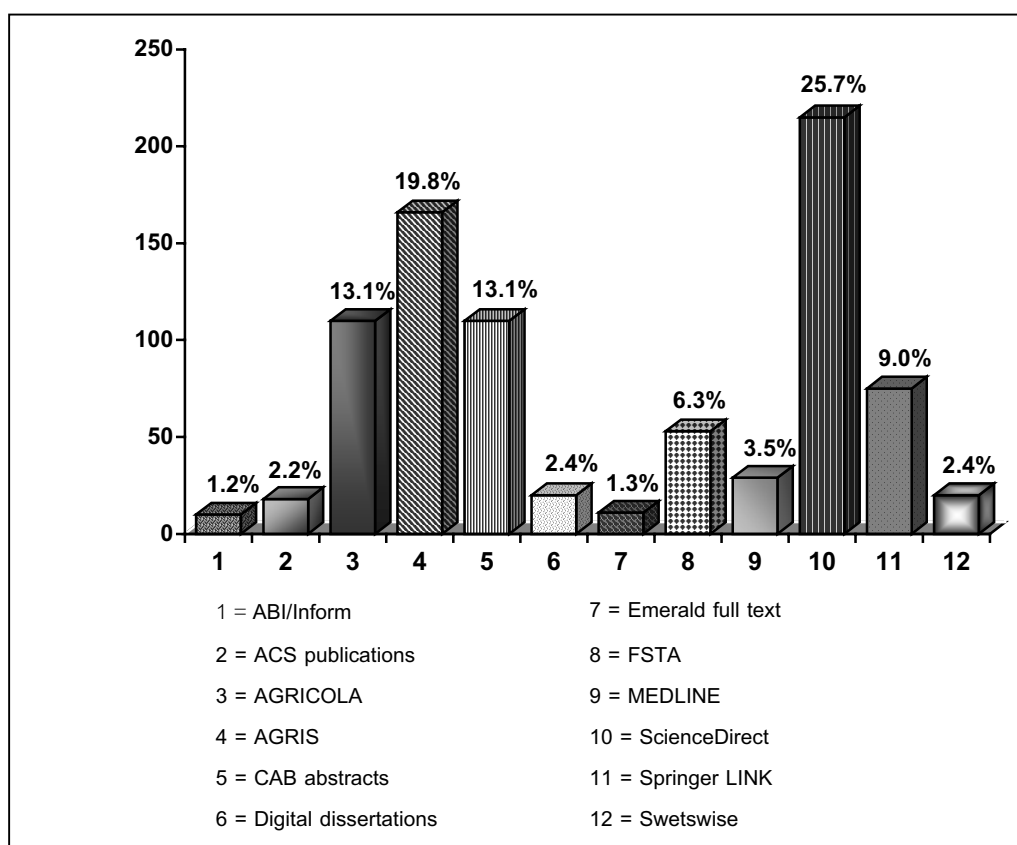
One most efficient way to collect the full text documents is by promoting self-archiving by the owners into the institutional repository or availing a digital copy of their work to the institutional library and information system outside. There is a strong movement – the Open Access Movement – that promotes this cause.

The Thai AGRIS Centre, with support from FAO and IDRC, has made full-text documents available to all bibliographic records in the database from 1980 using microfiche technology. This was the most appropriate technology at that time because of its low cost of production and the long shelf life. However, there are some limitations in the wider use of the microfiche technology because it requires specific tools for reading and a user has to physically go to the centre.

Results from a study on the use of online databases by graduate students in agriculture at Kasetsart University confirmed this point. Most students showed greater interest in the full-text databases. It was also noted that the greatest limitation of online bibliographic databases was the accessibility of the full-text documents. The study also indicated that the most popular database with full-text documents was ScienceDirect, which is an electronic journal database – though the content was not of great interest as an agricultural subject to the students.

AGRIS was the most popular database among the bibliographic databases with no full-text documents.

**Figure 6: Kasetsart agriculture students and the online databases they use**



| Index | Database | Index | Database |
|-------|----------|-------|----------|
| 1 = | ABI/Inform | 7 = | Emerald full text |
| 2 = | ACS publications | 8 = | FSTA |
| 3 = | AGRICOLA | 9 = | MEDLINE |
| 4 = | AGRIS | 10 = | ScienceDirect |
| 5 = | CAB abstracts | 11 = | Springer LINK |
| 6 = | Digital dissertations | 12 = | Swetswise |

Source: Krisna Bennukul. 2003. *Behaviour in using online databases for agricultural information searching by graduate students in agricultural science,* Kasetsart University

## Production of digital documents by scanning

Studies indicate that archivists face several challenges in archiving digital information. One of the greatest is technological obsolescence, which is the result of the evolution of technology: As newer technologies appear, older ones cease to be used. Currently new devices, processes and software are replacing the products and methods used to record, store, and retrieve digital information at breathtakingly short cycles of two to five years. Consequently, information that relies on obsolete technology becomes inaccessible.

Preserving digital information also focuses on the longevity of the physical media on which the information is stored. Even under the best storage conditions, however, digital media can be fragile and have limited shelf life.

At present, scanning technology is the best option for digitizing print documents at the Thai AGRIS Centre due to its low cost of production, more efficient capacities and cheap tools of storage. It also is widely applied even though the microfiche technology is more stable. As previously mentioned, the centre therefore has made available electronic full-text documents by scanning all print documents since 2001 and from transferring microfiche to electronic files from 2000 back to 1980.

## Table 3: Numbers of digital documents scanned by the Thai AGRIS Centre (as of July 2004)

| Type of documents | Scanned document | | |
|---|---|---|---|
| | Hard copies (pages) | Microfiche (pages) | Total (pages) |
| Agricultural documents | 122,432 | 106,517 | 228,949 |
| Technical documents from annual conferences | 65,392 | – | 65,392 |
| Technical documents from Kasetsart university expertise | 103,859 | 18,269 | 122,128 |
| Total | 291,683 | 124,786 | 416,469 |

## Problems related to document scanning

The main problem resulting from the scanned documents is clarity and length of time. A great deal of time is spent on embellishing the image documents using a graphic application.

## Time spent in production of digital documents by scanning

Scanning can be a somewhat time-consuming process due to its several steps: document scanning, image enhancing, image quality control and file management. The time spent depends on the quality of the documents, efficiency of the equipment and human resources available.

### Figure 7: Proportion of time used for producing e-documents
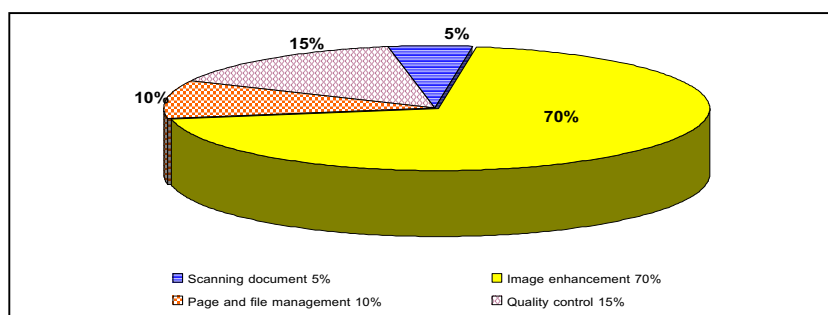


## Image-enhancement tools

The functions of image-enhancement tools are: deskew (the process that will remove skew from the image), auto crop (to eliminate vertical and horizontal marginal noise), auto align (to automatically align the text area into the paper centre) and noise removal. Using the figures presented in Table 4, it would take approximately six years to digitize all documents at the Thai AGRIS Centre, excluding the increasing number of new documents. Further development of efficient tools for Thai document image enhancement seems to be one way to solve the digitization-related problems.

### Table 4: Time saved by image cleaning tools

| Function | Time spent without tools (min / page) | Time spent by using tools (min / page) | Time saved (min / page) |
|---|---|---|---|
| Auto align | 3.30 | 1.30 | 2 (60%) |
| Noise removal | 6.36 | 3.25 | 3.11 (48%) |

# 7. Capacity building

It has taken a long time for information management to be accorded its value. The Thai AGRIS Centre has learned a lesson and as such has developed strategies and policies that promote management of information as a priority. Studies have been done, all aimed at improving the quality of service the centre offers. This started with improving the database and adding value-added services. The centre also is collaborating with other institutes on many projects that aim to improve the quality of the database and services. Annual requests for more staff, by the centre, bigger budgets and more equipment confirm the need.

The centre's services are promoted at every opportunity, including in occasional low-cost mediums such as leaflets and small ads in newsletters. Homepage linkage with other related Web sites is also extensively used. Exhibition and demonstration are frequently arranged in national and regional seminars. In addition, the centre invites guest speakers or lecturers on the use of the centre's agricultural database.

## Capacity-building activities

In 2000, Kasetsart provided the Thai AGRIS Centre two servers, which were installed with Linux and WebISIS programs to manage the Thai Agricultural Database with its 20 years of content. Three scanners also were allocated for digitalizing print materials from 1980 to present.

A project proposal for using the Thai agriculture thesaurus and knowledge-based development has been approved with financial support from Kasetsart University's Institute of Research and Development and in collaboration with FAO.

## Developing intelligent tools

The Thai AGRIS Centre envisages the development of a number of tools that will help improve the performance of the database and related information services in the future. These include: intelligent database management system, automatic indexing, automatic subject categorization and machine translator for agricultural documents, and image enhancement tools.

# 8. Enhancing Thai AGRIS management with NLP techniques

To enhance the performance of the Thai AGRIS management, both in knowledge repositories and service provision, the centre has collaborated with the Specialty Research Unit of Natural Language Processing and Intelligent System Development (NAiST), within the computer engineering department at Kasetsart University. Some of the cooperative work has been mentioned in the previous chapters but the following specifically looks at the efforts to enhance the performance of the Thai AGRIS system by applying natural language processing (NLP) techniques.

The Thai AGRIS system currently includes two important functions: storage and retrieval. In the storage subsystem, Kasetsart researchers have developed an annotation tool for attaching metadata to a document. Additionally, a converter tool is being used to generate text representation from major electronic file formats. An automatic indexing tool is used to extract important keywords from the document. An image-enhancement tool is used to remove noise from the document image. In the retrieval subsystem, the researchers have developed an intranet search engine that provides a standard Boolean-query mechanism. Furthermore, it provides query expansion using ontology knowledge for improving the retrieval process efficiency.

## Background and motivation

There are three main problems in Thai AGRIS management that affect the performance of both the information retrieval system and the machine translation system:

- Document preparation is done through manual typing and indexing, translating from Thai to English and English to Thai. Furthermore, 90 percent of Thai agricultural documents are hard copies that need to be scanned.

- Information access via Thai WebAGRIS is often difficult. Reasonable levels of precision and recall are only achieved when the search is done using keywords coming from controlled vocabularies. In addition, search engines can only retrieve data on a purely syntactic basis. It is not possible to embed common-sense, context or domain-specific knowledge into the search engines' queries.

- Thai-English AGROVOC needs constant maintenance in relation to translation, terms coverage (especially proper names and technical terms) and domain-specific restrictions (common words for scientific names, etc.). It is therefore necessary to extend the thesaurus database by recycling existing sources.

To enhance the Thai AGRIS management performance and address these three primary problems, the Thai AGRIS Centre has collaborated with Kasetsart's NAiST laboratory in a project called Development of Agricultural Information System for Knowledge Acquisition and Dissemination. It is a five-year project, which began in 2001, to integrate Thai language processing techniques to Thai agricultural information processing. It consists of:

- Developing annotation tools for supporting document preparation, such as automatic indexing and metadata translating;

- Developing semantic search engine for providing smart service with query expansion by utilizing Thai AGROVOC and ontologies extracted from AGROVOC;

- Developing tools for maintaining the Thai-English AGROVOC by extracting new terms from the corpus and technical dictionaries; and

- Developing machine-aided translation from Thai to English and vice versa.

## From STREDEO to Thai AGRIS management

STREDEO[3] multimedia, multilingual document storage, retrieval and delivery system for e-organization is a prototype document management system being developed by the NAiST laboratory (research started in 2001). Because the Thai AGRIS Centre was using the same concept model, the STREDEO was applied to the Thai AGRIS system management, consisting of automatic document indexing for both image and text, automatic document classification and dissemination, and intelligent search engine with query expansion:

- Based on automatic document indexing for both image and text, the document could be retrieved with full-text meaning instead of controlled vocabulary. This model could enhance both precision and recall of retrieved information.

- Based on automatic document classifying, the document will be classified according its domain concept. This model could support the process of meta-data annotation.

- Based on capturing the user's intention, the document will be disseminated according to his/her interest.

- Based on Thai AGROVOC ontology and NLP-based techniques, such as soundex and transcription, the search engine could work more efficiently both in precision and recall.

## Previous problems in document management systems of the Thai AGRIS Centre

In general, document management systems still have a few shortcomings that need to be resolved:

### Storage

Electronic documents usually store in a variety formats. Some of them are not open formats and the mechanisms for converting their content are not available.

Document representation is generally not enough for effective retrieval. A good designed tool must be developed to ease the task of annotation, such as including automatic indexing with the metadata annotating tool.

Maintenance of document images. Scanned documents may contain noise, which should be removed before further processes such as OCR.

### Retrieval

Use of controlled vocabularies in natural language processing on search queries causes low precision.

---

[3] Funded by KURDI (Kasetsart University Research and Development Institute).

In light of these problems, the STREDEO model has been applied to Thai AGRIS management as follows:

## Annotation tool for intelligent management of knowledge repositories

Figure 8 shows the overview of a developed system, consisting of an annotation tool, an image enhancement tool and a search engine module.

**Figure 8: Electronic document management system architecture**
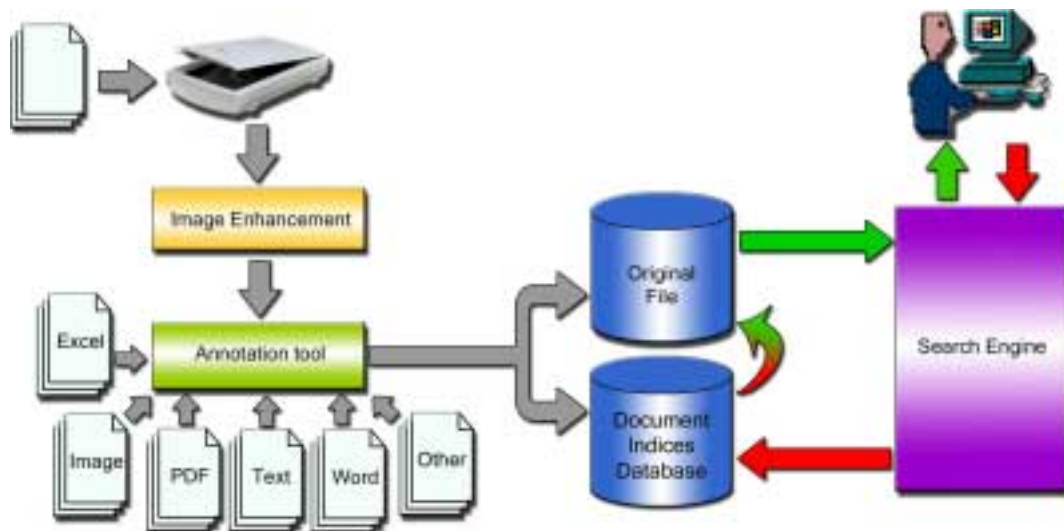


The annotation tool is responsible for generating document representation. It consists of three main components: converter (converts the electronic document to standard text file format), automatic indexing (used to generate keywords of document automatically) and annotation interface (allows user to attach metadata to the document).

To remove the noise from the document image, NAiST researchers have developed an image enhancement tool that is essentially a pipeline of three image enhancement processes consisting of deskew, auto crop and auto align.

## Smart search for smart service

The search engine module is derived from the Kasetsart University search engine, which uses a Thai text retrieval system for intranet document searching. It has two functions: basic search and advanced search. The basic search, like several search engines, provides a standard Boolean-query search mechanism.

The advanced search is used to increase the efficiency of a search engine. To improve the effectiveness of information retrieval, the query expansion has been developed. It uses knowledge from metadata information and an ontology in order to uncover indirect connections between query terms and document content. Both of them can improve searching performance.

## Automatic Thai ontology construction tool

Ontology plays an important role in many applications, such as information integration and document classification in taxonomies, including the information retrieval system. To enhance a search engine performance, ontology is used to boost the correctness and coverage of the system. However, there remains a problem of building large and adequate ontology within a short time frame and at low cost. The following present both a recycling approach to engineering ontology and general maintenance architecture for discovering new instances from text.

## An overview of the ontology construction and maintenance system

Figure 9 illustrates on overview of the ontology system, which contains four steps: lexical pre-processing (for identifying word strings), ontology term acquisition and relation extraction (to extract all single nouns and compound nouns from the tagged corpus and their relations), and term/relation organization (to organize the terms according to the relations).

**Figure 9: Thai ontology construction and maintenance system architecture**
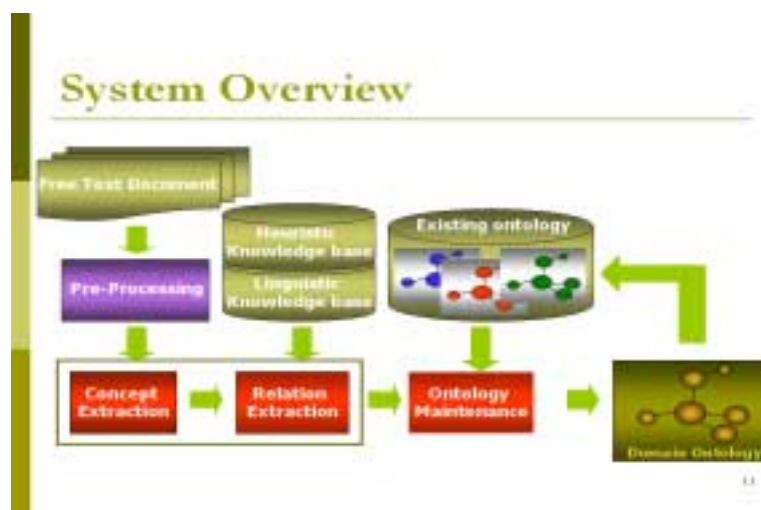


## From Thai AGROVOC to ontology relationship

The AGROVOC database was converted into the text file to recycle the original relation to ontology relationship (e.g. from BT/NT to IS-A). However, all BT/NT relations in AGROVOC could not be defined to the IS-A relation. Their semantic relations could be defined as "ingredient of" and "property of".

Heuristic method has been used to resolve this problem. A compound noun in NT will be processed to find a head noun; if head noun is consistent to BT, their relation will be defined to IS-A relation. If they are not compatible, the system will alert an expert to confirm this relation. This method will create many sub-trees where the root is a top term in AGROVOC. In the future those top terms will be extended by using the corpus-based ontology extraction algorithm for maintaining and constructing large ontology.

# 9. Knowledge management activities – Present and future plans

The Thai AGRIS Centre is pursing its goal to expand into an agricultural knowledge centre this year, which coincides with its 25th anniversary. The knowledge management activities conducted to date and its future plans are briefly explained in the following sections.

## Objective

The overall objective of the Thai AGRIS's current activities is to transform its function from an information and resources centre to an agricultural knowledge centre. With that target, a number of projects have been proposed in three areas of emphasis:

- Knowledge management
- Capacity-building
- IT development

## Knowledge management plan

## Knowledge acquisition

The plan is to assemble available agricultural information or to make it comprehensive and complete. Another activity will be to transform tacit knowledge, such as expertise of experienced academics or wisdom of local people or communities, into explicit knowledge. The acquisition process will be implemented through networking. Communication and information transfer will be carried out over the Internet using tools and methodologies developed on common standards. Coordination and collaboration will take place at three levels: organizational, personal and specialized group or interest group levels.

## Knowledge formalization

In the formalization process, assembled knowledge will be systematically classified, grouped, indexed and stored according to AGRIS standards, using available tools such as the AGROVOC thesaurus. Accordingly, the centre also plans to enlarge the Thai AGROVOC to a comprehensive Thai agriculture thesaurus as standard vocabulary for indexing Thai agricultural information.

## Knowledge dissemination

Disseminating knowledge to a wider audience will be done with support of tools such as web portals and intelligence retrieval service systems. To accommodate the system and make it more efficient in bringing knowledge to users, there is a necessity to cultivate a knowledge-sharing culture.

## Knowledge building and application

The centre plans to develop a knowledge mapping and capability management system.

## Capacity-building plan

Human resources are a critical factor for the Thai AGRIS Centre to achieve its goals. Their skills need to be developed and transformed from clerical to technical levels and therefore some routine duties will be outsourced to qualified persons. The centre's staff will have managerial responsibilities, including general management, monitoring and quality control.

Information specialists with agriculture expertise will be trained as knowledge managers to supplement other duties of indexing and cataloguing.

## IT development plan

IT development will be achieved in collaboration with the Specialty Research Unit in Natural Language Processing and Intelligent Information System Technology, computer engineering department and knowledge management research unit. The planned equipment acquisitions are as follows:

- In the knowledge acquisition phase, a scanner noise-cleaning system, knowledge extraction and generalization;
- In the formalization process phase, a Thai language processing system that will enhance the performance of automatic document annotation, document indexing and subject categorization;
- In the knowledge sharing and accessing phase, a machine translator for accessing Web-based English information; and
- In the knowledge tracking phase, a multi-viewpoint knowledge tracking system based on information extraction and ontology.

## Knowledge sharing via a multilingual Web-based active reading application

The objective of this project is to develop a reading assistant for the local farmers to access agricultural information on the Internet. The system consists of two components: active reading assistant and tabular translation. The first module applies the multilingual dictionary for word-level translation while the second module translates all tables in a Web page and converts all measurement units to more locally familiar units. The experiment is based on the FAOSTAT[4] corpus.

This project intends to overcome reading obstacles of users by developing a Web-based multilingual active reading system. Furthermore, a word-level translation with word reordering function will be provided for whole-page translation.

## Multilingual dictionary

In this project, the multilingual dictionary will serve as the knowledge base for word-level translation for many natural language processing applications, such as machine translation and cross-language information summarization. The multilingual dictionary from the Multilingual Service Working Group (http://ml-wg.cpe.ku.ac.th) will be used. This dictionary currently supports ten member languages: Bahasa Indonesian, Chinese, English, French, Italian, Japanese, Korean, Tagalog, Thai and Vietnamese. In order of generalization, English was selected as the intermediate language for lexical acquisition.

---

[4] For more information, visit http://apps.fao.org.

## Multilingual Web-based translation

The system consists of the following four principal components:

1.    Input analysis module to analyse information tables within the input page;

2.    Word-level translation module to translate each cell of tables with a multilingual dictionary;

3.    Unit conversion that converts each numeric cell whose measurement unit is not used in the target language into one of the target language; and

4.    Output generation to produce tables and display them to the end-users.

The Web-based multilingual active reading system was developed for information exchange. This research intends to surmount the reading obstacles experienced by the local users. Currently, the multilingual dictionary with English-Thai agricultural word alignments has been completed and paves the way for the launch of the English-Thai translation.

The techniques of language engineering, namely anaphora resolution, discourse processing, named entity recognition and knowledge engineering including information and knowledge extraction techniques, will be integrated for Thai AGRIS management. These techniques will develop an information extraction and knowledge summarization system focusing on Thai documents. Its expected role is to analyse and cluster crucial themes in order to produce a corresponding domain-specific knowledge.

To improve the performance of the document retrieval system, the following capabilities should be addressed:

**Information extraction.** This elicits thematic information of query results and performs as a decision clue because it allows users to decide reading more rapidly.

**Multiple-viewpoint knowledge tracking.** This determines the relationship of the query results and can represent it by multiple-viewpoint paths of continuity; for example, the result clustered by periods, the root causes of the problems, the methods to solutions, or the reader classes.

**Document taxonomies.** This subsumes and represents the query results in the taxonomic fashion.

**Knowledge discovery.** This extracts crucial knowledge from the documents.

Figure 10 presents the system architecture composed of two important parts: information extraction module and user-select viewpoint-generating module. The former is responsible for extracting the relevant information from interested document by using information extraction and natural language processing techniques. An ontology will be used in both modules to aggregate data.

**Figure 10: System architecture information extraction and tracking**
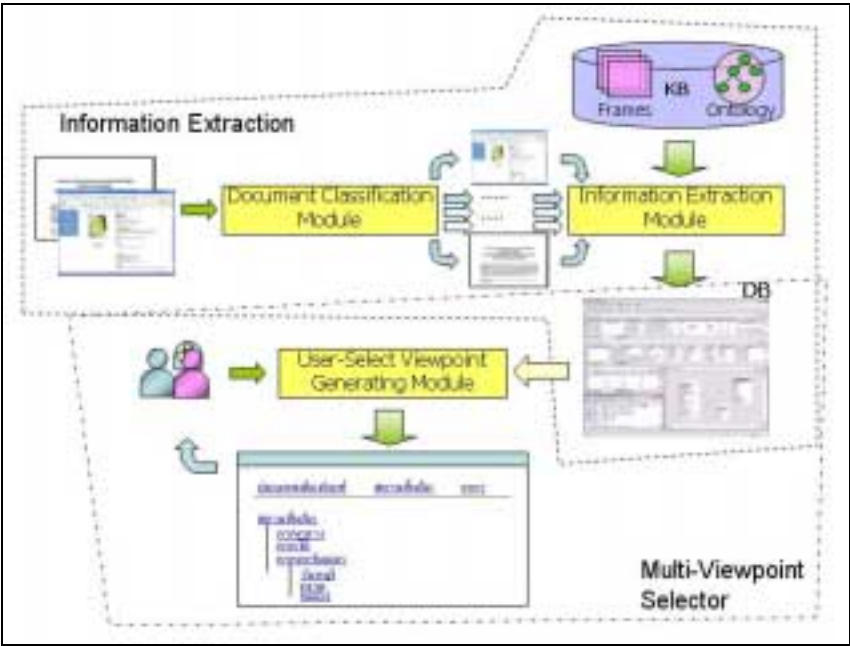
# References

1.  Bennukul, Krisna. 2003. *Behavior in using online databases for agricultural information searching by graduate students in Agricultural Science*, Kasetsart University. Thesis MA. Library and information science. Thammasat University, Bangkok. 130 p.

2.  FAO. 1994. *AGRIS/CARIS twentieth anniversary celebration proceedings*, FAO, Rome. 48 p.

3.  FAO. 2003. *The new vision for AGRIS*. Available source: http://www.fao.org/agris/.

4.  FAO. 2003. *AGRIS – A strategy for an international network for information in agricultural sciences and technology within the WAICENT framework*. Available source: http://www.fao.org/DOCREP/ MEETING/005/AC502E.HTM.

5.  Institute for Computer and Information Engineering. 2000. *WWW/ISIS technical reference manual*. v. 5.1.0/00-12-01. Warsaw. 125 p.

6.  Institute for Computer and Information Engineering. *WWW-ISIS: A painless way for implementing data entry for ISIS databases on WEB*, Tutorial, part II v. 1.0. Warsaw, June 2001. Available source: http://www.icie.com.pl/. 75 p.

7.  Kawtrakul, A. et al. "A statistical approach to Thai word filtering", pp. 398-406. in *The second symposium on natural language processing*, Bangkok.

8.  Kawtrakul, A., C. Thumkanon, T. Jamjanya, Muagyunnan, K. Poolwan, Y. Inagaki, 1996. *A gradual refinement model for a robust Thai morphological analyzer*, pp. 1086-1089. *In* COLING 96.

9.  Kawtrakul, A. et al. 1997. "Automatic Thai unknown word recognition", pp. 341-346. In *Proc. of NLPRS*.

10. Kawtrakul, A. et al. 1998a, Towards Automatic Multilevel Indexing for Thai Text Information Retrieval, pp. 551-554. *In* Proc. of IEEE.

11. Kawtrakul, A. et al. 1998b. Backward Transliteration for Thai Document Retrieval, pp. 563-566 *In* Proc. of IEEE.

12. Prachyothin, Narumol, Taveesak Koanantakul, Premin Chindavimollert. 1993. *Thesaurus and information system*. National Electronic and Computer Technology Centre, Bangkok. 188 p.

13. Thunkijjanukij, Aree. 1999. *Thai agricultural information management system*. Thai National AGRIS Centre, Kasetsart University, Bangkok. 54 p.

14. Thunkijjanukij, Aree. 2003. Working paper 1985-2003. Thai National AGRIS Centre.

15. Thunkijjanukij, Aree, Wanphen Srijankul. Thiranan Damrongson. 2003. *Thai WebAGRIS: Agricultural information management and retrieval system*. Thai National AGRIS Centre, Kasetsart University, Bangkok. 65 p.

16. Wattanapongse, Piboonsin and Aree Thunkijjanukij. 2000. *Thai AGRIS Centre: 20 year of development*. Kasetsart University, Bangkok. 74 p.

# Appendix 1
## Thai AGRIS Centre outputs

The Thai AGRIS Centre has been applying WebAGRIS for bi-lingual database management service both through the Internet and CD-ROM. Some major outputs are as follows:

1. **Thai agriculture database,** which is bi-lingual (Thai and English) with information dating from the present back to 1997 (from 1980-1996 is in Thai only). It is disseminated via the Internet. Only bibliographies with summaries are provided.

2. **The Kasetsart University annual conference database** is a record of research papers presented at conferences between 1962 to present. It contains the bibliographies and full text e-documents. It is disseminated through the Internet and CD-ROMs for purchase.

3. **Kasetsart University open archive database.** This is a bibliographical database with full text document. It includes academic papers of professors' literature.

4. **Agricultural experts database.** This is a proposed project with the objective of collecting academic papers and articles of lecturers, researchers and experts in agricultural science. It will offer bibliographies and e-full text document.

5. **Animal feeds database** contains animal feedstuffs research papers in Thailand. It is a cooperative project between the Faculty of Agro-Industry and the Thai AGRIS Centre.

6. **The database for fruit pests and diseases** consists of research papers from around the world on pest and pathogens of Thai fruits. The database concentrates on major export products such as mango, longan, lychee, rambutan and pineapple. The project is under the umbrella of a pilot project on pest-risk assessment for exporting Thai fruits, a collaborative effort between Kasetsart University and the Ministry of Agriculture and Cooperatives.

7. **Thai swamp buffalo database** collects articles and research papers regarding swamp buffalo in Thailand, in collaboration with the International Buffalo Information Centre.

8. **Thai cassava database** collects articles and research papers regarding cassava in Thailand, in collaboration with the Thai Cassava Foundation.

Thai AGRIS Centre timeline

## Development of AGRIS information system and the Thai AGRIS Centre

| Year | Development |
|------|-------------|
| 1969 | FAO, Commonwealth Agricultural Bureau and the National Agricultural Library conduct a study to establish AGRIS system. |
| 1974 | The international information system for the agricultural sciences and technology (AGRIS) is established and provides training to member countries. |
| 1975 | The Kasetsart University library is selected to host the national AGRIS Centre. Because of insufficient personnel, the initial responsibilities are given temporarily in January to the Thai National Documentation Centre of the Science and Technology Research Institute, although retaining close cooperation with the Kasetsart library. <br><br> The First Thai AGRIS input on paper form is submitted through the Agricultural Information Bank for Asia. |

| Year | Cooperation | Information Technology Development | Information Service |
|------|-------------|-----------------------------------|---------------------|
| 1980 | The Kasetsart University library becomes the Thai AGRIS Centre and initiates Thai agricultural information development at Kasetsart University (3/1980). | | • Provides agricultural information service through the bibliography-AGRINDEX (1975-1994), AGRIASIA (1977-1996). <br> • Initiates document delivery service through AGLINET. |
| 1982 | | • Created the Thai Agricultural Database. <br> • Published the first volume of Thai Agricultural Bibliography | |
| 1983 | Organizes the first workshop to create cooperation network amongst agricultural institutes, agricultural libraries and the concerned agencies. | | Have the first demonstrating on agricultural information retrieval: On-line service to AGRIS Processing Unit, Austria by connecting to International Database Access and Remote Computing Service. (28/10/1983). |
| 1985 | | The Kasetsart library obtains its first computer with the support of IDRC. Inputting using Dbase III Plus. | Provides AGRIS SDI services. |
| 1986 | | Submits first AGRIS input in electronic form, sending diskette via mail. | Starts the AGRIS online data retrieval service (AGRIS/STAIRS) by connecting to AGRIS processing unit in Vienna. The service was done through international communication service, Communication Authority of Thailand (9/12/1986). |

| Year | Cooperation | Information Technology Development | Information Service |
|------|-------------|-----------------------------------|---------------------|
| 1987 | | • Changes database application to Micro CDS/ISIS version 1.0.<br>• Micro CDS/ISIS version 2.3, in December 1989.<br>• AGRIS on CDS/ISIS version 2.0 and AGCHKE base on Micro CDS/ISIS version 2.3 in January 1991. (Technical note 26, 27).<br>• AGRIN version 2.0, in January 1993.<br>• AGRIN version 3.0 and AGCHKE version 3.0, in December 1997.<br>• WebAGRIS Beta version in January 2001. | |
| 1990 | Thai AGRIS Centre receives AGRIS and FHN CD-ROM from FAO. | With financial support from NECTEC, Thai AGRIS Centre receives its first PC. | |
| 1991 | | The Kasetsart library acquires a PC with CD-ROM drive and a set of printer for AGRIS CD-ROM service (7/1991). | Thai AGRIS Centre announced AGRIS CD-ROM data retrieval service. |
| 1992 | Submitted AGRIS input directly to APU with out sending data to AIBA in Philippines. | Started data entry for Thai Agricultural Database. Transfer data from University Computer Centre's server and re-entry some lost data. | |
| 1994 | • AGRIS/FAO announces sending AGRIS input via email, FTP.<br>• Internet access to AGRIS online and use telnet to IAEA (third technical consultation of AGRIS and CARIS participating centres: 6-10 June 1994) | | |
| 1994 | | Cables and installs local area network in the library building. Sets up server system to enlarge data retrieval service from AGRIS CD-ROM and Thai agricultural database (6/1994). | |
| 1995 | FAO establishes World Agricultural Information Centre (WAICENT) | | |
| 1995 | | Designs and installed database CDS/ISIS system on the LAN. Used search Interface program from CIAT for internal information retrieval. Installed CD-NET for AGRIS CD-ROM retrieval in local network for users to retrieve, print and download data. (1/1995). | Provides AGRIS CD-ROM data retrieval service via local area network. The data available are Thai Agricultural Database, AGRIS and FHN CD-ROM since 1975 (3/7/1995). |

| Year | Cooperation | Information Technology Development | Information Service |
|---|---|---|---|
| 1997 | Promotes cooperation between countries in the region by carrying out the agricultural information development project for the forestry department library in Lao PDR. | Implements KU CD-NET system for AGRIS and FHN CD-ROM information retrieval on Kasetsart University campus network (1/1997). | Tests KU CD-NET system, evaluating AGRIS CD-ROM information retrieval service and survey for user requirement. |
| 1998 | *Fourth technical consultation of AGRIS and CARIS participating centres:* Strengthening AGRIS network's role as a provider of information and knowledge in sustainable agricultural development, 8-11 June 1998. | | |
| | | Submits AGRIS data input via Internet using email and FTP (9/1998). | • Provides AGRIS CD-ROM retrieval system on campus network: KU CD-NET (5/1998).<br>• Assistance to the library in the forestry department of Lao PDR under the support of FAO/FORSPA to develop a Lao forestry database. |
| 1999 | • AGRIS processing unit in Vienna is terminated. AGRIS process transfers to Rome.<br>• AGRIS database is available on the Internet. | | |
| | | | Promotes access to AGRIS Database via Internet. |
| 2000 | • FAO provides an opportunity for participants of AGRIS input centres in the Asia-Pacific region to attend a WAICENT workshop in Japan.<br>• Dr. Stephen Rudgard is the keynote speaker for the 20th anniversary of Thai AGRIS Centre conference, 6-7 July 2000.<br>• The first consultation on agricultural information management (cir 2/2000) takes place 5-7 June 2000. | | |
| | • Representative from Thai AGRIS Centre attends WAICENT workshop in IWS2000, Tsukuba, Japan.<br>• Organizes 1st conference on Information Technology for Agriculture, 6-7 July 2000. | • Creates Thai National AGRIS Centre Homepage.<br>• Provides Thai AGRIS Database information retrieval service on web.<br>• Set up Linux and WebSIS server. | Establishes the agricultural Web portal with both local and international information. |
| 2001 | • Thai AGRIS Centre is nominated as a centre of excellence AGRIS network.<br>• Thai AGROVOC and Thai Agricultural Knowledge-Based Development Project was approved.<br>• Collaborates with FAO on the Thai AGROVOC development.<br>• Meets with Thai AGRIS network agencies on Development of Agricultural Information System for Knowledge Acquisition and Dissemination (8/2001). | • Kasetsart University provides two servers and a scanner for the centre's document digitalization project.<br>• Starts digitalizing documents.<br>• Replaces the archive collection in microform with digital scanning.<br>• Starts developing Thai AGROVOC.<br>• Creates Thai WebAGRIS for data entry in both Thai and English languages. | |

| Year | Cooperation | Information Technology Development | Information Service |
|---|---|---|---|
| | • Trains staff on development of the Thai AGROVOC and Thai WebAGRIS at FAO Rome, Italy (11/2001). | | |
| 2002 | Attends WebAGRIS workshop at the Institute for Computer and Information Engineering in Brussels, Belgium. | Thai AGRIS centre installs a system for providing agricultural e-document services (10/2002). | |
| 2003 | Studies and develops agricultural knowledge management system and multilingual ontology. With the collaboration of the specialty research unit of Natural Language Processing and Intelligent Information System Technology, Department of Computer Engineering, Kasetsart University, <br>• In collaboration with FAO, organizes the first *Workshop on agricultural information management: WebAGRIS system and Thai agricultural information networking.* | • Tests Thai WebAGRIS system for Thai information retrieval and data entry through Internet. <br>• Tests Thai AGROVOC service via Internet for public feedback and improve for the beta version. <br>• Testing Thai AGROVOC for Thai WebAGRIS data entry and indexing. | Provides database management application, Thai WebAGRIS and also the utilization of Thai AGROVOC. |
| 2004 | • Strengthen collaboration with other partner in AOS workshop, Beijing (4/2004). | • Finished first edition of Thai AGROVOC. <br>• Develop Thai agricultural thesaurus and Thai agricultural ontology. <br>• Finished developing and testing Thai WebAGRIS. <br>• Officially use WebAGRIS system for Thai AGRIS Centre data entry (1/12/2004). <br>• Transfer all Thai AGRIS database in AGRIN 3.0 format to Thai WebAGRIS system. | • Announce first edition of Thai AGRIVOC for indexing and query expansion. <br>• Provide Thai AGRIS database on Thai WebAGRIS system via Internet. <br>• Assistance to NAFRI, Lao PDR, under the support of SIDA and in collaboration with FAO to develop a Lao AGROVOC and Lao agricultural database. |
| 2005 | • Organized second conference on information technology for agriculture, 9-10 March 2005, in occasion of 25th anniversary Thai AGRIS Centre, in collaboration with FAO. | • Finished first phase of document digitization. | • Announced agriculture e-document service. <br>• Provided Thai agricultural knowledge-based service. |

# Appendix 2
## List of dictionaries for Thai AGROVOC translation

Osodsapa, Y. 1999. **Dictionary of Fertilizer.** Kasetsart University Publ., Bangkok.

Photipuk, P. and N. Chansae. 1997. **Veterinary Medical Dictionary English-Thai.** Chulalongkorn University Publ., Bangkok.

Prince of Songkla University and King Mongkut Institute of Technology Ladkrabang. 1992. **English-Thai Dictionary of Agriculture: Agro-industry.** Odeanstor Publ., Bangkok.

Prince of Songkla University and King Mongkut Institute of Technology Ladkrabang. 1993. **English-Thai Dictionary of Agriculture: Animal Husbandry.** Odeanstor Publ., Bangkok.

Prince of Songkla University and King Mongkut Institute of Technology Ladkrabang. 1994. **English-Thai Dictionary of Agriculture: Agricultural Machinery.** Odeanstor Publ., Bangkok.

Prince of Songkla University and King Mongkut Institute of Technology Ladkrabang. 1995. **English-Thai Dictionary of Agriculture: Crop Husbandry.** Odeanstor Publ., Bangkok.

Prince of Songkla University and King Mongkut Institute of Technology Ladkrabang. n.d. **English-Thai Dictionary of Agriculture: General Agriculture.** Odeanstor Publ., Bangkok.

Saivichain, P. 1994. **Home Economists' Word List.** Klangwaingkanpim Co., Chiangmai.

Smitinand, T. 2001. **Thai Plant Names.** Royal Forest Department, Bangkok.

Soil Science. 1998. **Glossary of Soil Science.** Kasetsart University Publ., Bangkok.

Suwanpratasd, S. and K. Panburana. 1999. **Dictionary.** Witayapat Co. Ltd,

Suwanwajokkasij. 1987. **Dictionary of Agriculture.** 2nd ed. Library Kasetsart University, Bangkok.

The Royal Institute. 1984. **English-Thai Dictionary of Forest: Forest Products.** Pisansinkanpim Publ., Bangkok.

The Royal Institute. 1998. **English-Thai Dictionary of Botany.** Arunkanpim Publ. Co. Inc., Bangkok.

The Royal Institute. 1999. **The Royal Institute Dictionary.** Nanmebook Publications Ltd, Bangkok.

Thiengburanathum, W. 1996 a. **Dictionary of Plants and Animals in Thailand.** 2nd ed. Raumsan (1977) Ltd, Bangkok.

Thiengburanathum, W. 1996 b. **English-Thai Dictionary.** Thaimitkanpim Publ., Bangkok.

Thiengburanathum, W. 1999 a. **Dictionary of Floriculture and Ornamental Plant.** 4th ed. Raumsan (1977) Ltd, Bangkok.

Thiengburanathum, W. 1999 b. **Dictionary of Thai Herb.** 5th ed. Raumsan (1977) Ltd, Bangkok.

Thai National Economic Information Coordinating Centre. 2001. **Economic thesaurus.** Chuanpim Publication Inc., Bangkok