

ESAKIA file, One of the Public Taxon-based
Entomology Database KONCHU Produced at the
Computer Center of Kyushu University

Tadauchi, Osamu

<https://doi.org/10.5109/2579>

出版情報 : ESAKIA. 34, pp.227-237, 1994-03-24. Entomological Laboratory, Faculty of
Agriculture, Kyushu University

バージョン :

権利関係 :

ESAKIA file, One of the Public Taxon-based Entomology Database KONCHU Produced at the Computer Center of Kyushu University¹⁾

Osamu TADAUCHI

Entomological Laboratory, Faculty of Agriculture,
Kyushu University, Fukuoka, 812 Japan

Abstract. A database file ESAKIA, one of a public "Entomology Database KONCHU", based on the journal Esakia is produced at the Computer Center of Kyushu University. The KONCHU is a general database name including various files of bibliographical and image databases. The former is based on main Japanese entomological journals and available by an on-line network to the world. It is a taxon-based database, treating a taxon as one record, i.e., species, genus, family, etc. Each record is constituted of 13 items, such as bibliographical, taxonomical, distributional data and key words. The ESAKIA file includes 6,326 records through 1993. Several retrievals on the ESAKIA file were executed and show that most records consist of taxonomic ones(80.30%), and of two orders, Coleoptera (63.71%) and Hymenoptera (25.88%), distributed in the Australian Region (New Guinea and Australia), as well as the Palaearctic and Oriental regions (Japan, Korea, China, Taiwan and South East Asia). Examples of use of the file using the database management system SIGMA are presented.

INTRODUCTION

Data on organisms have characteristics to provide for a database production. Structures and functions of organisms have been described continually, and the content of the data have been added, deleted and changed unceasingly. For the effective use of the enormous data accumulated, necessity and importance of the database production has been increased.

A biological database BIOSIS (Bioscience Information Service, Philadelphia) in the U.S.A. includes a great many bibliographical records since 1969. The BIOSIS began to edit "Zoological

1) Contribution from the Entomological Laboratory, Faculty of Agriculture, Kyushu University, Fukuoka (Ser. 4, No. 65).

Record" in cooperation with the Zoological Society of London since 1981 and has published from the journal of 1978, vol. 115. These data were processed for database production and available by an on-line network. However, entomologists must go back to the past and retrieve references on applied entomology as well as taxonomy. Especially taxonomists usually retrieve references published 50 years before and more for identifying species names. Although recent published records are available from BIOSIS, it only contains records within about 25 years. Moreover, most of the articles written in Japanese local journals are not cited in such databases. In entomology BIOSIS has limited value. For filling up the vacancy, it is thought to be important for entomology in Japan and Eastern Asia to produce a database containing old references, even if it would be a local database.

I produced an entomology database DMUSHI (the number of records: 8,870) based on 50 volumes (1928-1985) of the journal *Mushi* published by the Fukuoka Entomological Society and published a subject index from it (Tadauchi, 1985). The DMUSHI treats one **taxon** as one record unit with bibliographical data and key words. It became a public database at the Computer Center of Kyushu University and was opened to the public by an on-line network. After that I composed a program producing an entomology database KONCHU which is based on the main Japanese entomological journals and have organised to develop several files. I also have recently included an image database of insects into the KONCHU.

In the present paper outlines of a public database KONCHU, the ESAKIA file and the database management system SIGMA, with examples of use of the file are presented.

OUTLINES OF THE DATABASE KONCHU AND THE FILE ESAKIA

The database KONCHU is a general database name including various files of bibliographical and image databases. KONCHU is based on main Japanese entomological journals since 1900 and treats one journal as one file. It has a taxonomical feature as well as a bibliographical one because it treats one **taxon** as one record with information on taxonomy, biology, morphology, biogeography, ecology, etc. The records of KONCHU are written mainly in English, and in Japanese with katakana and kanji. The other is an image database based on various insect image data including photographs, direct images through a micro-lens or a microscope and various figures for each **taxon**.

The **ESAKIA** file is one of bibliographical databases based on the journal *Esakia*, Kyushu University Publications in Entomology. The journal was published from 1960 from the Hikosan Biological Laboratory, Faculty of Agriculture, Kyushu University. From 1978, No. 11, it has been edited in cooperation with the Entomological Laboratory, Faculty of Agriculture, Kyushu University. The number of records in the ESAKIA file through 1993 (No. 33, including a special issue) is 6,326. New records will be added to the file annually. The name of the ESAKIA file for retrievals is "S:A71414B.ESAKIA" (copyright of the ESAKIA file: O. Tadauchi).

At present, several files of the KONCHU based on the Japanese entomological journals will be prepared to be opened to the public at the Computer Center of Kyushu University, *i.e.*, MUSH1 (Journal of the Fukuoka Entomological Society), KONTYU (Journal of the Japanese Entomological Society), ODOKONA and ODOKONB (Japanese Journal of Applied Zoology and Entomology, A

and B), MATSUMURANA (Insecta Matsumurana, Hokkaido University Publications in Entomology), SHIKOKU (Journal of Shikoku Entomological Society), AKITSU (Journal of Kyoto Entomological Society), and so on.

ACCESS TO THE KONCHU

A user who wants to use the KONCHU must register at the Computer Center of Kyushu University in advance. Through inter-university networks, such as Nlnet and JAIN, etc, which are connected with an INTERNET in the world network, the KONCHU is available virtually anywhere in the world. A guide how to access to the Computer Center of Kyushu University is shown in "A guide to on-line databases at University Computer Centers in Japan" edited yearly (13th edition, 1993) by the Committee of Libraries and Databases, University Computer Centers, Japan.

TEXT DATABASE MANAGEMENT SYSTEM, SIGMA

1. Outline

The ESAKIA file of KONCHU is managed by a text database management system, "SIGMA" (Arikawa *et al.*, 1987, 1988) working at the Computer Center of Kyushu University. The system SIGMA was first implemented on FACOM M-780 OS IV/F4 MSP at the Computer Center in 1981 and has been serving researchers for natural language analysis (Higuchi, 1987), entomology database development (Tadauchi, 1987, 1988; Tadauchi *et al.*, 1990) and so on. It stores data simply as the form of text files, i.e., strings of characters, and exploits very precise and fast one-way sequential processing methods by pattern matching algorithms. It is designed to cope with text files and allows complicated retrievals, editing and refiling of retrieved records.

2. Spaces of SIGMA

The files in SIGMA consists of three parts, MEMO, SIGMA and EXTERNAL spaces. Each user can have these three spaces and makes files in a MEMO space for his own use. The files in SIGMA space are for sharing data with other users. All the files in the SIGMA space have "S." at the head of their names, for instance, "S. ESAKIA" for the ESAKIA file. The EXTERNAL space is for the files under the operating system, terminals, printers, etc. The sharing of files in the system SIGMA is shown in Fig. 1, where an arrow represents the permission of access.

The files in a MEMO space are further classified into five sections, MEMO, WORK, LOG, INDEX and BACK-UP. Each section except MEMO is also a stack of files of bounded depth. The roles of these sections are as follows: 1) MEMO: Private files of a user are stored; 2) WORK: Working area for a user. A file in this section is called a work file. Usually, a file in another section is moved or copied onto a work file for processing; 3) LOG: Most communications between SIGMA and a user are recorded in a file called a log file. This file, for example, helps to construct a command procedure from a sequence of commands which is already tested for correct run. In this

sense, the use of log files realizes the learning ability in the system; 4) INDEX: An execution of SEARCH command produces a file called an indexing file. Such files are stored in this section. The files in the INDEX section are the only files which are not text files; 5) BACK-UP: The files flowing over from the stacks in the WORK, LOG and INDEX sections, and the files deleted from MEMO section are collected in this section. Some of the files in BACK-UP are deleted only if all space in a MEMO space is exhausted.

3. Commands

Main commands used in the SIGMA system are listed in Table 1. Shortened commands underlined are available.

DATA FORMAT IN THE ESAKIA FILE

1. Items and tugs

One record treated in the ESAKIA file is the same in all the other files in KONCHU and is made for one taxon in a given paper. The ESAKIA file mainly consists of insect records, with a small amount of spider and mite records. One record includes 13 items with tugs in parentheses as follows:

- 1) **taxon** name (scientific name) (TAX)
- 2) **taxon** name (Japanese name) (JTAX)
- 3) author name (**AU**)
- 4) title (**T**)
- 5) journal (J)
- 6) volume number & page (**VNP**)
- 7) published year (**Y**)
- 8) order name (OR)
- 9) family name (**FAM**)
- 10) synonym (SYN)
- 11) distributional area (**DST**)
- 12) key words (KEY)
- 13) note (NOTE)

The example of one record is shown in Fig. 2.

2. Replaced characters

In the ESAKIA file there are some characters to which a user must pay attention, because of replacement of characters or letters.

- 1) Letters between capital and lower case are distinguished. Therefore, capital letters are used as in normal use, such as the first letters of a proper noun and a generic name, with the exception of a family name in "author item (AU)", where all letters are capital, for instance, "TADAUCHI".
- 2) Alphabetic and numeral letters are made by 1 byte, while Japanese letters, *i.e.*, hiragana, katakana

and kanji by 2 bytes. Therefore, a user must use 2 byte-characters when he retrieves by Japanese name of a given insect in katakana.

3) The following letters are not used in the SIGMA system and replaced other characters as follows:

á = 'a; â = ^a; ä = "a; à = ~a; å = %a; ç = *c

4) A blank is treated as one letter and has a meaning.

RETRIEVAL IN THE ESAKIA FILE

A command SEARCH in the SIGMA system is used for retrieval of the ESAKIA file. This command simultaneously deals with plural questions by using plural key words. The system SIGMA asks some questions after the session opens, as follows.

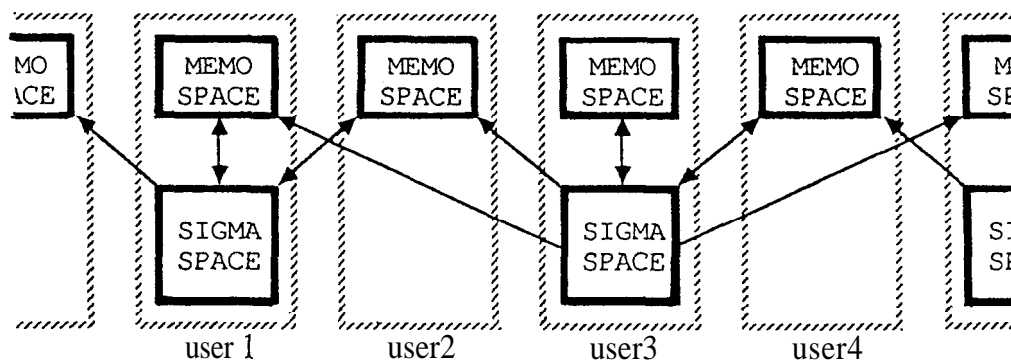


Fig. 1. File sharing in SIGMA system. Each user has his own MEMO space and shares a SIGMA space with the other users. An arrow represents the permission of access (after Arikawa et al., 1988).

```
#
(TAX) Stylops aburanae Kifune et Maeta
(JTAX) アブラナヒメハナバチネジレバネ
(AU) KIFUNE, Teiji; MAETA, Yasuo
(T) Ten new species of the genus Stylops (Strepsiptera, Stylopidae) parasitic
on the genus Andrena (Hymenoptera, Andrenidae) of Japan (Studies on the Japanese
Strepsiptera XIII)
(J) Esakia
(VNP) Special Issue (1): 98-99, (97-110)
(Y) 1990
(OR) Strepsiptera, ネジレバネ目
(FAN) Stylopidae, ネジレバネ科
(SYN)
(DST) Japan (Honshu)
(KEY) taxonomy; new species; description (female); host; remarks
(NOTE) type (female); type locality (ex. Todai, Ina, Nagano Pref., Japan); type depository
(Kyushu Univ., no. 2777); host (Andrena (Andrena) aburana)
#
```

Fig. 2. An example of a record in the ESAKIA file.

Table 1. A list of main commands in the SIGMA system.

SEARCH	:	search multiple keys
REPLACE	:	replace multiple keys
SORT	:	extract records and sort them
REFILE	:	make a text file from an index file
CATENATE	:	catenate files
COPY	:	copy a file
DELETE	:	remove a file and add it to BACK-UP
GET	:	move a file to WORK
KEYIN	:	write onto WORK
LOAD	:	load a file onto WORK
MOVE	:	move a file
PUT	:	put the top work file
SAVE	:	store the top work file
LOOK	:	show the topmost work file

1. Record delimiter

The command SEARCH distinguishes between two record delimiters and regards words or data sandwiched in between them as one record. As "#" is used for the record delimiter in the ESAKIA file, a user must reply to a system "#" about the record delimiter.

2. Item delimiter

If a user wants to retrieve the ESAKIA file by designating specific items, such as "AU"(author name), "DST"(distributional area), and so on, he must reply to a system "!" about an item delimiter. If a user does not designate items in the retrieval, he may reply only by a return key.

3. Key words

Next, the system asks about key words. In the command SEARCH, key words are assigned to keyword variables, A1, A2, A3 . . . in inputting order. The triple dot "..." represents a wild card.

4. Logical formulas

Logical formulas are used for questioning. A logical formula is composed by using keyword variables (A1, A2, . . .). formula variables (Z1, Z2, . . .). integers, logical operators ("," (or), "." (and), "^" (not)), comparison operators (<,>, . . .). arithmetic operators (+, -, *, /) and parentheses. A logical formula is assigned to each logical variable. For each logical variable, the SEARCH command retrieves the records defined by its logical formula. Representatives are as follows:

Z1 := A1.A2	A1 and A2
Z2 := A1,A2	A1 or A2

Z3 := ^A1 not A1

5. Designation of a file

After registering logical formulas, the system asks whether a user wants to display the results of retrieval("Y") or not("N"). Next, SIGMA asks the name of a file. The name of the ESAKIA file is formally "S.'A71414B.ESAKIA". But if a user designates "A71414B" by using a subcommand PREFIX of a command PROFILE for the first time, a shortened name, "S.ESAKIA" is accepted.

6. Display of results

The system displays the result of retrieval after showing the number of record and the CPU time (sec/1000) for each logical formula. Next, as shown in Example 1, the system asks some questions and then displays the results, if a user wants.

RESULTS OF SOME RETRIEVALS OF THE ESAKIA FILE

Some retrievals were executed using some key words for examining the ESAKIA file.

1. Fields (Fig. 3)

The ESAKIA file mainly consists of taxonomic records (key word: "(KEY)...systematics,taxonomy"; 5,080 records), followed by biological ones (233) and morphological ones (40).

2. Orders (Fig. 4)

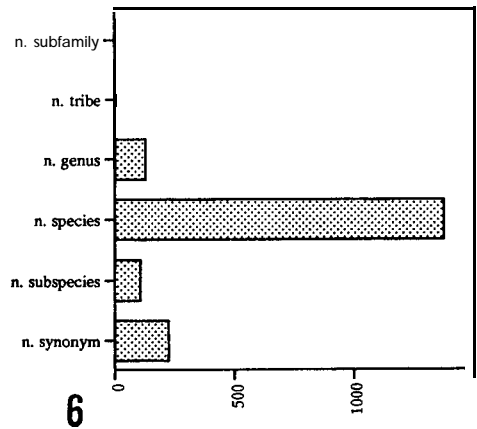
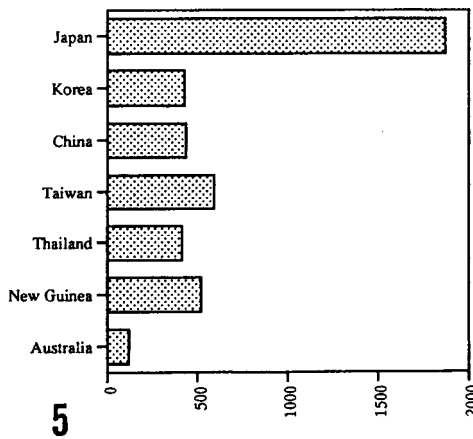
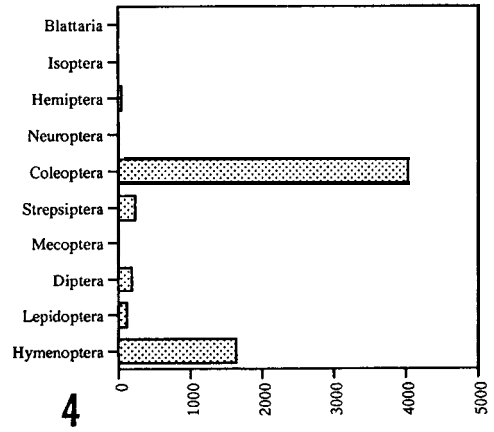
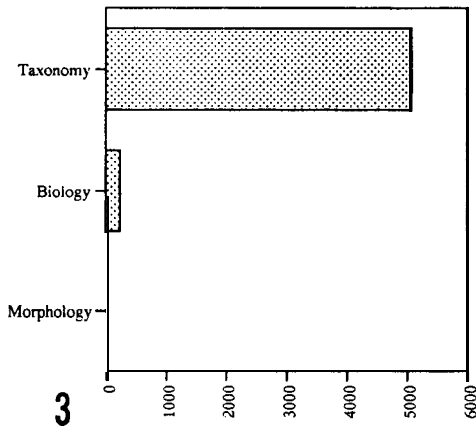
The records of the file are composed of two main orders, Coleoptera (key word: "(OR)...Coleoptera"; 4,030 records) and Hymenoptera (1,637), followed by Strepsiptera (224), Diptera (176), Lepidoptera (113), Hemiptera (48), Blattaria (5), Mecoptera (2) and Isoptera (1) in decreasing order with Arachnida.

3. Distributional area (Fig. 5)

The distributional data in the item (DST) in the ESAKIA file are made up of records from Japan (key word: "(DST)...Japan"; 1,871), followed by Taiwan or Formosa (592), New Guinea (520), China (438), Korea (432), Thailand (417) and Australia (120).

4. New taxa (Fig. 6)

New taxa in the ESAKIA file are as follows: new subfamily (key word: "(KEY)...new subfamily"; 2), new tribe (3), new genus (126), new species (1374), new subspecies (104) and new synonym (226).



Figs. 3-6. The number of records for some key words retrieved on the ESAKIA file. 3: fields; 4: orders; 5: countries; 6: new taxa.

ACKNOWLEDGEMENTS

My sincere thanks are due to Prof. K. Morimoto and Prof. Emeritus Y. Hirashima of Kyushu Univ. for their suggestions in developing of the ESAKIA file. I am indebted to the following persons in the SIGMA project group in Kyushu University: Prof. S. Arikawa, Prof. T. Miyahara, Mr. Y. Kawasaki and Mr. J. Inoue. I am thankful to Prof. Emeritus G. T. Riegel of Eastern Illinois Univ. for his reading the original manuscript. This work was supported by a Grant-in-Aid Scientific Research from the Ministry of Education, Japan.

EXAMPLES OF RETRIEVAL OF THE ESAKIA FILE

Examples 1 and 2 show the use of the retrievals of the ESAKIA file.

```

READY
SIGMA
SIGMA>SEARCH
RECORD DELIMITERS
D01:=#
D02:=
ITEM DELIMITERS
D02:=!
D03:=
KEYWORDS
A01:=(DST)...Japan
A02:=(DST)...Korea
A03:=(DST)...China
A04:=(DST)...Taiwan
A05:=(DST)...Formosa
A06:=(DST)...Thailand
A07:=(DST)...New Guinea
A08:=(DST)...Australia
A09:=
LOGICAL FORMULAS
Z01:=A1
Z02:=A2
Z03:=A3
Z04:=A4,A5
Z05:=A6
Z06:=A7
Z07:=A8
Z08:=
REPORT(Y/N)?Y
FILE:=S. ESAKIA
RETRIEVED TEXTS
QUESTION 01 (201) = 1871 1871
QUESTION 02 (202) = 432 432
QUESTION 03 (203) = 438 438
QUESTION 04 (204) = 592 592
QUESTION 05 (205) = 417 417
QUESTION 06 (206) = 520 520
QUESTION 07 (207) = 120 120
TOTAL = 3507 3507
CPU (SEC/1000) = 579 579
FILE:=
DO:REFILE
REPORT(Y/N)?N
QUESTION:=2
NEW RECORD DELIMITER:=#
NUMBERING(N/Y)?N
OUTPUT-FILE:=-T
#
(TAX) Chrysomelidae
(AU) KIMOTO, Shinsaku;KAWASE, Eiji
(T) A list of sane chrysomelid specimens collected in E. Manchuria and N. Ko
rea
(J) Esakia
(VNP) (5): 39-48
(Y) 1966
(OR) Coleoptera
(FAM) Chrysomelidae
(DST) E. Manchuria;N. Korea
(KEY) systematics;list;include new synonyms, 1 new sp.;local fauna
#

```

Example 1. An example shows a retrieval of the numbers of records on the ESAKIA file using key words of Japan, Korea, China, Taiwan, Thailand, New Guinea and Australia using distributional tugs (DST) and displays the first record.

```

READY
SIGMA
SIGMA>SEARCH
      RECORD DELIMITERS
D01:=#
D02:=
      ITEM DELIMITERS
D02:=!
D03:=
      KEYWORDS
A01:=new species
A02:=n. sp.
A03:=(OR)...Hymenoptera
A04:=(OR)...Coleoptera
A05:=Simandrena
A06:=
      LOGICAL FORMULAS
Z01:=A1
Z02:=A2
Z03:=Z1,Z2
Z04:=A3
Z05:=A4
Z06:=Z3.Z4
Z07:=Z3.Z5
Z08:=A5
Z09:=Z3.Z8
Z10:=
REPORT(Y/N)?Y
FILE=$.ESAKIA
      RETRIEVED TEXTS
QUESTION 01 (201) =      1372      1372
QUESTION 02 (Z02) =       625        625
QUESTION 03 (203) =      1374      1374
QUESTION 04 (204) =      1637      1637
QUESTION 05 (Z05) =      4030      4030
QUESTION 06 (Z06) =       208        208
QUESTION 07 (207) =       819        819
QUESTION 08 (Z08) =        15         15
QUESTION 09 (Z09) =         7          7
TOTAL =      6014      6014
CPU (SEC/1000) =       617        617
FILE=#
DO:REFILE
REPORT(Y/N)?N
QUESTION:=9
NEW RECORD DELIMITER:=#
NUMBERING(N/Y)?N
OUTPUT-FILE:=T
#
(TAX) Andrena (Simandrena) yamato Tadauchi et Hirashima, n. sp.
(AU) TADAUCHI, Osamu;HIRASHIMA, Yoshihiro
(T) New or little known bees of Japan (Hymenoptera, Apoidea) IV. Supplements
to Andrena (Simandrena)
(J) Esakia
(VNP) (20): 82-86
(Y) 1983
(OR) Hymenoptera
(FAM) Andrenidae
(DST) Japan(Hokkaido, Honshu, Shikoku, Kyushu, Sado Is., Tsushima Is., Yakushi
ma IS.)
(KEY) systematics;in key;new species;description(female, male);type(female:Ky
shu Univ., No. 2422);type locality(Kuroubaru, Chikuho, Fukuoka Pref.);floral
records;flight records
#

```

Example 2. An example shows a retrieval of the numbers of records on the **ESAKIA** file using key words of the new species, Hymenoptera, Coleoptera, *Simandrena* and displays the first record of the new species of *Andrena (Simandrena)*.

REFERENCES

- Arikawa, S. *et al.*, 1987, A text database management system SIGMA, Version 2. *Koho, Computer Center Kyushu Univ.*, 20: 517-581 (In Japanese).
- *et al.*, 1988, SIGMA: A text database management system. *RIFIS Techn. Rep., Res. Inst. Fundam. Inform. Sci., Kyushu Univ.*, CS-3: 1-16.
- Committee of libraries and databases, Univeisity Computer Centers, Japan, 1993, A Guide to on-line databases at University Computer Centers in Japan, 13th edition. 2.51 pp. (In Japanese).
- Computer Center, Osaka Univ., 1991, BIOSIS Search Guide, 2nd edition. 121 pp. (In Japanese).
- Higuchi, T., 1987, On Thomas Mann File. *Res. Rep. Computer Sci., Computer Center Kyushu Univ.*, 4: 37-39 (In Japanese).
- Tadauchi, O., 1985, General index to **Mushi** Vol. 1(1928) – Vol. 50 (1985). Subject index. A list from a database DMUSHI produced in the Computer Center of Kyushu University. *Mushi*, 50: 47-175.
- , 1987, A public database of entomology KONCHU and its usage. *Koho, Computer Center Kyushu Univ.*, 20: 597-614 (In Japanese).
- , 1988, Entomology database KONCHU. *Kyushu Univ. Library Bull.*, 24: 28-30 (In Japanese).
- , T. Miyahara, M. Takeda & T. Uchida, 1990, Database of "A Check List of Japanese Insects". *Res. Rep. Computer Sci., Computer Center Kyushu Univ.*, 7: 9-18 (In Japanese).