Employing the Unified Modeling Language (UML) and Relational Modeling Concepts for the Development of a Schema for Fruit Flies

M.W. Mwatawala
Department of Crop Science and Production,
Faculty of Agriculture,
Sokoine University of Agriculture,
Morogoro,
Tanzania
Phone: +255-23-260-3681
mwatawala@giant.suanet.ac.tz

A. Lindow and M. Gogolla
Database systems group,
Department of mathematics and computer science,
University of Bremen,
Germany
Phone: +49-421-218-3495 (Secretariat), Fax: +49-421-218-3054
{lindow | gogolla}@informatik.uni-bremen.de

M. De Meyer
Entomology Section,
Royal Museum for Central Africa,
Tervuren,
Belgium
Phone: +32-2-769-5360
marc.de.meyer@africamuseum.be

ABSTRACT

This paper is about the analysis, conceptual design, and implementation of a fruit fly database containing information about fruit flies of Morogoro Tanzania, their description, host range, infestation levels in different fruits and the proposed management techniques. The analysis and design of the database is described with *Unified Modeling Language, Entity Relationship diagrams, Extended Entity Relationship diagrams* and tabular views of *relational model*. So this paper is an example of a practical and useful application of Information and Communication Technology in solving a pest problem in a developing country, whose economy is largely dependent on agriculture.

Key Words: Relational Databases, Scientific Databases, Application, Requirements/ Specifications, Design, Modeling, Fruit Fly, Pest Management

1. Background: Fruit Flies Problem

Numerous fruit flies (Diptera: Tephritidae) are economically important pests of fruit trees [18]. According to Gupaul et. al.[7] fruit flies rank among the world's most serious pests of Horticultural crops . The female of the fly oviposit eggs underneath the skin of the fruit. The hatched larvae (maggots) spend their life feeding inside the fruit. The fruit will later on rot and fall to the ground. The result is both quantitative and qualitative losses of fruits [8, 18]. Fruit flies have been reported to occur in many parts of Tanzania, including Morogoro region [3, 12, 18]. However the actual number of species occurring in the region and their host range are not known for certainty. Full year survey cycles and collection the fruits for rearing could establish the fruit flies species present and fruits attacked or associated with them. Knowing the fruit fly fauna and their host range will help in designing a proper management system called Integrated Pest Management (IPM). One way of presenting the results from such studies is through the use of Information and Communication Technology (ICT), more specifically, databasing. The applications advantages of using ICT in agricultural entomology, IPM and decision making in pest control have been reviewed by many workers [1, 2, 9, 13]. Potential for utilization of ICT in IPM in Tanzania has been presented by Mwatawala [11]. IPMDBASE is an example of database of IPM in selected crops that was developed in Tanzania [10]. Other examples of ICT applications in entomology and pest management include databases by Thompson [16] and [3]. The aim of this paper is to present the schema for the fruit flies database to be used in Morogoro, Tanzania. Section 2 describes the conceptual design of database including the use case view. Section 3 shows the logical design of the database, first by using the Unified Modeling Language (UML), and then transforming it into Entity Relationship (ER) model. The ER model is then enhanced and mapped into a relational database schema. Finally, the schema is normalised. Section 4 presents the implementation of the database in MS Access. Discussion and conclusion are given in sections 5 and 6 respectively.

2. Conceptual design of the database

2.1 Requirements collection and analysis

The identity of fruit fly species, their host range, agro-ecological zones in which they are found as well as their population numbers are among key pre-requisite for arriving at a correct decision in managing the fruit flies. The relationships among these prerequisites as related to the management of fruit flies can be explained as follows:

Fruits

There are many fruit species that can be attacked by fruit flies e.g. mango, citrus, guava, wild fruits [18]. Within species there are different varieties, which posses additional characters that differentiate them from the other varieties in the major group. The botanical name uniquely identifies a fruit crop. A combination of botanical name and variety name uniquely identifies a fruit variety. A fruit variety can be attacked by many species of fruit flies, in many different locations. While several fruit species can be attacked by a single fruit fly species, in different locations [3, 18]. The database should keep track of species name, variety name, common name, family name, economic status as well as morphological description of stem, inflorescence, leaf and fruit.

Fruit flies

There are many fruit flies attacking different fruits [3, 18]. Distribution and classification has been given by White and Elson-Harris [18]. The species name uniquely identifies a fruit fly. Some fruit fly species can attack a number of different fruit species i.e. They

are polyphagous [18]. The database should keep track of species name of fruit fly, its common name, tribe, sub-tribe and genus as well as morphological description of wing, thorax, abdomen, legs and antenna.

Location
Fruit orchards can be found in different locations, in different agro-ecological zones. All the surveyed locations, in different agro-ecological zones of Morogoro, have been found to be infested by a number of different fruit flies. The database should keep track of location name, district, geographical position and altitude.

Management options
There are many management techniques for managing fruit flies [3, 8, 18]. The management options to be included into the database, are those that have been found to be effective in managing at least one fruit fly species. The options fall under cultural methods, chemical methods, physical methods, biological methods, Integrated Pest Management (IPM). The database should provide information on various management techniques that can be used to control given fruit flies species.

Weather
Different weather parameters can be recorded in different locations and these could have effect on fruit flies fauna as well as population numbers. However, some locations may lack weather records. The database should keep track of weather parameter namely rainfall, temperature, relative humidity and the period of the year when these parameters are recorded.

2.2 Uses of the Database
The developed database should be able to provide the following:

(i) Identity of different fruit flies, fruits they attack, and locations they can be found.
(ii) Identity of different fruit varieties, locations in which they are grown, fruit fly species attacking them and the level of infestations (incidence).
(iii) List and description of different management techniques that can be used to control different fruit fly species.
(iv) Description of different locations in which fruits are grown as well as weather parameters recorded in these locations.

2.3 Use Case View: Actors and Use Cases
In the following sections, we describe the database by the notations, introduced by Rumbaugh et al [14]

The Actors
The actors and their description are presented in table 1. It includes researcher, student and extension agent who is a link between the system and the farmer

The Use Cases
There will be the following use cases: Identify a fruit fly, Get host range of fruit flies, Provide incidence a fruit fly species in a fruit variety, List fruit flies attack a fruit variety, List fruit flies in a location, Give incidence of fruit fly species in a fruit variety and Select management technique (Figure 1). It can be seen that the researcher will be interacting with all functionalities of the system, because of the nature of his job, which also include consultancy. The extension agent, who in conventional way is a bridge between the farmer and the researcher, will need to identify a fruit fly, get its host range, incidence and finally select the management techniques for controlling the fruit fly. This is because in IPM the losses have to be economically justifiable before opting for a particular management technique [6, 15]. On the other hand, farmers will mostly interact with the database through extension agents, while students will mostly interact with three use cases as presented in fig. 1.

The Actors
In this paragraph we want to give very short descriptions of the actors, appearing in the use case diagram in Figure 1.
We distinguish between four kinds of actors**.** The first role, in which a user can interact with the system, is the **Researcher/**

Scientist, who generates knowledge for solving farmers problems, as well as for academic purposes. Another role is the **Extension Agent**, who transfers knowledge (practical information) from the researcher to the farmer i.e. a link. The **Farmer** role characterizes the end user of the practical knowledge (clientele) generated by the researcher, and brought to him by extension agent. The **Student** is recipient of knowledge as part of the academic curriculum.

Schematic representation of the use cases:

Name: **Identify a fruit fly species**
Goal: To identify a fruit fly species
Precondition: Details about the fruit fly species exists
Post-condition: The fruit fly species identified
Actors: Researcher, extension agent, student
Triggering event: Researcher, extension agent or student requests fruit fly identification
Description: A real world researcher, extension agent or student wants to know the identity of a fruit fly species
Alternatives: A real world researcher, extension agent or student wants to confirm the identity of a fruit fly

Name: **Get host range of a fruit fly species**
Goal: To list the fruit varieties attacked by a fruit fly species
Precondition: Host range details are available
Post-condition: Host range of a fruit fly species determined
Actors: Researcher, extension agent and student
Triggering event: Researcher, extension agent or student requests the host range of a fruit fly species
Description: A real world researcher, extension agent or student wants to know the list of fruits attacked by a fruit fly species
Alternatives: A real world researcher, extension agent or student wants to know if a fruit fly species attacks a fruit variety

Name: **Provide incidence of fruit fly species in a fruit variety**
Goal: To get the level of incidence of a fruit fly species in a fruit variety
Precondition: Incidence levels exists
Post-condition: Incidence level provided
Actors: Researcher and extension agent
Triggering event: Researcher or extension agent requests the incidence data of a fruit fly species in a fruit variety
Description: A real world researcher or extension agent wants to know the number of fruit flies emerging from a unit weight of fruits or a unit number of fruits.

Name: **List of fruit flies attacking a fruit variety**
Goal: To get the list of all recorded fruit flies attacking a fruit variety
Precondition: The list of fruit flies attacking a fruit variety exists
Post-condition: The fruit flies attacking a fruit variety listed
Actors: Researcher and extension agent
Triggering event: Researcher and extension agent requests the list of fruit flies attacking a fruit variety
Description: A real world researcher or extension agent wants to know the list of fruits attacked by fruit fly species.
Alternatives: A real world researcher or extension agent wants to know if a fruit variety is fruits attacked by a fruit fly species.

Name: **List of fruit flies in a location**
Goal: To get a list of fruit flies occurring in a location
Precondition: Fruit flies occurring in a location recorded
Post-condition: Fruit flies occurring in a location listed
Actors: Researcher and extension agent
Triggering event: Researcher or extension agent requests the list of fruit flies in a location
Description: A real world researcher or

extension agent wants to know the list of fruit flies found in a location

Alternatives: A real world researcher or extension agent wants to know if a fruit variety is grown in a location

Name: **Select management techniques**

Goal: To select management techniques for controlling the fruit flies

Precondition: The management techniques exists

Post-condition: The management techniques selected

Actors: Researcher and extension agent

Triggering event: Researcher or extension agent requests the selection of management techniques

Description: A real world researcher, extension agent or student wants to know the management options that can be used to control a fruit fly species or a list of fruit fly species

Alternatives: A real world researcher, extension agent or student wants to know if a fruit fly species can be controlled by a management option

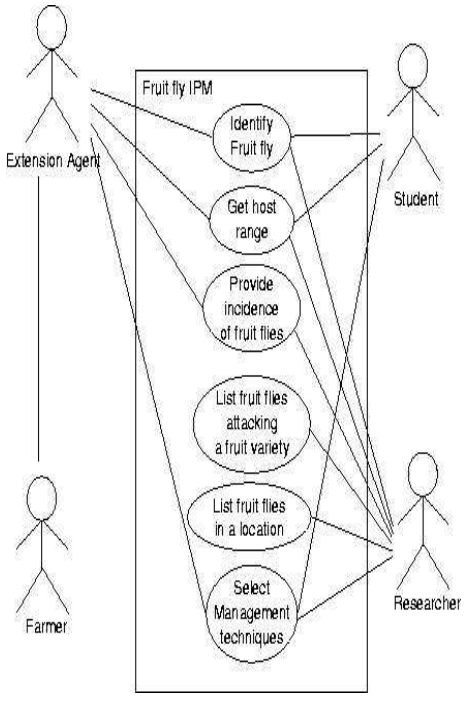

Figure 1. Fruit flies database: Use case view

3. Logical Database Design

3.1 Development of the Schema Using UML Concepts

Static View: Classifiers and their Associations

The database will have the following classes and attributes (attributes belonging to the key are underlined):

The class 'FruitFly' has the attributes sppN (scientific name), ffComN (fruit fly common name), tribe, subTribe, genus, head, thorax, abdomen, wing, antenna, and leg. The class 'Fruit' has the attributes botanicalName, varietyName, frtComName (fruit common name), fruit, leaf, stem, inflorescence, family, and commercialStatus. The class 'Location' has the attributes geoPosition (geographical Position), locationName, districtName, and altitude. The class 'ManagementOption' has the attributes chemical, biological, sterileInsectTech-nique, cultural, trapping, attractants, and legislative. And the class 'WeatherRecord' has the attributes weatherId, month, rainfall, minTemp (minimum temperature), maxTemp (maximum temperature), and r/h (relative humidity).

3.2 Transforming The UML Schema to ER and Subsequently EER Schema

3.2.1 The ER Schema: Entities and their Attributes

The ER schema will have five entity types, which were classes under the UML concepts. The attributes of the entity types, will be the same as those under UML concepts. All the associations in the UML concept becomes relationship types under the ER modeling concept. The multiplicities are adopted as structural constraints of the relationship types, and are indicated on opposite sides to that of the UML static view schema (figure 3)

3.2.2 ER to EER (Enhanced Entity

Relationship Model)

The ER model was enhanced to include the concept of category. This is because the entity type ManagementOption contain attributes that can form entities of their own. These entity types will have different attributes and different primary keys. Thus the entity type ManagementOption was transformed into a category, and the attributes became superclasses (figure 4)

We differentiate between the following seven Management Options and their attributes (again the key attributes are underlined):

The first is the ChemicalMethod with the attributes a.i, tradeName, insecticideForm, and insecticideDose.

The second is the BiologicalMethod with its attributes bioAgentName, modeOfAction, approach, and precaution.

Another Method is the SterileInsectTechnique with its attributes sterilizingAgent, and releaseDetails.

Trapping is charactarized by the following attributes: trapName, manufacturer, and specifi-cation.

The method Attractant has the attributes attractantName, attractantFom, composition, and attractantDose.

There is also the CulturalMethod with properties like cultMethodName and cultMethodDetails.

Last but not least, there are LegislativeMethod with the attributes actNo. and actStatement.

3.2.3 EER to Relational Mapping

(i) Mapping Entity Types

For each strong entity type, in the ER schema, a relation R was created, that included only the simple (atomic) attributes [4, 5]. As showcase we translate the strong entity type FruitFly, the other cases are similar:

FruitFly:

sppN	ffComN	tribe	subTribe	genus	Head	thorax	abdomen	wing	antenna	legs

Relations1. Entity relation created from the strong entity type 'FruitFly'

(ii) Mapping Category ManagementOption and its superclasses

In mapping the category ManagementOption, the surrogate key *mgtId* was created, because the resulting superclasses have different keys. The surrogate key, *mgtId* was included as a foreign key in each of the superclass to specify the correspondence in values between the surrogate key and key of each superclass [5]. We show exemplarily the translation of ManagementOption and Chemical Method:

ManagementOption:

mgtId	IPM

ChemicalMethod:

a.i.	tradeName	insecticideForm	insecticideForm	mgtId

Relations 2. Mapping superclasses of the category ManagementOption

(iii) Mapping the Relations

For each binary many to many relationship R, a new relation S was created. The primary keys of the relations that represent the participating entity types, were included as foreign keys in the new relation [4, 5]. As example we present the relationship 'Attacks'.

Attacks:

sppN	botanicalName	varietyName	Incidence

Relations 3. Relation created from the M:N relationship 'Attacks'

3.2.4 Normalization of the Database and its Final Mapping

After mapping the relations, normalization was used as a measure of appropriateness of the design of the database. A series of normal form tests were carried out to individual relation schema, and the database was normalized to the third normal form. The entity relations in the relational schema have all the properties of a 3NF as outlined

by Elmasri and Navathe [5]. This means all attributes are atomic and there are no nested relations. Also a primary key contain multiple attributes that are functionally dependent on part of the primary key. Lastly, no relation in the relational schema has a non-key attribute functionally determined by another non-key attribute. The relational schema is also in Boyce-Codd Normal Form (BCNF) because every determinant is a superkey [5].

4. Implementation in MS Access

The database schema described so far in the sections above was implemented in MS Access, where the normalized relations were realized as tables and the attributes as datafields, the datatypes of the attributes were all available in MS Access. Primary, Foreign keys, and NOT NULL constraints were used to ensure the referential integrity. Forms and other GUI elements were created to support the users of the database inserting, updating and deleting records in the database.

Actually the Database is used to collect data from the field.

5. Discussion

The schema of the database has been designed to accommodate the data from various experiments, both in the field and in the laboratory. The informal measures of quality relation schema as well as normalization tests were applied [5]. A good part of the database will be populated with information from standard texts. Based on this design, different users can interact with the database to get the information that is useful to them. Different users will be able to get information on identity of different fruit fly species, their host range, their incidence and consequent damage in commercial fruits as well as the various management techniques for controlling the fruit flies. The population dynamics study will provide data for the list of fruit flies occurring in a location including the peak population period for the various fruit flies. The rearing experiment will provide data for the host range and list of fruit flies attacking a particular variety, which will also include the incidence of fruit flies in various fruit flies varieties. Weather information from various locations will be obtained from the Tanzania Meteorological Agency (TMA). The geographical position will be determined using the Geographical positioning System (GPS). Fruit flies morphological details, classification and management techniques will be obtained from standard literature. Once populated, the database will minimize one the problems that has been reported to limit the implementation of IPM i.e. lack and poor access to information on crop pests in Tanzania.

6. Conclusion

The designed fruit fly database schema, once populated, will help in alleviating one the problems that has been reported to limit the implementation of IPM i.e. lack of information. IPM program is knowledge intensive[2]. ICT systems of knowledge dissemination like databases could help very much in improving the access of information [9]. Such systems are also quick and easy to update. The usefulness of the database will depend much on the computer infrastructure in the country, and the computer literacy of the extension staff in the Ministry of Agriculture and Food security. Tanzania's agriculture is a rural phenomenon, and majorities of Tanzanian's live in rural areas. The country's ICT policy commits itself to the provision ICT services in rural areas [17]. At moment there are few rural telecentres, expansion of these into many parts of the country will tremendously improve access of information to the farmers through ICT. Farmers in Tanzania can access the information contained in the database through the extension agent, who will have access computers at the telecenter. The database can be stored in a CD ROM or connected to the internet. However, for the database to be much more useful for farmers in Tanzania, a Swahili version of the database will have to be developed, so that an extension can choose, download and print specific information and deliver it to the farmer.

7. Acknowledgments

Authors would like to acknowledge the support and assistance of the following:
The German Academic Exchange Services (DAAD).
The administration and staff of Database systems group, Informatics and Mathematics department, University of Bremen.

References:

[1] W.I. Bajwa and M. Kogan, *Internet-based IPM informatics and decision support*, www.ipmworld.umn.edu/chapters/Bajwa.htm site visited on 27/6/2003

[2] A. Bartlett, "ICT and IPM", In: FAO Farmers, FAO and Field Schools: Bringing IPM to the grassroots in Asia. The Food and Agriculture Organisation, Rome, Italy, 2002.

[3] CABI. *Crop protection compendium*, Commonwealth Bureau of Agriculture International, Wallingford, UK, A CD ROM, 2004.

[4] T. Connoly and C. Begg, *Database solutions*, Addison-Wesley, Munich, Germany, 2004.

[5] R. Elmasri and B.N. Navathe, *Fundamentals of database systems* (3rd ed), Addsion-Wesley, New York, USA, 2000.

[6] R. Daxl, V. Niels, K. Carlos, L. Rolf and W. Hermann, *Integrated pest management*, TZ Verlags Gesellschaft mbH, Rossdorf, Germany, 1994.

[7] S. Gupaul, N.S. Price, R. Soonoo, J. Stonehouse and R. Stravens, "Technologies of fruit fly monitoring and control in Indian ocean region", In: N.S. Price (ed), Technologies of fruit fly monitoring and control in Indian ocean region, Indian ocean commission, Mauritius, 2000, pp. 3-15.

[8] D. Hill, *Insect pests of the tropics and theirs control*. Longman publishers, London, UK, 1983.

[9] S.E. Miller, "Information age and agricultural entomology", Bulletin of entomological Research 83, 1993, pp. 471-474.

[10] M.W. Mwatawala, *Development of Integrated Pest Management database in selected crops for use in Tanzania*, MSc thesis, Sokoine University of Agriculture, Morogoro, Tanzania, 1998.

[11] M.W. Mwatawala, "Potential for utilisation of ICT in IPM in Tanzania", African Journal of Library, Archives and Information Sciences. In Press, 2005.

[12] M.W. Mwatawala, I.M. White, A.P. Maerere, F.J. Senkondo and M. De Meyer, "A new invasive *Bactrocera* species (Diptera: tephritidae) in Tanzania", African entomology 12(1), 2004, pp. 154-158.

[13] G.A. Norton and J.D. Mumford (Eds), *Decision tools for pest management*, University Press, Cambridge, UK, 1993.

[14] J. Rumbaugh, I. Jacobson and G. Booch, *The Unified Modeling Language Reference Manual*, Addison-Wesley, New York, USA, 1999.

[15] K. Sorby, G. Fleischer and E. Pehu, "Integrated pest management in development: Reviews of trends and implementation strategies", Agriculture and rural development working paper 5, World Bank, Washington DC, USA, 2003.

[16] F.C. Thompson (ed), *Fruit fly expert identification system and systematic information database,* Backhuys Publishers, Leiden, Netherlands, 1999.

[17] URT (2003). *National information and communication technologies policy,* Ministry of Communication and Transport, United Republic of Tanzania (URT) www.tzonline.org/pdf/ictpolicy2003.pdf site visited on 5/ March/ 2005.

[18] I.M. White and M.M. Elson-Harris, *Fruitflies of economic importance: their identification and bionomics*, CAB International, Wallingford, UK, 1992.

[19] www.communityipm.org site visited on 20/ July/ 2004.

Figure 2. Fruit flies database: Static View

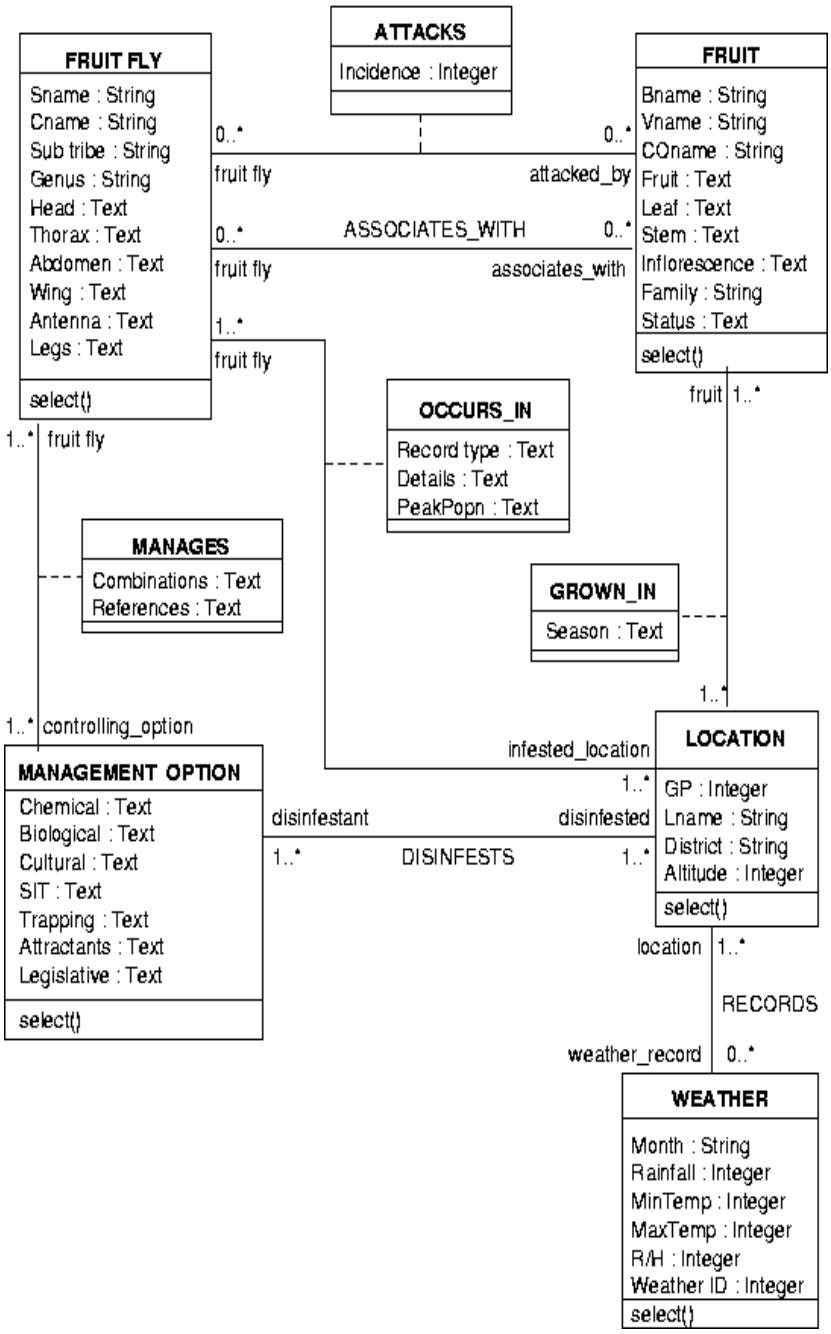

Figure 3. Fruit flies database: ER schema

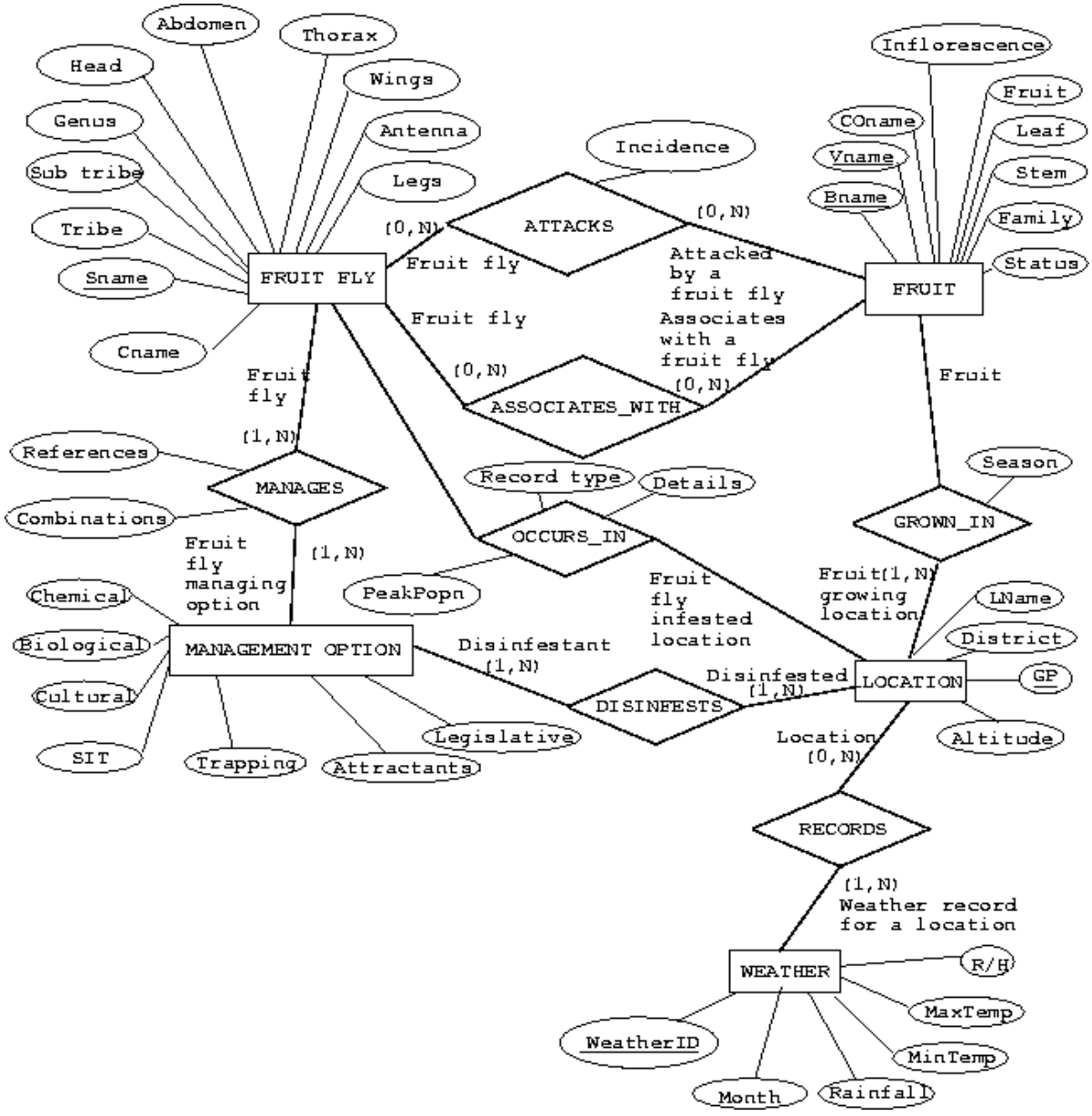
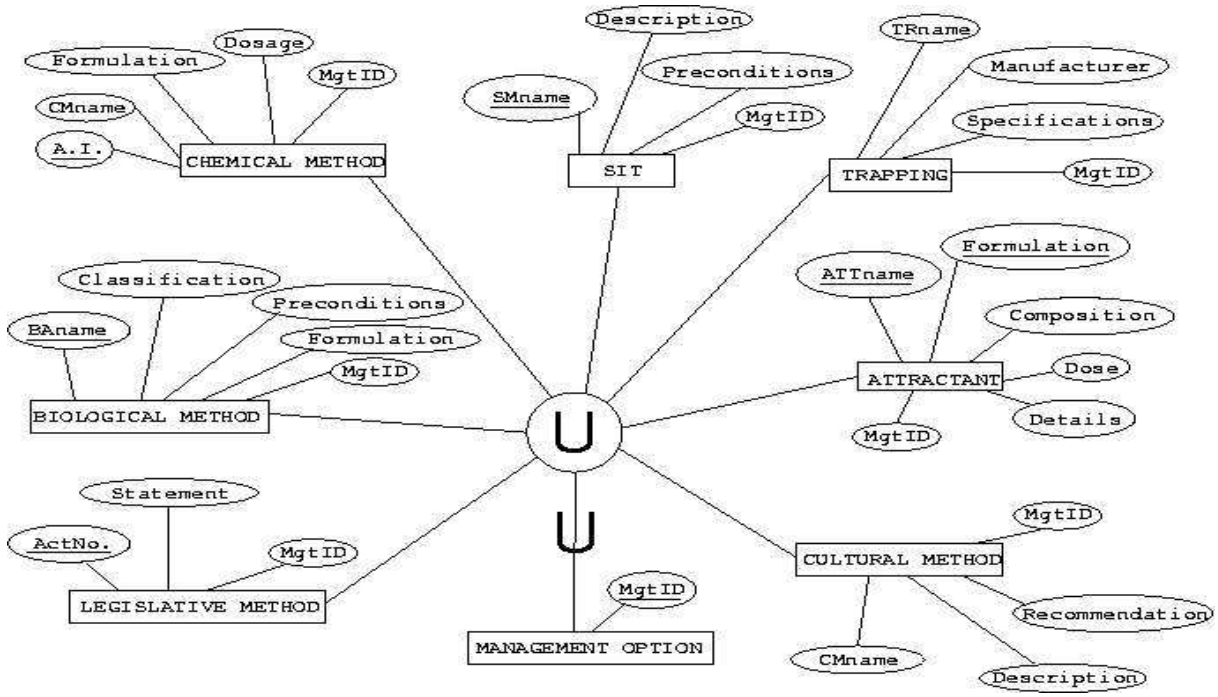
Figure 4. Fruit flies database: Superclasses of the ManagementOption Category