CHAPTER FIVE - KNOWLEDGE BASE MANAGEMENT

5.1 INTRODUCTION

Now that the relationships between stages in the process of knowledge base creation have been outlined, this section discusses iterative evaluation of the knowledge base during its development. The repeated evaluation of the growing knowledge base to drive further elicitation is distinct from the evaluation of the representativeness of the knowledge base considered to be complete in relation to the objectives for its creation (see 6.1) and the correctness and utility of its content. While these processes have tasks in common, their motivations and outputs are different.

Iterative evaluation of the knowledge base occurs throughout the formulation of unitary statements, formal representation and the development of lists of formal terms and hierarchies. The process demands consideration of both individual unitary statements and sets of unitary statements as well as the specification of formal terms and the relationships between formal terms.

So, for example, the statement:

Orange caterpillars can cause sickness in livestock

is incomplete and probably contains implicit meaning. Reference to the knowledge source shows that it can be more completely represented as:

Feeding orange caterpillars to livestock can cause sickness

This still does not really capture the meaning of the source knowledge, which is that:

> Feeding leaves to livestock can cause sickness to those livestock if there are orange caterpillars on the leaves.

Whilst this is now explicit representation of a piece of knowledge, the word ‘can’ along with a lack of any information about the circumstances under which the statement is held to be correct makes it of limited utility. This suggests a need for further knowledge elicitation, possibly revealing that this always occurs, i.e.

> Feeding leaves to livestock causes sickness to those livestock if there are orange caterpillars on the leaves.

This statement can now be meaningfully formalised. Leaves, livestock and orange caterpillars are objects. ‘Feeding’ is an action. Sickness might be viewed as a process or to be a change (decrease) in the value of an attribute (health). This statement might be formalised in many different ways, all capturing essentially the same meaning but with different emphases, for example:

> action(feeding, leaves, livestock) causes att_value(livestock, health, decrease) if att_value(orange_caterpillars, location, on_leaves[1])

or:

> action(feeding, leaves, livestock) causes att_value(livestock, health, decrease) if att_value(caterpillars, location, on_leaves) and att_value(caterpillars, colour, orange)

In other contexts a more precise definition of the results of consuming orange caterpillars may be necessary - sickness may be considered to be a particular type of decline in health that has particular consequences, the difference between the sickness caused by orange caterpillars and other sicknesses might be significant. All these factors will affect the formal

[1] Note that in this case ‘on_leaves’ is seen as a value of the attribute ‘location’

representation of the knowledge and depend on the consideration of this unitary statement with other unitary statements in the knowledge base.

Considering a unitary statement in the context of the other unitary statements in the knowledge base may reveal repetition, contradiction, incomplete sets of unitary statements, or inconsistent use of the terms. In this example, there are statements that:

> *Green caterpillars cause sickness in livestock*

> *Black caterpillars cause sickness in livestock*

There is also information in the object hierarchies that there are three types of caterpillars; orange, black and green. If it is assumed that this is the finite set of caterpillars in this context then the three can be replaced with the single statement that:

> Feeding leaves to livestock causes sickness if there are caterpillars on the leaves.

If it is suspected or found that there are other types of caterpillars that may not, or do not, cause sickness if fed to livestock, the three statements may be replaced by the single statement:

> Feeding leaves to livestock causes sickness if (there are caterpillar on the leaves) and (those caterpillars are green, or those caterpillars are black or those caterpillars are orange).

Alternatively, green, orange and black caterpillars might be classified as being poisonous caterpillars such that:

> Feeding leaves to livestock causes sickness if there are poisonous caterpillars on the leaves.

Considering this unitary statement in relation to other statements may also demand consideration of linkages - for example are there statements, which describe the consequences of a decline in the health of livestock? Apparently contradictory statements may be identified and resolved.

For example, does the statement that:

> Leaves of fodder trees that are attacked by caterpillars do not cause sickness in livestock

mean that caterpillars do not cause sickness or does it mean that the leaves of those trees that are attacked by caterpillars do not cause sickness, provided that there are no caterpillars on them?

At this juncture, consistency of use of terms can be ensured. The term ‘livestock’ is, in the current context, defined in the object hierarchy as cows, goats and buffalo. Do all the statements that use the term ‘livestock’ refer to all three of these, or should some be replaced, for example, with ‘cattle’ or ‘milking livestock’ (i.e., cows, goats or buffalo that are currently lactating)? Precise use must be considered in terms of precise definition. The term ‘feeding’ is used in this statement. It may be that this term is defined as meaning the action of providing fodder for livestock that are stalled. This might be distinguished from the process of ‘browsing’ where livestock actively seek out and select fodder on trees. This process of selective browsing may mean that fodder effects relating to stall-fed livestock do not occur for free-range livestock.

This example illustrates something of the range of activities that can be involved in iterative evaluation and improvement of the knowledge base. The following sections anatomise some of the key processes occurring in iterative evaluation.

5.2 EVALUATING INDIVIDUAL UNITARY STATEMENTS

Individual unitary statements may be evaluated in terms of validity of representation, relevance, utility and ambiguity.

5.2.1 VALIDITY OF REPRESENTATION

Some evaluation of the validity of representation (to check whether the knowledge statement has been correctly encoded) is built into the process of creating formal statements directly or through diagramming. This is due to the restricted structure of the formal grammar, the use of a parser and the stylised English generation in formal representation by AKT.

It cannot be overemphasised how important it is that when formulating unitary statements prior to creating a formal statement, these unitary statements should be rigorously checked to ensure that they conform to the definition of a genuine unitary statement.

For example, the statement:

Nutrient uptake by fodder trees planted on crop land causes a decrease in soil fertility.

Can be broken down into two statements:

Planting fodder trees causes an increase in soil nutrient uptake

An increase in soil nutrient uptake causes a decrease in soil fertility.

Statements which do conform to the fundamental definition of a unitary statement may often be improved through use of condition. For example:

Thorny leaves are unpalatable.

may be stated as:

Leaves are unpalatable if they are thorny.

This is important because of its subsequent impact on formal representation. Formal representation of the former would demand that 'thorny leaves' be treated as an object. This is less flexible than a formal statement in which leaves are the object and their palatability and thorniness are two attributes of that object.

5.2.2 RELEVANCE AND UTILITY

Evaluation of articulated knowledge against objectives provides an important means of maintaining the quality (fitness for purpose) of the knowledge base. Effective filtering of articulated knowledge irrelevant to the knowledge base during knowledge elicitation and formal representation greatly reduces the need for subsequent rationalisation. Unitary statements may be relevant, in as far as they are concerned with the domain under consideration, but irrelevant in that they are not useful (in the context of the knowledge base as a whole). So, for example, the statements:

Manure production influences crop production

and

Some species of tree fodder have a beneficial effect on milk production.

Might be deleted from a knowledge base in which both relationships were expanded in much greater detail in other parts of the knowledge base.

5.2.3 AMBIGUITY

Unambiguous (precise and complete) articulation and representation of a unitary statement is required if a useful knowledge base is to be created. Most ambiguity encountered in a knowledge base is resolvable, being an artefact of articulation and representation. However, knowledge may be inherently ambiguous, so some intrinsic ambiguity will remain in any knowledge base.

The level of resolvable ambiguity in the knowledge base is influenced by:

- the nature of the knowledge source;
- the nature of the knowledge elicitation process;
- the type of interface used for entering the knowledge (i.e. the statement dialog or the diagram dialog);
- the experience of the knowledge base developer; and
- the clarity of objectives for creating the knowledge base.

Resolvable ambiguity is most frequently exposed during the processes of formal representation and the creation of the glossaries and hierarchies. Both processes demand a precise use of terms. Where ambiguity is the result of inadequate unitary statements, it may be resolved through reference to interview material. However, where ambiguity is a result of incomplete or muddled articulation by the knowledge source, further knowledge elicitation is demanded.

Unambiguous representation depends on:

- **complete specification of the meaning** of a unitary statement;
- complete representation of the **context of application** of the unitary statement; and
- **precise use of terms** within that unitary statement.

5.2.3.a Complete specification of meaning

Unitary statements may frequently contain an implied meaning. Even where it can be assumed that this is understood by potential users of the knowledge base, it may significantly constrain automated reasoning.

For instance, while it might be valid to assume that users understand that the statement, in a knowledge base about tree fodder,

> *Celtis australis leaves can cause sickness*

means that feeding the leaves to livestock may cause sickness in livestock, automated reasoning will not pick up this implication. Even simple reasoning such as answering the question:

> *What is the consequence of feeding Celtis australis leaves to goats?*

cannot, therefore, be automated. As a result, formal representation must state explicitly any implied meaning, in this case maybe as:

> *action(feeding, part('Celtis australis', leaves), livestock) causes1way att_value(livestock, health, decrease)*

5.2.3.b Context of application

Knowledge is inescapably contextual at some level, even though the AKT approach to knowledge base creation is based on the disaggregation of sets of unitary statements. Recording conditional information associated with a unitary statement provides important contextual information about the application of that unitary statement. However, exhaustive recording of the context of application is impractical: some level of understanding from common knowledge for the evaluation of the knowledge base as a whole must be assumed. Judgement of the completeness of conditional information is subjective and dependent upon context.

It is important that where a statement is considered to be unconditional this is explicitly recorded.

5.2.3.c Precise use of terms

Another source of ambiguity is the imprecise use of terminology. The definition of terms and the consistent use of terms are discussed later (see 5.3.4). Here, the concern is with the imprecise use of terms within a statement. A widespread example in existing knowledge bases is the use of the term ‘shade’. A unitary statement often refers to one aspect of shade (for example a change in light levels, light quality or temperature) rather than a composite of all shade effects.

5.2.3.d Intrinsic ambiguity

In a fundamental sense, virtually all knowledge is ambiguous at some level. In the current context, however, intrinsic ambiguity is of interest where it has an impact on the practical utility of knowledge.

Ambiguity is intrinsic where the source of an ambiguous statement cannot resolve the ambiguity in that statement. This is generally because the ambiguity is not clear to the informant, frequently because the issue in question is at the margins of his or her comprehension. Intrinsic ambiguity often occurs in the meaning of terms within a statement. Clearly, distinguishing intrinsic ambiguity from the inability of the knowledge base developer to understand the concepts under consideration is a matter of judgement.

5.3 EVALUATING SETS OF UNITARY STATEMENTS

The content of a knowledge base is greater than the sum of the content of the individual unitary statements. As a result, iterative evaluation must include the consideration of sets of unitary statements as well as individual statements. Sets of unitary statements should be evaluated in terms of:

- repetition
- contradiction
- completeness, and
- consistency in use of terms.

5.3.1 REPETITION

A compact knowledge base is significantly more tractable, and, therefore, more useful, than a less compact version. Compaction is achieved by identifying and removing repetition in the knowledge base. Two types of repetition can be identified, strict repetition and deducible repetition. Strict repetition is where a piece of information is stated more than once. Deducible repetition is where a statement in the knowledge base can be deduced from other statements in the same knowledge base and is therefore superfluous.

5.3.1.a Strict repetition

AKT will not allow exactly the same statements from various sources to be added to the knowledge base. The user can add extra sources for a single statement if it is repeated by different informants. While some repetition may be self evident, the identification of repetition will best be achieved by reference to the objectives of the use of the knowledge base. So, two different unitary statements using different terms but capturing very similar information may or may not be defined as repetition depending on the potential impact of the different formulations on the use of the knowledge base.

Selection facilities to identify all statements concerned with a particular formal term or combination of formal terms reveal much repetition. Comparing sets of formal statements also reveals repetition, as formal representation often causes convergence of apparently distinct statements. Such comparisons are most effective when care has been taken to minimise the use of equivalent terms. Common causes of repetition are inconsistent use of terms, or spelling mistakes. For example, the following three statements would all be accepted by the knowledge base, but would all be identical, were it not for inconsistency and carelessness:

Trampling by goats causes soil creep
Trampling by goat causes soil creep
Trampling by gaots causes soil creep

It is particularly important to decide early on whether objects should be entered into the knowledge base in the singular or plural - in the above example, whether statements should refer to 'goat' or 'goats'.

5.3.1.b Deducible repetition and the use of hierarchies in compacting the knowledge base

The identification of hierarchical relationships between objects provides a means of

- capturing the hierarchical nature of knowledge;
- enabling the inheritance of properties by objects up and down the hierarchy, which facilitate the development of a compact knowledge base without loss of information; and
- facilitating hierarchically structured exploration of the knowledge base.

5.3.1.c The use of hierarchies in compacting the knowledge base

Hierarchies provide a means of compacting the knowledge base. They allow knowledge to be recorded at its most general level of application, yet be used to consider more specific instances. This is achieved by considering hierarchical relationship between terms. If, for example, 'wheat', 'barley', 'maize' and 'fava beans' are identified as being annual crops within a hierarchy, the information that annual crops only live for one year is best recorded as generic information about annual crops, rather than for each type of annual crop individually[2]. This influences the size and tractability of the knowledge base. Compare Tables 5.1 and 5.2. Each captures the same information; in Table 5.1 the information is explicitly stated, in Table 5.2 it is more implicitly captured.

[2] This is known as 'property inheritance' and is discussed more fully in Chapter 1.7.

Table 5.1 A complete set of 43 unitary statements

Crops are economically useful	Legumes photosynthesise
Legumes are economically useful	Chick peas photosynthesise
Root crops are economically useful	Pigeon peas photosynthesise
Cereals are economically useful	Cowpeas photosynthesise
Chick peas are economically useful	Legumes have roots
Pigeon peas are economically useful	Chick peas have roots
Cow peas are economically useful	Pigeon peas have roots
Crops are deliberately cultivated	Cowpeas have roots
Legumes are deliberately cultivated	Legumes have leaves
Cereals are deliberately cultivated	Chick peas have leaves
Root crops are deliberately cultivated	Pigeon peas have leaves
Chick peas are deliberately cultivated	Cowpeas have leaves
Pigeon peas are deliberately cultivated	Legumes transpire
Cowpeas are deliberately cultivated	Chick peas transpire
Crops are plants	Pigeon peas transpire
Legumes are plants	Cowpeas transpire
Cereals are plants	Legumes are crops
Root crops are plants	Roots crops are crops
Chick peas are plants	Cereals are crops
Pigeon peas are plants	Chick peas are legumes
Cowpeas are plants	Pigeon peas are legumes
	Cowpeas are legumes

Table 5.2 A compacted statement of knowledge based on a hierarchical structuring of the objects, reducing the number of statements to 13

1 Crops are economically useful	8 Legumes are crops
2 Crops are deliberately cultivated	9 Root crops are crops
3 Crops are plants	10 Cereals are crops
4 Legumes photosynthesise	11 Chickpeas are legumes
5 Legumes have roots	12 Pigeon peas are legumes
6 Legumes have leaves	13 Cowpeas are legumes
7 Legumes transpire	

A mechanism exists for capturing the hierarchical nature of taxonomic statements, therefore all the statements not explicitly stated can be deduced by applying the general rules to lower orders of the hierarchy.

Suppose for example, that in a particular knowledge base:

- four crop species are classified as annual crops;
- these four are the only annual crops represented in the knowledge base; and
- all are recorded as only living one year.

Ensuring that knowledge is recorded at the highest level can greatly reduce the size of the knowledge base.

Deducible repetition can also occur where the implications of linked sets of unitary statements are explicitly stated in the knowledge base. Where a statement contains the assertions that:

a) Fertiliser application increases soil fertility if........

b) An increase in soil fertility causes an increase in crop yield if ....

The statement that:

c) Fertiliser application increases crop yield.

is a deducible repetition.

Similarly it can be logically deduced from b) that:

A decrease in soil fertility causes a decrease in crop yield if....

Systematic identification and removal of repetition in a knowledge base may often halve the number of unitary statements in the knowledge base.

5.3.2 CONTRADICTION

Contradictory unitary statements may be of two types - conflicting and inconsistent. Conflicting knowledge is knowledge from two different sources which is contradictory. Inconsistent knowledge is knowledge from a single source which is contradictory. Until contradiction is resolved, either by rejecting one unitary statement in favour of another or by demonstrating that apparent contradiction does not represent actual contradiction, the two contradictory unitary statements can be viewed as being competitive.

Where contradictory knowledge in the knowledge base is identified, it may be resolved through further knowledge elicitation or flagged within the knowledge base. Contradictory unitary statements may become apparent at any stage during the creation of the knowledge base. Equally, contradictions may only become apparent once the implications of the knowledge have been illuminated through reasoning with that knowledge.

Once contradiction has been identified it may be resolved by one of two means.

- Apparent contradiction may be resolved through clarification of the meaning of contradictory statements: in particular, two apparently contradictory statements may be distinguished by specification of the conditions under which they are held to be true.

- Contradictions may be resolved by demonstrating that there is a significant difference between the reliability of two statements such that the less reliable statement can safely be rejected in favour of the more reliable. (However, it is important to remember that judging 'reliability' of statements subjectively by the knowledge base developer is difficult. It is best to opt for the rule of thumb: consider true unless proven wrong.)

Box 1. An example of apparently inconsistent unitary statements

In an investigation into forest gardens in the Kandy district of Sri Lanka (Southern, 1994) the following statements were given by the same informant on different days:

Vegetable diseases are reduced in the shade (Abeysinghe, 21.4.92)
Disease problems are not influenced by the degree of shade
(Abeysinghe, 26.4.92)

However, after returning to the source to clarify this apparent contradiction, it became clear that the second statement related to a conversation about a part of the garden where vegetables were not grown. Thus by appending the conditions in which the statements were relevant, they were no longer contradictory.

Inconsistent knowledge should, in principle, be resolved by one of the above mechanisms. Conflicting statements may frequently not be resolved: under such circumstances the two unitary statements may be flagged by attaching a Memo to both statements, stating that these are alternative views. Automated reasoning tools are useful for identifying some types of inconsistency, for example the ‘inconsistent att_value statements’ and the ‘inconsistent causal statements’ tools provided in the macros.

5.3.3 COMPLETENESS

The incompleteness of a set of unitary statements (as opposed to an individual statement: see 5.2.3) is assessed through the identification of apparent gaps in the knowledge in relation to objectives. Gaps may occur in a knowledge base because the knowledge needed to fill those gaps is unknown or because the relevant knowledge has not been elicited. Therefore, the identification of gaps demands further knowledge elicitation and, if the necessary knowledge is found to be available, addition to the knowledge base. Completeness can only be defined in relation to objectives in the creation of the knowledge base. Even then, iterative evaluation of completeness tends to be subjective.

Generating diagrams provides a powerful means of facilitating the identification of gaps in sets of knowledge. For example, a diagram constructed as a result of one of a series of interviews with farmers in Nepal provided a set of questions for further interviews (such as 'what are the stem strength properties of the different crop species grown and the different varieties of each species?'; 'what influence does crop head size have on crop yield?'; 'What are the consequences of an increase in straw height?'; and 'does an increase in shade always result in an increase in pest incidence?'). This helped to identify topics for further discussion with the informant.

5.3.4 CONSISTENCY AND PRECISION IN THE USE OF TERMS

Consistent and precise use of terminology is important if the knowledge base developer is to make effective use of the knowledge base. Comparison of the use of the same term in different statements ensures the consistency of meaning. Comparison of apparently similar terms identifies overlaps in terminology.

Terminology in knowledge systems about agroforestry is rarely consistent or precise. A demand for an exacting consistency when creating knowledge bases may lead to a proliferation of terms. There is then a danger of increasingly unrepresentative use of those terms by source communities. Here again the strategy to be taken depends on objectives in the creation of the knowledge base. They may be primarily to study the current state of knowledge of a target community, or to develop a knowledge base for use in providing decision support.

5.4 EVALUATING HIERARCHICAL STRUCTURES

As well as evaluating and modifying unitary statements, iterative evaluation of the knowledge base demands evaluation of the formal terms specified, the relationships between those formal terms and their definitions.

5.4.1 FORMAL TERMS

Identifying and removing repetitious terminology can significantly increase the utility of the knowledge base. The lists of formal terms available in AKT encourage consistent use of terminology, particularly through formal representation. Nevertheless, regular comparison of the various formal terms may reveal overlapping terminology. The ability to identify equivalent terms for the same object (i.e. synonyms), provides a useful means of allowing a natural articulation of knowledge, while still identifying equivalent meaning. Synonyms can also be applied to processes, attributes, actions, values and links.

5.4.2 RELATIONSHIPS BETWEEN OBJECTS

The development of object hierarchies is one of the most demanding tasks in the creation of the knowledge base (see 4.3). Regular evaluation of object hierarchies, including testing those hierarchies on source communities has proved to be valuable.

Equally, assessment of the set of attributes that are related to a particular object or process, and the set of values that are linked to an attribute, is important in ensuring consistent terminology.

In one knowledge base for example, two separate objects, 'leaves' and 'manure' have an attribute 'texture'. However, the set of values that this attribute can take differs for the two objects ('soft' and 'hard' for leaves, 'loose' and 'firm' for manure). As a result, one or both of the attributes must be renamed to avoid this apparent repetition. By contrast, both 'leaves' and 'manure' also have the attribute 'water content', but the values that this attribute can take ('high' or 'low') are the same for both.

5.4.3. DEFINITION OF FORMAL TERMS

The definition of a formal term may be improved iteratively during the creation of a knowledge base. After each refinement of the definition of a term, all existing uses of that term in unitary statements must be checked to ensure a valid use of terminology.

Key points of Chapter 5:

- Formalised unitary statements must be unambiguous and a valid representation of the original statement.
- Iterative evaluation of sets of statements for avoiding repetition and contradiction and for ensuring completeness and consistent use of terms leads to a concise knowledge base which optimises its utility. Hierarchical arrangement of objects in the knowledge base is a powerful feature which enables valid representation of object classification and this helps compact a knowledge base significantly.