Chapter 8

8 Evaluating the System

The goal of this thesis is the development of a KBS approach that supports the reuse of knowledge for a wide range of activities. RDR was chosen as the starting point of the investigation due to its various strengths including its demonstrated ability to support the expert in KA, maintenance and inferencing. While a discussion of these activities was offered in Chapter Seven it is not necessary to offer an evaluation in this thesis as this has been provided by other researchers. For a current evaluation of the MCRDR technique for these type of activities the reader is referred to Kang (1995) and Kang et al (1998).

The aim of this chapter is to provide some evaluation of the reflective activities. Chapter Seven included discussion of each activity and how MCRDR/FCA measures up to the criteria identified from the literature. This chapter is concerned with a different type of evaluation. Many of the reflective activities rely on the line diagram and this chapter concentrates on an evaluation of the FCA line diagram from a number of perspectives. The first one is generality. In Chapter Six it was claimed that a T-Box can be generated from any A-Box that can be converted into a decision table. As examples of this, in Section 8.1.1 two CLIPS rule-bases are converted into decision tables (cross tables) and line diagrams produced to describe these domains. Section 8.1.2 considers a different aspect of generality by evaluating whether the line diagrams drawn with FCA can be seen as a generally acceptable knowledge representation method. To achieve this the FCA diagrams are compared to a number of other graphic rule representations. The reader will of course decide which representations they prefer, and why, but some statistics from a survey, which is described in section 8.2.1.4 are given on how well each of these representations scored using a set of four criteria: pleasing appearance, comprehensibility, good structure and ability to determine implications.

In Section 8.2 a second set of tests has been performed to evaluate how well the diagrams can be used for instruction or tutoring. The first evaluation in Section 8.2.1 uses a survey to determine how quickly an individual can learn to read a line diagram, how the line diagram representation compares to using a rule trace of the same information and whether the line diagram can be used to learn something about the domain and/or knowledge base. The survey is reinforced in Section 8.2.2 by using domain experts to comment on what I, as a novice, have been able to teach myself about each domain from the FCA models based on the MCRDR rules. First we look at the issue of generality.

8. 1 The Generality of the MCRDR/FCA Modelling Tool

Two claims are evaluated in this section. The first is that Formal Concept Analysis can be used on any knowledge representation that can be mapped into a decision table and is considered in section 8.1.1. The second claim is that the modelling representation itself is comparable to other visual representations of the same knowledge and is discussed further is section 8.1.2.

8.1.1 The Generality of Using FCA to Model different Knowledge Representations.

This thesis has claimed that FCA can be used to display concepts in rule-based systems or any representation that can be converted into a decision table. Colomb (1993) has proven formally that any propositional KBS or decision tree can be mapped into a decision table. Once we have the knowledge as a binary decision table or crosstable, the mapping from a decision table to a formal context is a direct translation. In this section the viability of applying FCA to KBS representations other than RDR KBS is shown by using two KBS built using a commonly used and available knowledge representation, known as CLIPS. Two of the sample KBS that are supplied with CLIPS version 5.1 have been selected. The first KBS is an 84-rule KBS that classifies animals and the other is a 61-rule KBS that assists a user in selecting an appropriate wine.

8.1.1.1 The CLIPS Animal KBS.

The CLIPS Animal KBS is easily converted into a decision table. Such a conversion requires the expansion or *inflation* (Colomb 1993) of each rule into its primitive conditions. Conversion of the CLIPS KBS into decision table format requires all abstractions to be removed from the original KBS. Figure 8.1 shows each rule as a set of primitive concepts which is used as input into MCRDR/FCA.

Many interesting views of the animal classification system were found using this data. In Figure 8.2 the condition HOOVES=YES was selected. We can see that any animals that have hooves are also vertebrates (BACKBONE=YES), mammals (WARM.BLOODED =YES) & (HAS.BREASTS=YES) and herbivores (CAN.EAT.MEAT=NO). We can also see that the number of toes is the next consideration, which is then broken up into various categories such as if they have horns and how many, if they have plating, if they have fleece, if they live in the desert and whether they are domesticated. It is possible to use the diagrams to see what are the relevant features in classifying animals within an area of interest.

```
1 %BIRDP(backbone=yes);(warm.blooded=yes);(has.breasts=no);
2 %FLATW(backbone=no);(live.prime.in.soil=yes);(flat.bodied=yes);
3 %WORML(backbone=no);(live.prime.in.soil=yes);(flat.bodied=no);
4 %FISH0(backbone=yes);(warm.blooded=no);(always.in.water=yes);(boney=yes);
5 %SHARK(backbone=yes);(warm.blooded=no);(always.in.water=yes);(boney=no);
6 %CMINS(backbone=no);(live.prime.in.soil=no);(body.in.segments=yes);
   (shell=no);
7 %BAT00(backbone=yes);(warm.blooded=yes);(has.breasts=yes);(can.eat.meat=
   yes);(fly=yes);
8 %TURTL(backbone=yes);(warm.blooded=no);(always.in.water=no);(scally=yes);
   (rounded.shell=yes);
9 %FROG0(backbone=yes);(warm.blooded=no);(always.in.water=no);(scally=no);
   (jump=yes);
10 %SALAM(backbone=yes);(warm.blooded=no);(always.in.water=no);(scally=no);
   (jump=no);
11 %LOBST(backbone=no);(live.prime.in.soil=no);(body.in.segments=yes);
   (shell=yes);(tail=yes);
12 %CRAB0(backbone=no);(live.prime.in.soil=no);(body.in.segments=yes);
   (shell=yes);(tail=no);
13 %JELLY(backbone=no);(live.prime.in.soil=no);(body.in.segments=no);
   (digest.cells=yes);(stationary=no);
14 %PROTO(backbone=no);(live.prime.in.soil=no);(body.in.segments=no);
   (digest.cells=no);(multicelled=no);
15 %CRCAL(backbone=yes);(warm.blooded=no);(always.in.water=no);
   (scally=yes);(rounded,shell=no);(limbs=yes);
16 %SNAKE(backbone=yes);(warm.blooded=no);(always.in.water=no);(scally=yes);
   (rounded,shell=no);(limbs=no);
17 %SANEM(backbone=no);(live.prime.in.soil=no);(body.in.segments=no);
   (digest.cells=yes);(stationary=yes);(spikes=yes);
18 %CORAL(backbone=no);(live.prime.in.soil=no);(body.in.segments=no);
   (digest.cells=yes);(stationary=yes);(spikes=no);
19 %SNAIL(backbone=no);(live.prime.in.soil=no);(body.in.segments=no);
   (digest.cells=no);(multicelled=yes);(spiral.shell=yes);
20 %MONKY(backbone=yes);(warm.blooded=yes);(has.breasts=yes);
   (can.eat.meat=yes);(fly=no);(opposing.thumb=yes);(prehensile.tail=yes);
21 %RHINO(backbone=yes);(warm.blooded=yes);(has.breasts=yes);
   (can.eat.meat=no);(hooves=yes);(two.toes=no);(plating=yes);
22
          %HORZE(backbone=yes);(warm.blooded=yes);(has.breasts=yes);(can.eat.
   meat=no); (hooves=yes);(two.toes=no);(plating=no);
23 %WHALE(backbone=yes);(warm.blooded=yes);(has.breasts=yes);
   (can.eat.meat=no);(hooves=no);(lives.in.water=yes);(hunted=yes);
24 %DOLPH(backbone=yes);(warm.blooded=yes);(has.breasts=yes);
   (can.eat.meat=no);(hooves=no);(lives.in.water=yes);(hunted=no);
25 %CLAMO(backbone=no);(live.prime.in.soil=no);(body.in.segments=no);
   (digest.cells=no);(multicelled=yes);(spiral.shell=no);(bivalve=yes);
26 %SQOCT(backbone=no);(live.prime.in.soil=no);(body.in.segments=no);
   (digest.cells=no);(multicelled=yes);(spiral.shell=no);(bivalve=no);
27 %MAN00(backbone=yes);(warm.blooded=yes);(has.breasts=yes);
   (can.eat.meat=yes);(fly=no);(opposing.thumb=yes);(prehensile.tail=no);
   (nearly.hairless=yes);
```

Figure 8.1 Part of the CLIPS Animal KBS expanded into its primitive form

One more example is a search for animals closest to man. There were two ways that this could be performed. One was to select the Use Related Conclusion option on the Selection screen, see Section 5.2.1.4, for the MAN conclusion code %MAN00. This, however, resulted in the inclusion of any rules which had any of the same rule conditions as the man rule. This meant that all vertebrates were chosen and resulted in too many rules being selected. To overcome this problem the selection was restricted to the aspects of man that were of primary interest. This meant cutting and pasting the parts of the rule that were seen as most relevant to man. The condition: "(backbone=yes) ;(warm.blooded=yes); (has.breasts=yes); (can.eat.meat=yes); (fly=no); (opposing.thumb=yes); (prehensile.tail" were selected and the Generate from Clause option on selection screen 2, see Section 5.2.1.6, was used. A nice extension to the MCRDR/FCA tool for this purpose would be to allow the user to click on a concept and have the system display those concepts that were the closest.

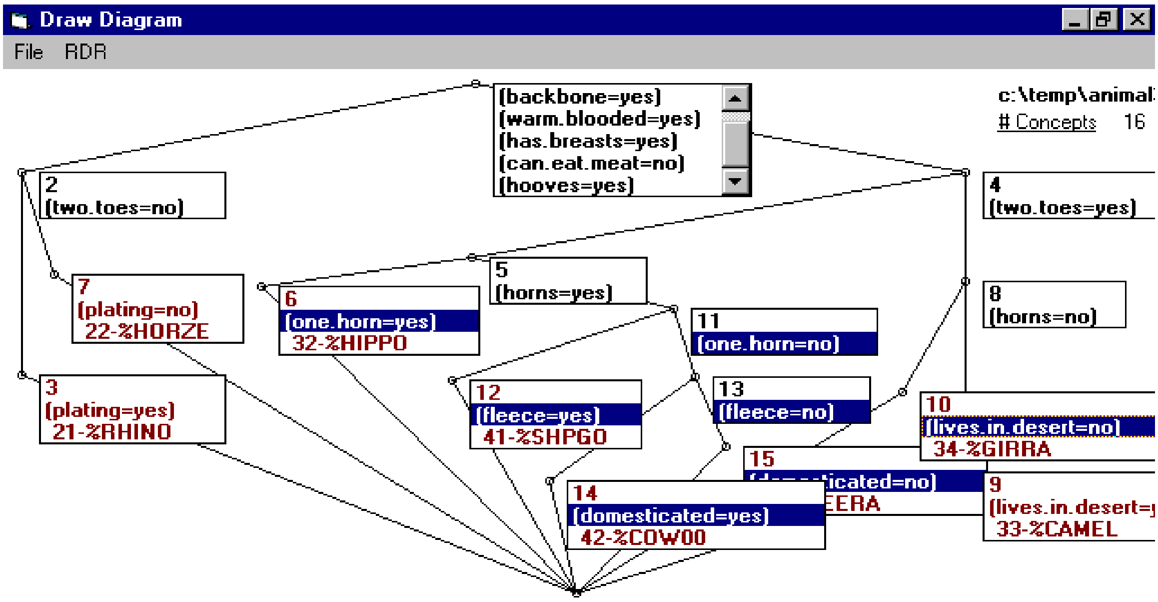


Figure 8.2The MCRDR/FCA diagram for the selection HOOVES=YES in the CLIPS Animal KBS.

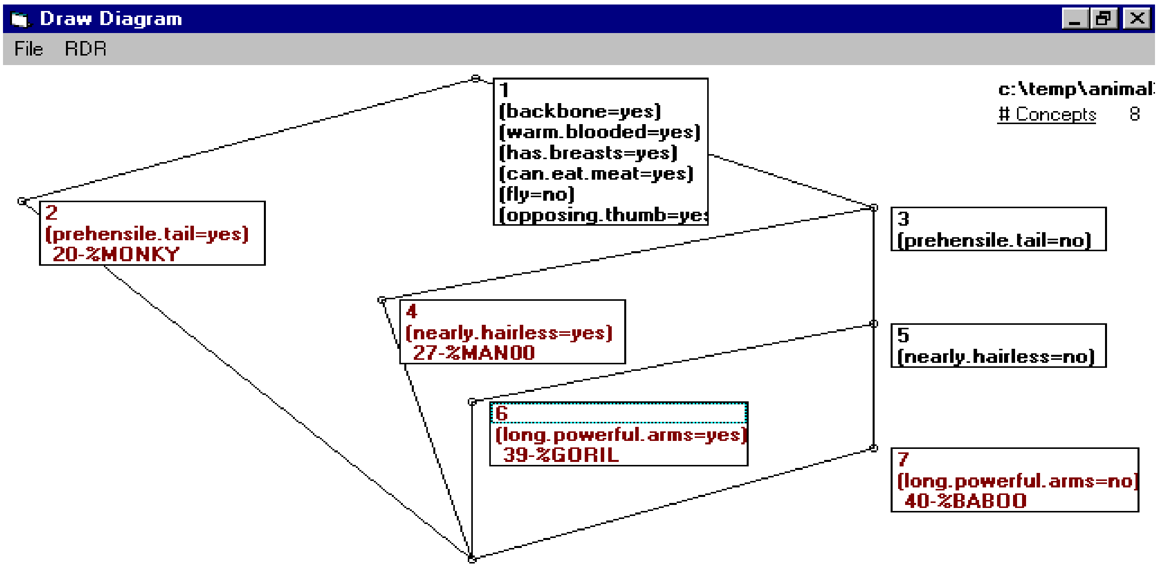


Figure 8.3 Finding the animals closest to man using MCRDR/FCA on the CLIPS Animal KBS.

As would be expected, in Figure 8.3 we can see in the top concept that all animals included have the selection criteria. The monkey is separated the most from man, then the gorilla and baboon are most similar with the difference being that the gorilla and baboon have body hair and man is nearly hairless. The gorilla and baboon are further differentiated by having long or not long, respectively, arms. If man was classified using the long.powerful.arms attribute it would appear closer to either the gorilla or baboon depending on the attribute value.

```
1      %GAMAY (recommended-color=red);(recommended-body=medium);(recommended-
sweetness=medium);
2      %GAMAY (recommended-color=red);(recommended-body=medium);(recommended-
sweetness=sweet);
3      %CHABL (recommended-color=white);(recommended-body=light);(recommended-
sweetness=dry);
4      %SBLAN (recommended-color=white);(recommended-body=medium);(recommended-
sweetness=dry);
5      %CHARD (recommended-color=white);(recommended-body=medium);(recommended-
sweetness=medium);
6      %CHARD (recommended-color=white);(recommended-body=medium);(recommended-
sweetness=dry);
7      %CHARD (recommended-color=white);(recommended-body=full);(recommended-
sweetness=medium);
8      %CHARD (recommended-color=white);(recommended-body=full);(recommended-
sweetness=dry);
9      %SOAVE (recommended-color=white);(recommended-body=light);(recommended-
sweetness=medium);
10     %SOAVE (recommended-color=white);(recommended-body=light);(recommended-
sweetness=dry);
11     %RIESL (recommended-color=white);(recommended-body=light);(recommended-
sweetness=medium);
12     %RIESL (recommended-color=white);(recommended-body=light);(recommended-
sweetness=sweet);
13     %RIESL (recommended-color=white);(recommended-body=medium);(recommended-
sweetness=medium);
14     %RIESL (recommended-color=white);(recommended-body=medium);(recommended-
sweetness=sweet);
15     %GTRAM (recommended-color=white);(recommended-body=full);(sauce=spicy);
16     %CBLAN (recommended-color=white);(recommended-body=light);(recommended-
sweetness=medium);
17     %CBLAN (recommended-color=white);(recommended-body=light);(recommended-
sweetness=sweet);
18     %VALPO (recommended-color=red);(recommended-body=light);
19     %ZINFA (recommended-color=red);(recommended-sweetness=medium);
20     %ZINFA (recommended-color=red);(recommended-sweetness=dry);
21     %CBSAU (recommended-color=red);(recommended-sweetness=medium);
22     %CBSAU (recommended-color=red);(recommended-sweetness=dry);
23     %PINOT (recommended-color=red);(recommended-body=medium);(rec-
sweetness=medium);
24     %BURGY (recommended-color=red);(recommended-body=full);
25     %BFULL (sauce=spicy);(recommended-body=full);
26     %BLIGH (tastiness=delicate);(recommended-body=light);
27     %BLI30 (tastiness=average);(recommended-body=light);
28     %BME60 (tastiness=average);(recommended-body=medium);
29     %BFU30 (tastiness=average);(recommended-body=full);
30     %BME40 (tastiness=strong);(recommended-body=medium);
31     %BFU80 (tastiness=strong);(recommended-body=full);
32     %BME40 (sauce=cream);(recommended-body=medium);
```

Figure 8.4 The CLIPS Wine KBS in MCRDR format

Some of the features used to classify animals in this KBS appear a little unusual and have occurred due to the purpose of this KBS which is to correctly classify an animal not to show the evolution of man or the food chain. For example, a bird was not defined in terms of having feathers or being able to fly but is defined in Rule 1 as : %BIRDP (backbone=yes); (warm.blooded=yes); (has.breasts=no). The reason this KBS could give the correct classification is that animals which become incorrectly classified as birds were further distinguished on the basis of not being able to fly (as shown in the man example in Figure 8.3) or the bat was distinguished as being able to eat meat and fly (See rule 7 in Figure 8.1). Of course many birds eat meat and most fly. This raises the question of whether the knowledge in this KBS could be reused for other purposes such as teaching a student the structure of the animal kingdom. In this case it appears that this KBS would require major modifications.

8.1.1.2 The CLIPS Wine KBS.

The CLIPS Wine KBS was structured differently to the Animal KBS. There was domain knowledge which included how to decide which wine to select and there was control knowledge which could be further divided into rules that controlled the user interface and other rules that controlled how and when user preferences should influence the general domain rules. The expert domain knowledge was easily converted to decision table format. The user interface rules were ignored. The user preference rules used many variables. Conversion to a decision table would result in a lot of similar rules with only one value different to cover all possible values of the variable. For the purposes of this evaluation only the domain rules have been converted into a decision table and used as input to the MCRDR/FCA modelling tool. A table with 43 rows was developed. The first 32 rules are shown in Figure 8.4. The wine KBS also differs from the animal KBS because the rule conclusions use confidence factors. To handle confidence factors a conclusion such as %BLI30 indicates that the recommendation is body=light with a confidence of 30% or %CWH90 is a recommendation of colour=white with a confidence of 90%. The confidence factors are treated as additional information associated with a conclusion and in the line diagram does not involve any sort of propagation of weights through the KBS.

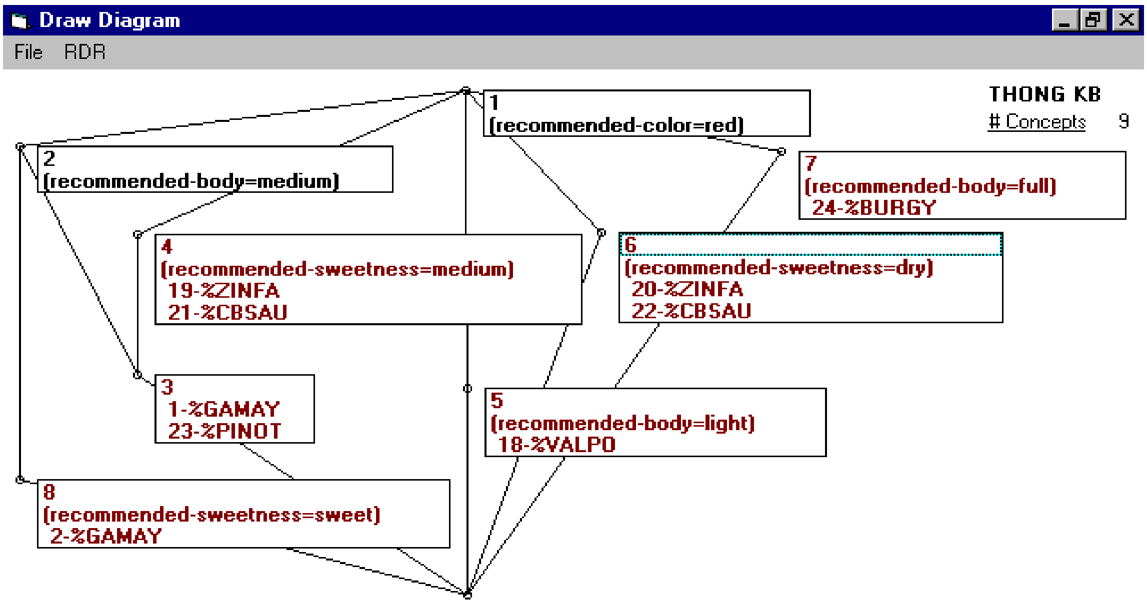


Figure 8.5 The MCRDR/FCA diagram for the selection COLOUR=RED in the CLIPS Wine KBS.

In Figure 8.5 the clause COLOR=RED has been selected. This diagram shows us not only which wines are red but also the BODY and SWEETNESS where it is part of the decision. For example, we can say that if the RECOMMEND-BODY=MEDIUM and the RECOMMENDED-SWEETNESS=MEDIUM the red wine GAMAY or PINOT NOIR would be suitable selections. However, if the RECOMMEND-BODY=MEDIUM and RECOMMENDED-SWEETNESS =SWEET the wine GAMAY would be a better choice.

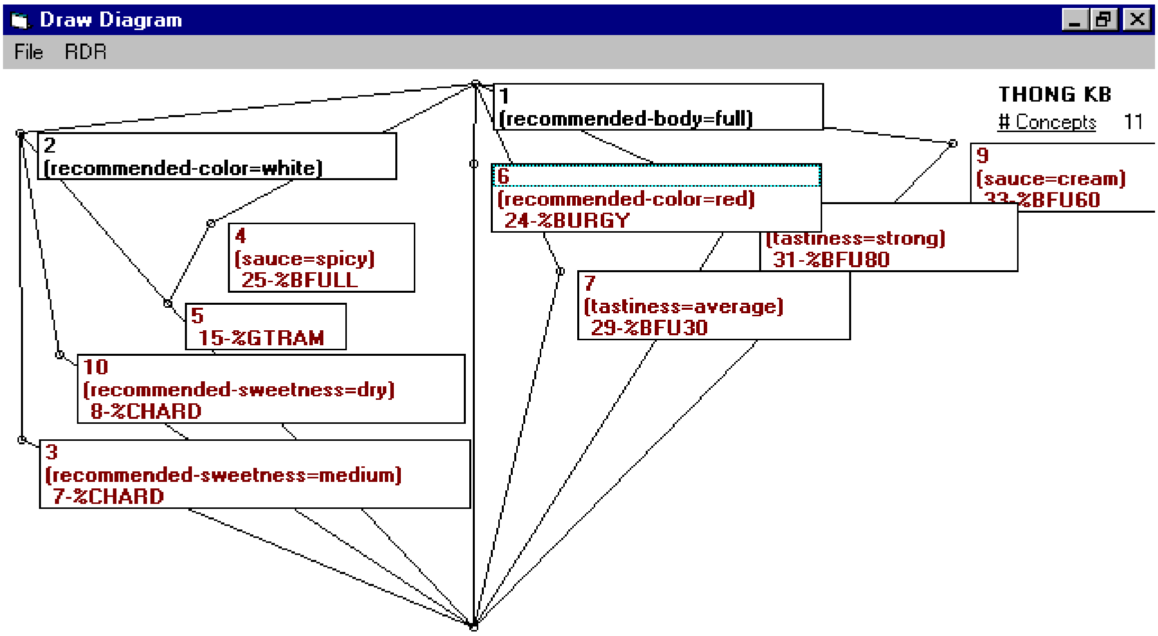


Figure 8.6 The MCRDR/FCA diagram for the selection RECOMMENDED_BODY=FULL.

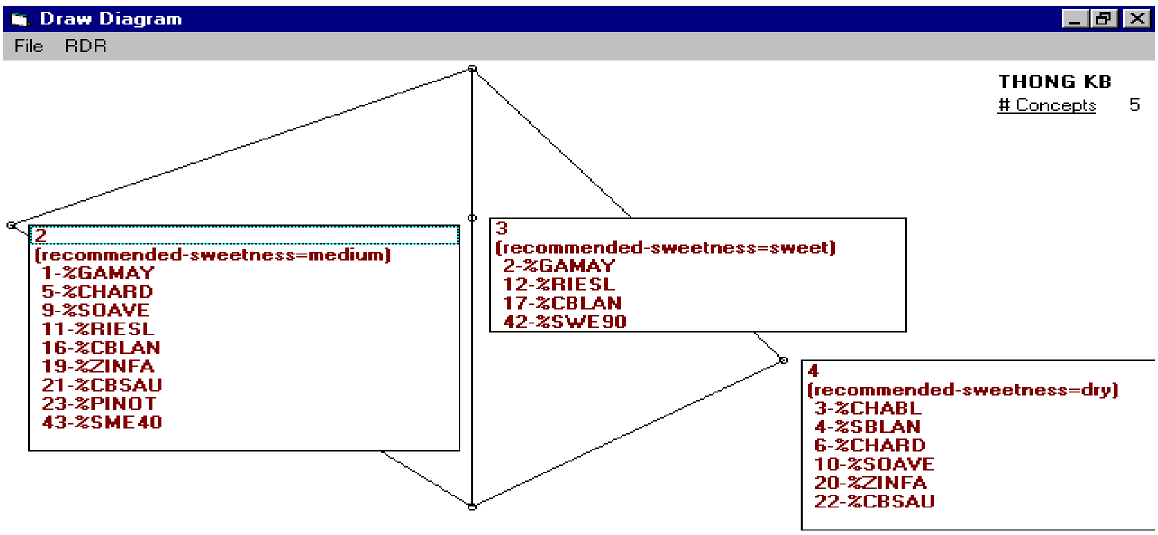


Figure 8.7 The MCRDR/FCA diagram for the selection RECOMMENDED_SWEETNESS using the option Clause with no related attribute in the CLIPS Wine KBS.

In Figure 8.6 the clause "RECOMMDED_BODY=FULL" has been selected. The diagram shows which aspects of a dish are important in this decision. As mentioned, the Wine KBS uses confidence factors (CF). BFU30 means BODY=FULL with a CF of 30%. Therefore, if tastiness is strong then a full body is more highly recommended than when the tastiness is average. If there is a spicy sauce a full-bodied wine should definitely be selected but if the sauce is creamy there is only a 60% confidence that a wine with a full body should be selected.

In Figures 8.5 and 8.6 all attributes used in rules with the specified attribute were included. If the extra information is not of interest the screen will be unnecessarily cluttered. In Figure 8.7 the attribute "RECOMMENDED-SWEETNESS" was selected with the option to not include related attributes. Using this selection we can focus on the sweetness of a wine to see which wines are sweet, medium or dry.

8.1.1.3 Summary

The animal and wine CLIPS KBS have shown that knowledge in rules-bases not structured as MCRDR KBS can be utilised to develop line diagrams from which a better understanding of the knowledge, the relationships between concepts, the higher level concepts and the structure of the KBS can be gained. These are only examples but there is no reason why it would not apply to all propositional systems. We now consider generality from the aspect of visual representation.

8.1.2 The Generality of the Visual Representation

In this section a comparison is made between various representations of the same knowledge to show the compatibility of the line diagram with other possible visual representations.

8.1.2.1 Comparing against Other Taxonomic Representations

In this section we explore the suitability of the FCA line diagram as an ontological and particularly a taxonomic representation. Ontologies are often equated with taxonomies that can be used to classify categories or concept types in a knowledge base (Sowa 1991). A taxonomy contains a stronger concept of structure and organisation than the word ontology as the word is derived from the Greek *taxis* and *nomos*, which respectively mean arrangement and law. Woods (1991) notes that: conceptual taxonomies which allow probabilistic and default rules and abstract and partial definitions maybe "useful for indexing and organising information and for managing the resolution of conflicts" (Woods 1991, p.45). The line diagram was used in Chapter Six for conflict resolution. Taxonomies include a classification system which provides structure that allows new concepts to be positioned. In the FCA framework, any new concept may be located based on the description of its *intents* and *extents*, the attributes and objects that describe the concept. In the MCRDR/FCA framework the taxonomy is recreated each time from the specified formal context and each individual concept is ordered in relation to each other concept. Limited default reasoning is supported in the line diagram through the inheritance of properties (attributes) but the use of roles for default reasoning is not well supported as there are not different links or arcs for different types of relationships between concepts. The line diagram can be described as a concept map with typed nodes but untyped arcs that is a hybrid system which includes a spatial layout and location system, node based system and a link-based system (Kremer 1998, Lambiotte et al 1984).

The line diagram is a complete lattice. Woods remarks that lattices offer "expressive power and the detection and resolution of conflicts in inherited information" (Woods 1991, p.80) that otherwise might not have been discovered until runtime. The concept lattice also satisfies the informal definition of a semantic net given by Shappiro (1991, p.137) as a "labelled directed acyclic graph (DAG) in which nodes represent entities and labelled arcs represent binary relations between entities". The line diagram in FCA is a labelled DAG, although the arcs are not labelled but can be read as a binary relation which means "isa" when traversing the graph from bottom to top for extensions. Nets also provide "clustering of properties around concepts and the incorporation of inheritance hierarchies" (Schubert 1991, p.102) and inference propogation. Each of these features are supported by the FCA line diagram. Woods (1991) sees semantic networks as different to other KR because of the use of links to record facts and store associations which can be used for reasoning. Schubert, however, points out that many, if not most, KR are able to provide many of these features and tries to distinguish a semantic net from other representations. He draws the conclusion that semantic nets should not only be graphic taxonomic reasoners but that they should use graph theoretic notions. This point is made by Shaw and Gaines 1991b in this section's closing quote.

The use of extensional definitions in FCA can be seen as problematic when we come to describe situations for which there are no real examples. However, say that we wanted to compare a horse, zebra and unicorn. There is nothing to stop us describing these three objects in terms of their real or imagined properties in a formal context. FCA is an intensional logic that uses descriptions of the attributes of objects as a means of classifying and ordering objects. Of course, there may be a concept that can not be found in any object, real or hypothetical, and it would not be possible to include this concept in our subsumption hierarchy unless the properties of that concept are contained in the intersection of other examples. This is a limitation of FCA. As can be seen in the example below it was necessary to treat the higher level concepts (collie, cat, robin and starling) as examples in order to include them in the taxonomy.

For comparison we look at some taxonomies produced by Shaw and Gaines (1991b) which they developed to demonstrate how their visual language reflects linguistic semantic arguments. They make a comparison between taxonomies based on Cruse's (1986) lexical semantics and the use of the disjoint and subsume relationships in their visual language. Gaines and Shaw point out that the visual language and other representations of subsumption hierarchies support many of the distinctions that can be made using Cruse's taxonomic lexical hierarchies such as branching or non-branching, the existence of a number of levels, generalisation at higher levels and differentiation and convergence at lower levels. In this section it is shown how the line diagrams drawn using FCA offer a suitable representation for taxonomies.

Primitive(creature)
Primitive(animal, creature)
Primitive(animal, bird, creature)
Primitive(dog, animal)
Primitive(dog, cat, animal)
Primitive(collie, dog)
Primitive(robin, bird)
Primitive(collie, spaniel, dog)
Primitive(robin, blackbird, bird)
Primitive(robin, blackbird, starling, bird)
Primitive(male)
Primitive(person)
Primitive(parent)
Primitive(male, female)
Primitive(man, male, person)
Primitive(woman, female, person)
Primitive(father, man, parent)
Primitive(mother, woman, parent)
Individual(Charles, spaniel)
Individual(Tweetie, blackbird)

Figure 8.8a Knowledge Structure Definitions and Assertions Compiled from Figure 8.8b (from Shaw and Gaines 1991b)

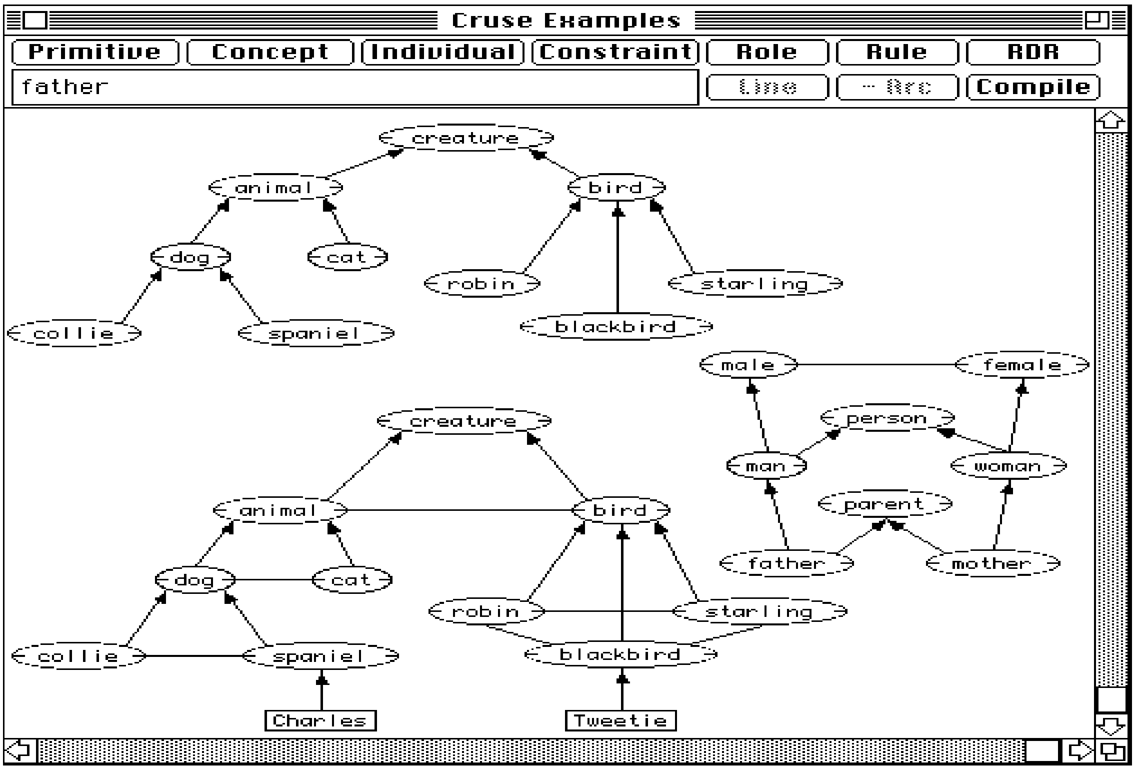


Figure 8.8b Taxonomies in Lexical Semantics in a Tool Based on the Visual Language (from Shaw and Gaines 1991b)

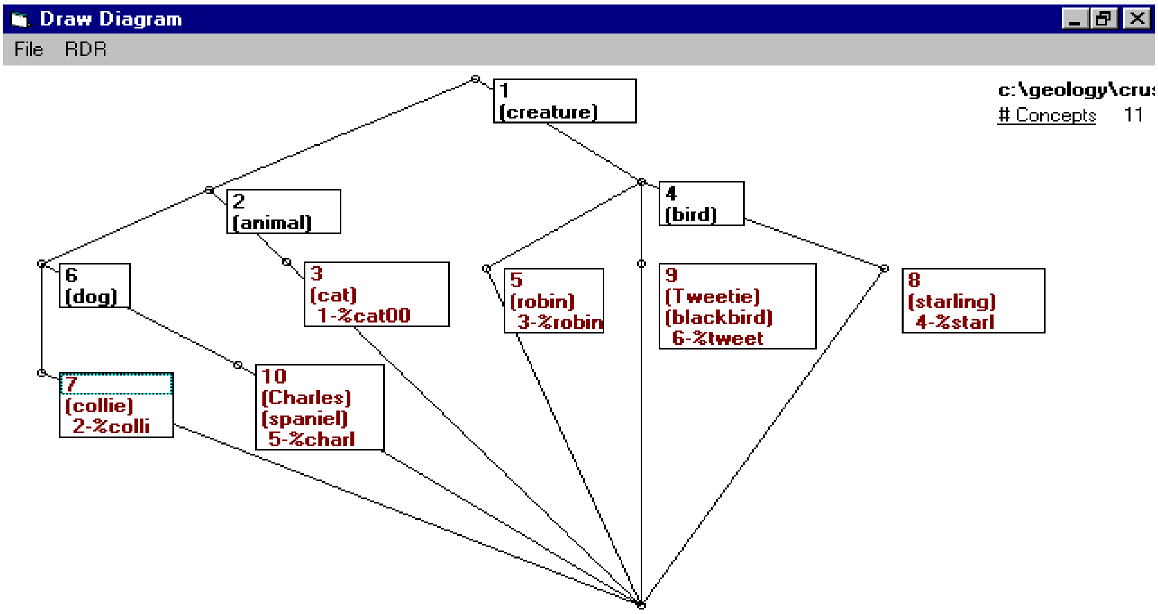


Figure 8.9a: A line diagram to represent the creature taxonomy expanded in the visual language in Figure 8.8a.

1	%cat00	(cat);(animal);(creature)
2	%colli	(collie);(dog);(animal);(creature)
3	%robin	(robin);(bird);(creature)
4	%starl	(starling);(bird);(creature)
5	%charl	(Charles);(spaniel);(dog);(animal);(creature)
6	%tweet	(Tweetie);(blackbird);(bird);(creature)

Figure 8.9b: The knowledge structure used to generate the line diagram in Figure 8.9a.

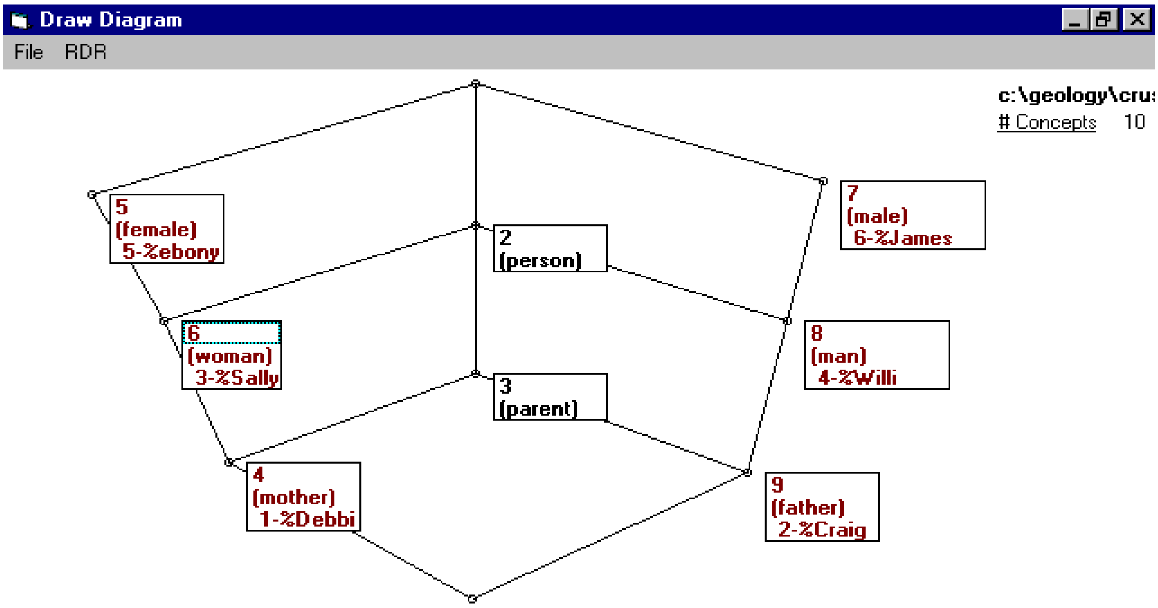


Figure 8.10: A line diagram to represent the father-mother taxonomy in Figure 8.8b.

In Figure 8.8b, the top taxonomy follows the creature taxonomy in Cruse (1986, p. 146) and the centre right father-mother taxonomy is based on the discussion in Cruse (1986, p.262). In the bottom of Figure 8.8b the creature taxonomy is expanded using the visual language. The visual language supports explicit representation of opposites through the disjoint relation which was not obvious in Cruse's taxonomy. The creature and father-mother taxonomy are represented using line diagrams in Figures 8.9a and 8.10, respectively. The key difference between the line diagram in Figure 8.9a and the taxonomy and the visual language graph in 8.8b is the extensional definition associated with each node at or below the *generic level* which is "the ordinary everyday names for things and creatures: cat, oak, carnation, apple, car, church, cup, etc" (Cruse 1986, p.146). In the creature visual language graph, Charles and Tweetie are instances of spaniel and blackbird, respectively. FCA relies on having examples to compute the higher level concepts. If there were instances of each type of animal or bird then the structure shown in Figure 8.9b which includes rules to represent the higher level concepts of cat, robin, starling and collie, would not have been necessary and would have resulted in a line diagram that only had extensional definitions shown at the bottom nodes. The closeness of the structure in Figure 8.9b to the KL-ONE-like structure produced by the visual language shown in Figure 8.8a is notable. Apart from the input being more succinct in the MCRDR/FCA rule representation (six rules compared to twelve based on the visual language), the key difference is that with MCRDR/FCA pathways[4] we started with the rules and derived the diagram and using the visual language the knowledge structures in figure 8.8b were derived from the graph in figure 8.8a.

The similarity to and strength of the visual language and the line diagrams over Cruse's representation is well summarised by Shaw and Gaines (1991b):

> *At one level they are isomorphic visual representations of a taxonomic structure. At another level, the computational representation has the added feature that the underlying data structure exists within the computer and hence is readily subject to graph-theoretic analyses of its fine structure. At another level, it is also an operational knowledge structure within the computer supporting automated deductive inference. At another level, this knowledge structure is available for composition with other knowledge structures to provide a knowledge-based system suitable for testing theories and supporting practical applications. (Shaw and Gaines 1991b, p. 12).*

In the next subsection we expand the comparison beyond taxonomic representations to a number of other graphic representations of the same knowledge. The focus is more on the acceptability and usefulness of the representation for human reasoning.

4 Pathways means the exception structure has been flattened by picking up all rule conditions leading back to the head rule.

8.1.2.2 Various representations of Cendrowska's Lens Data

One of the claims of this thesis has been that the line diagram is a useful representation for understanding the contents of a rule-base. In this section we consider a number of visual representations of the same set of rules and their various merits. We consider such things as clarity and the ability to determine implications as well as the ability to show relationships between concepts and the knowledge structure. The discussion is kept to a minimum as the results of a survey, described in Section 8.2.1.4, act as a good summary of the author's intuitions of these various representations. The six representations are: an MCRDR KBS (Figure 8.11), a formal context or crosstable (Figure 8.12), a concept matrix (Figure 8.13), a line diagram (Figure 8.14), Gaines and Shaw's (1993) visual language (Figure 8.15) and Gaines' Exception Directed Acyclic Graph (EDAG) (Figure 8.16). Rules for Cendrowska's lens data are used in all diagrams for consistency and comparison. Each diagram is shown together with a brief description of its considered strengths and weaknesses.

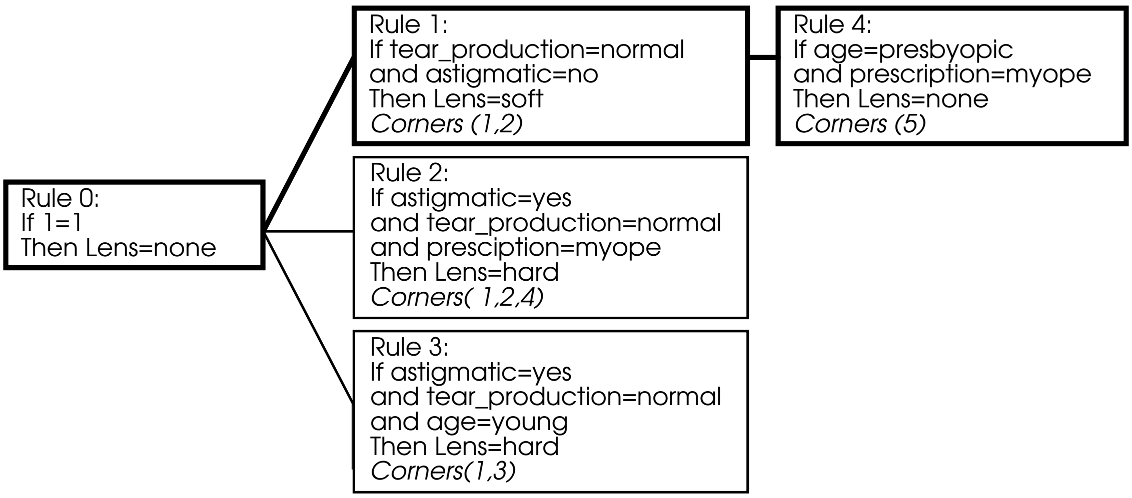


Figure 8.11. An MCRDR KBS for the Contact Lens Prescription Domain. The highlighted boxes represent rules that are satisfied for the case {age=presbyopic, prescription=myope, astigmatic=no, tear_production=normal}. The classification given is Lens=none. As this domain only deals with mutually exclusive conclusions we only get one conclusion, but if the domain was extended to cover spectacles and bifocals then this case could lead to multiple conclusions being given. The cornerstone case numbers associated with a rule are shown as Corners().

The MCRDR KBS shown in Figure 8.11 shows the exception structure of the representation. An added feature of the MCRDR KBS is the storage of cases associated with each rule which is a valuable aid in providing the context in which the rule applies as well as an aid in KBS maintenance and validation. While the format of the diagram shown is only one possible way that the MCRDR structure can be visually represented, the diagram is adequate since it displays the key aspects that need to be shown which are the exception structure and the associated cases.

In Figure 8.12 the same rule pathways have been flattened into a formal context. The formal context, or crosstable, provides the benefits of a decision table structure. Each rule can be compared using the same features and in the case of rule bases that use intermediate conclusions the decision table format results in a normalisation of the data to the one base level. Decision tables are a useful format for further processing by a large number of techniques including rough set theory and formal concept analysis as described in Chapters 4 and 5, respectively.

	1=1[5]	astigmatic = no	tear_production = normal	age = presbyopic	prescripton = myope	astigmatic = yes	age = young
1-%LENSN	X						
2-%LENSS	X	X	X				
3-%LENSN	X	X	X	X	X		
4-%LENSH	X		X		X	X	
5-%LENSH	X		X			X	X

Figure 8.12: Context of "MCRDR Contact Lens Rules"

As described in this thesis the data/rules in the crosstable may be transformed into a concept matrix, which is shown in Figure 8.13, by using FCA to find the higher level concepts represented in the intersections of sets of attributes and the sets of objects that share them. The concept matrix shows some of the structure of the concepts but because some concepts are children or parents of more than one concept the ordering will affect the appearance of which concepts are related. This limitation is overcome in the line diagram or concept lattice. The researchers at Darmstadt do not seem to consider the concept matrix as a major representation but see it more as an intermediate representation of the preferred concept lattice. This study sees the concept matrix as more than just an intermediate representation and is particularly useful as a more concise representation of the data that in some situations may be more useful than the line diagram. For example, in the Lotus domain a comparison of the concept matrices revealed patterns and differences between the KBS developed by the four agricultural advisors as shown in Figure 5.15. To be able to visually detect differences in the patterns it is necessary to order the attributes so that they are in the same order in each KBS. Similar to the preliminary work done on comparing conceptual models in Chapter Six, the matrix can be further exploited to support automatic comparison of concepts.

As described in Chapter Five, the concepts in the concept matrix can be ordered and drawn as a complete lattice as shown in Figure 8.14. The main strengths of the concept matrix is the ability to see higher level concepts from the primitive rule concepts and the structuring of the concepts into a subsumption hierarchy.

[5] 1=1 is the condition in the default rule which give the default conclusion as shown in Figure 8.11.

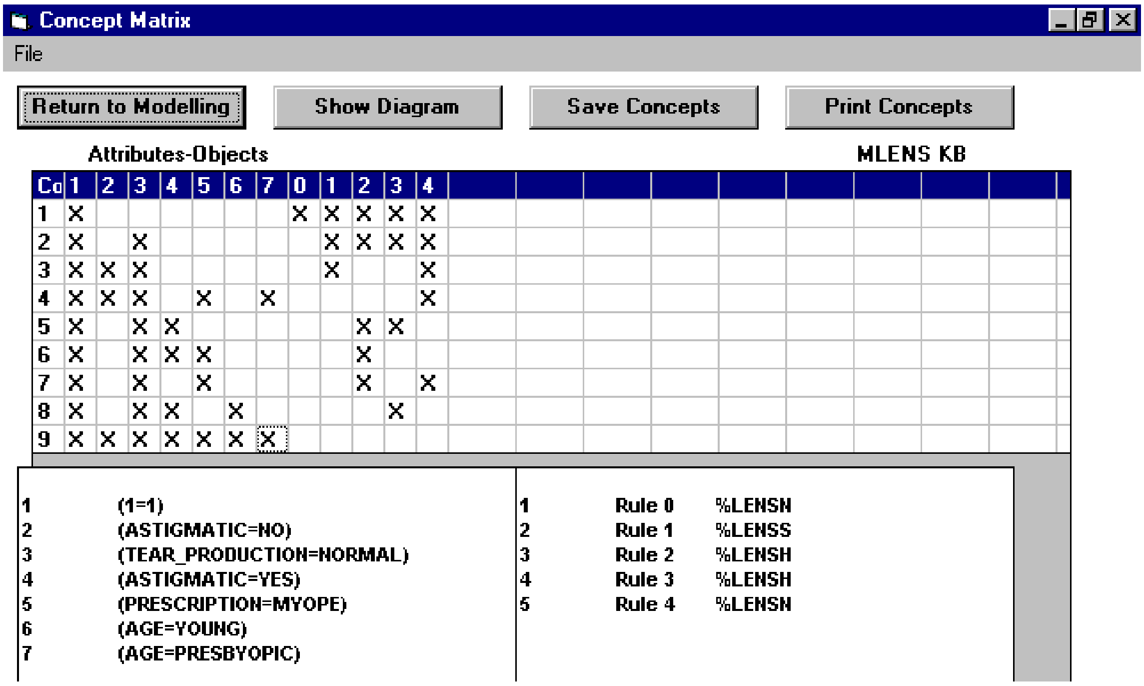


Figure 8.13: The concept matrix screen in MCRDR/FCA for the Formal Context "MCRDR Contact Lens Rules" given in Figure 8.12.

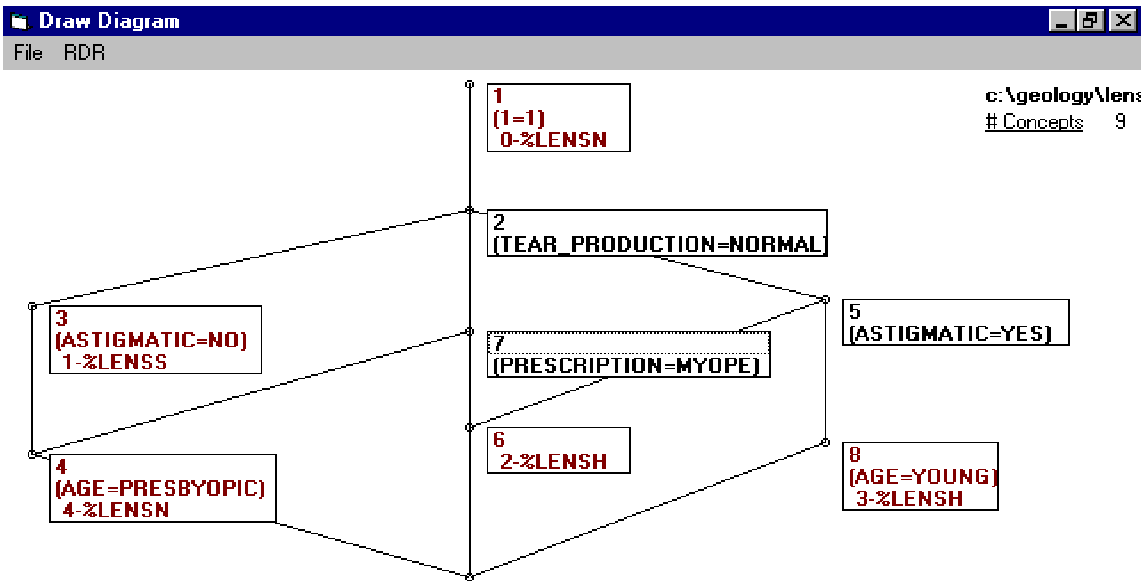


Figure 8.14: The Diagram Screen in MCRDR/FCA which shows the Concept Lattice for the Formal Context "MCRDR Contact Lens Rules" given in Figure 8.12.

The visual language described in Section 3.3.1.3 (Shaw and Gaines 1991b and Gaines and Shaw 1993a) and shown in Figure 8:15 offers a graphic representation of the subsume and disjoint logical relations. The visual language provides a domain model that is not present in the other representations and shows how the rules fit into the model. This provides a rich representation but as to be expected as more information is included the diagram becomes more complex compared to the other representations.

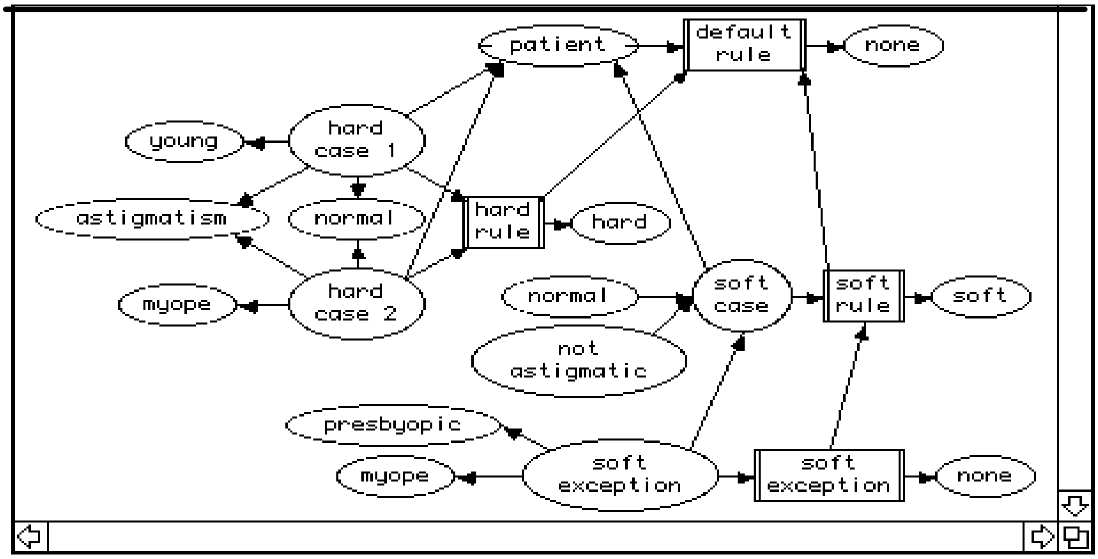


Figure 8.15 Contact Lens Rules Represented in the Visual Language (from Gaines and Shaw 1993a)

The final diagram to be considered is Gaines' (1996) Exception Directed Acyclic Graphs (EDAGS). The similarity between the MCRDR/FCA diagram and the EDAG in Figures 8.14 and 8.16, respectively, warrant a more in depth discussion of the EDAG and how they compare to the FCA line diagram. The development of the EDAG was motivated by the desire to develop knowledge representations that encode knowledge in a similar way to the expert (Gaines 1991a) and to make knowledge comprehensible (Gaines 1996). As a warning that the objectives stated by Miche and Quinlan may not always be achievable, Gaines notes:

> *"It is not obvious that an excellent performance system necessarily implies the existence of a comprehensible knowledge structure" (Gaines 1996, p. 206).*

Instead, compactness, coherence of the structure and the use of familiar concepts are all relevant when determining comprehensibility (Gaines 1996). Gaines considers the use of default reasoning and rules with exceptions to be a closer representation to the way experts reason than decision trees or production rules. Gaines (1996) describes how by generalising parts of the tree or generalising rules with exceptions it is possible to develop more compact knowledge structures. In the case of a tree representation, generalisation into partial orders results in branches rejoining and avoids the replication of sub-trees. The similarity with RDR is noted by Gaines with the difference identified as the inability of the RDR structure to allow multiple conclusions. This difference, however, does not exist with the newer MCRDR representation. Using the features of an

EDAG described in Gaines (1996) we now compare them to the MCRDR rule pathways and the MCRDR/FCA line diagrams.

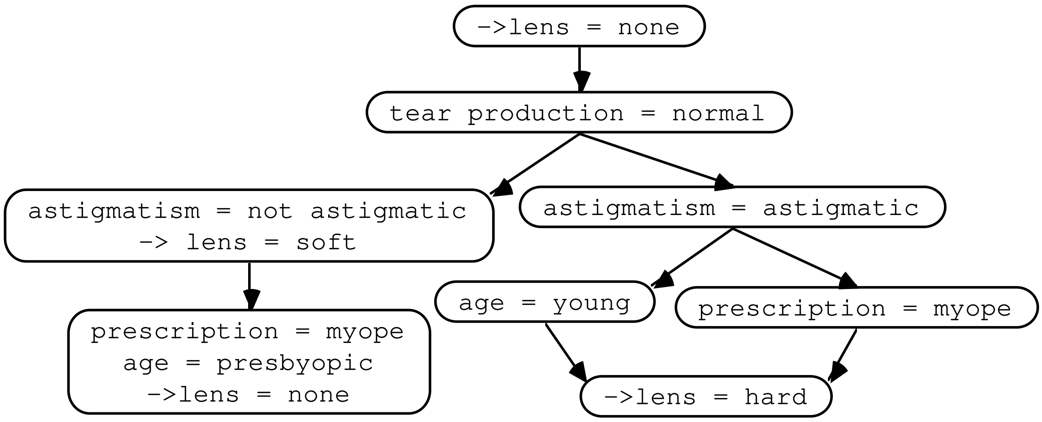


Figure 8.16 Induct rules with exceptions, factored as EDAG (from Gaines 1996)

1. An EDAG is not rooted and may have several disconnected parts. If a default rule is not used, which is not needed in many domains, an MCRDR is not rooted and each pathway at the first level could be treated as an individual disconnected part. A concept lattice is a complete network with a top and bottom root and is not disconnected.
2. Concepts in an EDAG and an MCRDR do not have to be mutually exclusive and multiple conclusions can occur. In both an EDAG and an MCRDR, a rule that has no exceptions will be concluded unconditionally if the premise is true.
3. The structure of an EDAG and an MCRDR is not binary but may have more than two nodes branching from a node.
4. The structure of an MCRDR is a tree, no branches may rejoin. However, the structure of a concept lattice which uses the intersections of sets of attributes and objects from the MCRDR rules is not a tree and, therefore, like an EDAG.
5. An MCRDR is made up of nodes which contain one or more conditions in the premise and one conclusion. A node in an EDAG or a concept lattice does not have to contain both a premise and a conclusion, but a premise may represent a common condition in a number of rules or a conclusion may represent a default or common conclusion.
6. A null node may be used in an EDAG to avoid crossing of lines. This approach is also used by the Darmstadt group in their drawing of concept lattices. Although this technique has not been used in the MCRDR/FCA diagrams it would be a satisfactory solution to reduce diagram complexity where line crossing can not be removed by repositioning nodes.
7. The notion of a premise as a potentially decidable predicate holds for both an EDAG and MCRDR. Possibly RDR (PRDR) was developed (Mulholland et al 1993) to handle an ion chromatography configuration problem where a premise could fire in a number of ways such as true (T), possibly true (P), unknown (U) or false (F).
8. The relationship between using an EDAG for inferencing by following the arrows and treating the EDAG as a semantic network when the arrows are reversed is compatible with the use of the MCRDR structure for inferencing and the concept lattice as a semantic network. By reversing the arrows in the EDAG or line diagram we can see the 'isa' relationships by following the inheritance of properties. Multiple inheritance pathways are treated as disjunctive concepts (Gaines 1996). In a

pathology system currently under development the MCRDR is more of an EDAG-like structure without an implicit OR.
9. The interpretation of a node in an EDAG, MCRDR and concept lattice all require tracing paths back to the root and consideration of the children nodes.

In this section a number of representations of the contact lens data have been shown including the MCRDR rules in Figure 8.11, a formal context based on the MCRDR rules in Figure 8.12, the concept matrix and lattice derived from them using FCA in Figures 8.13 and 8.14 respectively, the visual language of the PCP concepts and rules in Figure 8.15 and an EDAG in Figure 8.16. The similarity between the concept lattice and the EDAG was noted, with the concept lattice providing not only the inference paths but also an extensional definition as does the visual language in Figure 8.15. The MCRDR rule structure provides the inference paths more clearly, shows the cases associated with each rule and is the simplest representation but does not contain any of the higher-level concepts shown in Figure 8.15, which is the most complex diagram.

To evaluate my findings of the various strengths of these representations a number of computer science and engineering students and academics completed a survey of these six representations. This evaluation is found in part four of the survey which is described together with the results and discussion in section 8.2.1.4.

Section 8.2 Using Line Diagrams for Learning

The purpose of this section is to evaluate the usefulness of the line diagram for teaching and instruction. Two approaches to testing this claim have been taken. One is from the viewpoint of the novice the other is from the viewpoint of the expert. In the first evaluation a survey was used to discover how well a novice can use the line diagram for learning. In the second evaluation experts are used to comment on what has been learnt by a novice.

8.2.1 The Survey

A survey was conducted using computer science and engineering postgraduate students and academics that were attending a postgraduate workshop. Most of the participants were students from the Artificial Intelligence Department but all participants, except for one, did not know anything about line diagrams, formal concept analysis or the domains that were being used. Sixteen questionnaires, as can be found in Appendix B, were given out but only 12 were answered. In some cases questions were missed. The questionnaire was divided into four sections, each testing a different aspect of the line diagram. I piloted the survey on a Master of Applied Science (Information Studies) student who identified parts of the questionnaire that needed to be reworded and which parts needed a verbal explanation. Each section is now discussed. The survey was handed out as part of a computer science workshop for postgraduate students and academics. A slot of 25 minutes was allocated and, although the time was extended to 40 minutes over the lunch break, each section was rushed and insufficient time was provided to explain each section adequately, possibly affecting the quality of the results obtained.

8.2.1.1 Learning to read a line diagram - Section One of the survey

The first section of the survey was concerned with giving the subject some familiarity with the line diagram as well as testing how quickly the line diagram could be learnt to be read. I showed a line diagram on the overhead projector and explained briefly what a concept in formal concept analysis was and how the diagrams could be read. I selected two different concepts on the diagram to explain how the set of attributes and the set of objects for a concept could be found. Each subject was then required to look at their questionnaire giving the set of attributes and the set of objects for two concepts on one diagram and two concepts on another line diagram. On the questionnaire itself it was again stated how attributes and objects were found. Also an example using one of the concepts on the first line diagram was given which listed the set of attributes and the set of objects for that concept. About 8 minutes was spent on this section.

	WV	CS	MK	PC	SL	LK	JWK	AEL	MH	XYZ	MP	KD
Q1	6/6	5/6	6/6	6/6	6/6	6/6	6/6	6/6	5/6	2/6	4/6	6/6
Q2	5/5	5/5	4/5	5/5	5/5	5/5	4/5	5/5	4/5	3/5	4/5	5/5
Q3	6/6	5/6	5/6	6/6	6/6	6/6	5/6	6/6	6/6	3/6	4/6	6/6
Q4	7/7	5/7	5/7	7/7	7/7	7/7	4/4*	7/7	5/7	4/7	5/7	7/7

Table 8.1. The results for section one of the survey.
The denominator is the total number of attributes and objects. The numerator is the total number of correct attributes and objects listed by the subject. If extra attributes or objects were added they were subtracted from the correct number. Subjects 10 and 11 did not grasp what was to be done or how to read the diagram since extra information was given that indicated they were guessing.
* This subject missed the second part of the question which asked to list the set of objects.

As shown in Table 8.1, from the twelve subjects, six (6) got all of the descriptions correct. Another four (4) got most of the descriptions correct and two (2) obviously hadn't grasped how to read the line diagrams because they wrote down extra things or left things out. So 10/12 gained a reasonable to good understanding of how to read a line diagram after a one minute verbal introduction and a brief written description and example in the questionnaire. In the pilot study I had been able to correct a misunderstanding which I was not able to do with this large group. I realised that misunderstandings at this point may jeopardise the other results. Two options were to correct the individual answers and discuss any misconceptions or show the group the corrects answers. However both options would reduce the time allowed on the other more interesting sections and the second option ran the risk of subjects changing their original answers which would alter the results.

8.2.1.2 Comparing Line Diagrams with Rule Traces - Section Two of the survey

The purpose of Section Two of the survey was to determine whether the line diagram provided information in a more concise and understandable form than browsing a rule trace. The questions to be asked had to be simple and answerable from one diagram since the subjects were novices in the domain and with the representation and the subjects did not have access to the tool to perform their own inquiries. Since the line diagrams were based on the rules in the knowledge base it was obvious that with investigation most questions could be answered from the rules. The higher level concepts would be harder to find by browsing the rules but even this information could be gleaned by a skilled person. This meant that the speed with which a question could be answered was the key aspect to be measured. It was not possible to ask someone to first solve the problem using a rule trace and then solve the same problem with a line diagram, or doing these tasks in reverse order, since they would already know the answer after they solved the problem the first time. It was decided to split the subjects into two equal groups, each group using either the line diagram or rule trace to answer the question. Two sets of questions were asked and each group was assigned to a different source for each question. For example if group B used a line diagram for question one they used the rule trace for question two.

Group A were given a set of sheets which included printouts of the rules for questions 1.1 and 1.2 in Section Two and a line diagram for answering questions 2.1 and 2.2. The other group, Group B, were given a line diagram to answer questions 1.1 and 1.2 and printouts of the rules for answering questions 2.1 and 2.2. Both groups were given the chance to use the different sources (but only one source for each question) so that there was no particular bias that some people were better at using one source over another.

I had some reservations over the nature of the rule trace that should be given. I decided that I would provide the rule trace that is available in MCRDR/FCA which is what is provided in the basic MCRDR for Windows system. I realised that having to look at five screen dumps as compared to one line diagram had an obvious disadvantage. To minimise the amount of paper shuffling I put two screens on a page so that there were three pages at which to look. This may also be criticised but the point is that it does take time to browse through rules and I had already cut down the user time since they did not have to search for the relevant rules and they had all the information in front of them at one time. In the real system you only got one rule at a time. Perhaps to reduce the paper shuffle I should have provided a dump of the rules in the format shown in Figure 8.17.

```
21 1 0 19 %PL000 : (GRAIN_SIZE = COARSE)
22 1 0 21 %VC000 : (GRAIN_SIZE = FINE)
23 1 0 22 %PL000 : (GRAIN_SIZE = MEDIUM) & (SILICA = INTERMEDIATE)
24 1 0 23 %VC000 : (GRAIN_SIZE = MEDIUM) & (SILICA = LOWISH)
25 1 0 24 %PL000 : (GRAIN_SIZE = MEDIUM) & (SILICA = VERY_HIGH)
```

Figure 8.17 A dump of the rules in the SISYPHUS III KBS that used the conclusion plutonic or volcanic.

However, the user does not have access to the knowledge in the format shown in Figure 8:17 as part of the MCRDR system so I could have also been criticised for not comparing against the existing facilities in the tool. Two different participants felt that the rule trace I had provided did not show the structure of the rules. This was what I was trying to demonstrate. The point I was trying to make is that by structuring the knowledge into a line diagram, showing the hierarchical relationships between rules and higher level concepts, we are better able to understand the rules and the relationships between rules. The assumption by those that criticised the nature of the rule traces offered was that the rules themselves had structure which provided information on the relationships between them. As can be seen in Figure 8.17, this was not the case with the examples used and none of the rules were exceptions to other rules besides the default rule. Another point is that I would have assumed greater familiarity by most subjects with the use of a rule trace than the use of a line diagram for browsing rules. I would not have thought that the page layout would have made a substantial difference especially since I showed an overhead of the rule trace and pointed out where to find the conditions, the conclusion and conclusion description before anyone commenced, even though each of these were labelled on the screen printout they received.

As noted earlier, 16 questionnaires were taken. Care was taken to give out equal numbers of Group A and Group B type questionnaires. However, only 12 were filled out and of these, 7 were in group B and 5 were in group A. This made comparison a little harder and the results are quite spread. Some people obviously had a reasonable grasp of the line diagram and others hadn’t worked out how to use it properly. With proper understanding of the line diagram these questions can be answered in seconds. For example question 2.2 which asks which attributes/conditions are shared by all concepts can be answered immediately by looking at the top concept.

The results for each question are given in Tables 8.2(a - f). The numbers shown in the tables are the number of minutes and seconds it took to answer the question. An 'X' marked in front of the time indicates the answer was wrong. A question mark shows that the question was not answered or the answer was incomprehensible. The results for each subject are shown in the same order in each figure for comparison. A brief synopsis of the results under each figure are given.

Question 1.1	Group A 0 Rule Trace						
	3:00	0:52	?2:13	1:00	1:45		
	Group B - Line Diagram						
	3:30	2:00	X2:12	X3:00	0.15	0.20	?2:00

Table 8.2(a): The time in seconds taken to answer Question 1.1 in Section Two.
The two fastest times and correct answers used the line diagram. Three of the long times using the line diagram got the wrong answer indicating a poor understanding of how to read the diagram. There were three reasonable medium times using the Rule Trace.

Question 1.2	Group A -Rule Trace						
	1:30	1:07	?1:36	2:00	1:15		
	Group B - Line Diagram						
	?1:30	?1:00	1:00	0:30	1:00	1:45	?4:00

Table 8.2(b): The time in seconds taken to answer Question 1.2 in Section Two.
This question required listing the conditions under which a medium-grained rock should be classified as plutonic or volcanic. The substantially shortest time is again using the line diagram, but most times using either method are similar. Note that it is a different person to the two people with the shortest times in question 1.1. Questions 1.1 and 1.2 both required searching through most of the problem space in either technique so that may account for inconclusive results.

Question 2.1	Group A - Line Diagram						
	0:45	0:48	0:17	1:00	0:45		
	Group B - Rule Trace						
	4:20	?	3:05	1:30	1:00	1:45	?4:00

Table 8.2(c): The time in seconds taken to answer Question 2.1 in Section Two.
The times are definitely the best using the line diagram.

Question Four	Group A - Line Diagram						
	0:25	0:38	X0:32	0:20	0:30		
	Group B - Rule Trace						
	0:45	?	0:37	X0:60	1:00	0:40	X4:00

Figure 8.2(d) The time in seconds taken to answer Question 2.2 in Section Two.
Again, the times are definitely the best using the line diagram. Questions 2.1 and 2.2 are more concerned with relationships between rules/concepts which may be why we see a dramatic improvement. The lowish times using the rule trace may occur because by the time you get to 2.2 the subject has possibly already worked out which attributes are shared when they were looking for the closest concept. It probably would have been better to put the questions in the reverse order.

As described in the commentary with each table, the results for the first question were inconclusive and may have been due to the nature of the questions which related directly to the primitive concepts and did not require any type of reasoning or abstracting to be performed. The second set of questions involved relationships between concepts and the structure of the rules and the results strongly favour the line diagram over the rule trace.

In addition to answering the above questions the subjects were asked to make optional statements regarding the rule traces and the line diagrams. These comments are included for interest in Figure 8.18.

General comments
The questions are ambiguous I feel. Does 2.2 mean that the conditions are shared in value or just both shared in the rule antecedent ? What is closer ?

Might have been fairer comparison with better description of the use of the MCRDR method. Your results should still work, I'd think.
What do you mean by a closer conclusion

Comments regarding rule traces -
No implied notion of closeness
Familiarity bound - takes time to learn how to read. Still harder to make easy comparison
Pretty incomprehensible
Impossible to read, but more due to the layout than rule structure.

Comments regarding line diagrams -
Not clear how to derive conclusions
Restricted to planar concepts (for humans) Note consider effect of line diagram practice versus no rule traces practice. Rule trace seems to have higher learning curve.
Seem to prompt intuition a bit better, though still not very understandable.
Poor layout of rules may make this test unfair. Also these line diagrams are simple. What happens if you need several pages ?
How do I know when something is a conclusive classification or just an intermediate step ?
I like the line diagrams more.

Figure 8.18: The optional comments regarding rule traces and line diagrams in Section 2.

The comments in Figure 8.18 tend to favour the line diagram over the rule traces but no conclusions or claims are made based on the comments. The comment that the rules appear to have a higher learning curve than the line diagram is interesting as it was assumed that most computer science graduates, in particular AI postgraduates, were familiar with similar representations.

8.2.1.3 Line Diagrams for Learning - Section Three of the Survey

This section was the hardest for the subjects and the most subjective to evaluate because it has the least structure or guidance and required some deeper understanding of how to read the line diagram and how to reason about it. Only eleven (11) of the 12 subjects answered this section.

Firstly, the subject was shown a line diagram using the Lotus agricultural domain and a description of what I could see from the diagram. They were asked, optionally, to comment on my analysis and then they were shown a different diagram from the same domain which they were asked to analyse and comment on. The comments regarding my description are shown in figure 8.19. My annotations are shown in italics. Generally, there is agreement with what I have said and in some cases there seems to be understanding of why I had made a particular statement. There is obvious difficulty with understanding the representation. In particular the labelling for objects (the rule number and conclusion code) was found to be problematic. These difficulties are to be expected as the line diagrams do take some time to understand and as was noted in Chapter Five, there are limitations on the appropriateness of the extensional definition. The hierarchical and semantic net nature of the line diagram is the hardest to grasp but is also what gives the representation its strength. It may have been helpful to replace the conclusion code with a brief description of the conclusion in the extension labelling although the description of each code was provided on the page that it was used.

Do the observations seem reasonable ?
Yes, but not entirely. For instance, you said that concept 2 is an indicator of poor conditions for Maku. Specifically that Low-PH and not Waterlog aren't good for Maku. But on the right, Maku is recommended in concept 6, but negated in 7, when the WATERLOG condition is added. This might mean that in concept 2, the waterlog condition is not important but the low PH is the actual indicator of poor conditions for Maku ?
Sort of, except that I don't understand why concept 4 is linked to concept 5 (*doesn't understand the subsumption hierarchy*)- we already know it shouldn't be Maku Lotus. Also it doesn't cover all situations.

Makes a lot of assumptions.

Underlined sentence - "this hypothesis could be tested further by looking at all rules that use these two conditions" - may also want to see whether there are other rules not recommending Maku.

XYZ- Initially it is hard to say (what distinguishes a concept from a conclusion). Even after a few minutes this is hard to understand for a lay person. Arcane notation (such as 13-%CLOVR) is well-high impossible to remember (*but I have it on the same page ?*), very distracting and off-putting and requires learning. Of course, once learnt, it can be used to judge the reasonableness of the conclusions (I imagine) without much difficulty. But there is little incentive to learn it.

MH - yes

AEL - Does it follow that given 2 (not 3) implies S?
LM- Yep
SL - The last statement may not be true, as you have explicitly filtered for rainfall, but low precipitation areas may still be fertile, to be independent of rainfall.
PC- okay but very hard to tell. There is a potential conflict between 2 3 5 and 6 - can subsume

Figure 8.19: The comments regarding my analyses of a line diagram in section 3.

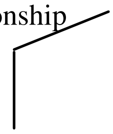
It was interesting to see if others could follow my reasoning based on the line diagrams and the comments in Figure 8.19 indicate that this was possible. The comments showed that the line diagrams could be used to initiate discussion about a domain and to suggest further aspects of the domain knowledge to be considered. This is consistent with the claim that the line diagrams can be used to assist in the negotiation process required during conflict resolution of requirements as suggested in Chapter Six. Further comments from the subjects which allowed for free comment are shown in figure 8.20 which primarily contain comments regarding interpretation of the line diagram.

Further comments ?

Why do things have to connect to the dot at the bottom ? It just makes the diagram more confusing (same person that didn't understand the lattice as a subsumption hierarchy).
What if I get stuck midway in the tree eg rainfall < 1200 but > 1000 ?
Should the bottom node be labelled %NC000.

AEL - on surface may many missing cases. However, some of these may be impossible
Guessing 

'or' relationship



'and' relationship

is this mutually exclusive ?

SL - Why is the box '5' there at all ? 4 already concludes Shalw, 2 conclude nothing, so you could move the conclusion from 5 to 2?? This, if true, would also mean that '2' is not a high level concept. Actually what is a primitive concept ?

PC- These probably only become useful if there was a motivation for learning. Otherwise pretty hard.
WV- Much of it seems to be reasonable. However, there seems to be insufficient respect given tot he symmetry between the two ancestors of concept 5 - the suggestion is that concept 2's application has priority over the (possibly redundant) concept 4; I imagine this is a mistake. I would certainly agree with the last paragraph above- that the information is incomplete.

Figure 8.20: The optional comments regarding the use of line diagrams for learning in section 3.

However, it was the next part of section three that was of greater interest to this survey. In the next part of section three of the survey the subject describes their understanding of the domain from the line diagram. To me the most important thing is not how correct the analysis is but the fact that all 10 of the 11 subjects were able to make some comments regarding the domain knowledge. The comments are reproduced in Figure 8.21. From the comments it can be seen that eight of the subjects seem to have gained some knowledge of the domain and four of these have got a feel for the structure and order in which decisions should be made. A major benefit is not just the answers it can give but the questions that it raises.

KD- Well this graph blows my last analysis out of the water, since concept 2 explicitly says that Low-PH is good for Maku. Low-Ph =no indicates the conclusion %Salin, I guess, but further expansion of the graph to cover the not-waterlogged case should be considered. (*good for hypothesis testing*)

CS- Concepts above subsume concepts below.
If low-ph = yes then Maku except
  if waterlog=no then Sharn except
    if rainfall >= 1000 then Sharn=no then Kenya
    else if rainfall <=1000 and soil then Shalw
  else if flooding >=3 then flood except
else if low-ph and waterlog=yes then Salin

MP- There are no higher level concepts. Ph does not affect Maku. (*This was a subject who had difficulty reading the line diagram in part one of the survey)*

XYZ- I'm not sure what this tells me.

MH- All groups have a classification associated.
Ph is a high-level attribute
- with low_ph=no and waterlog, %Salin - Maku Ok, but others should be considered.
- missing knowledge - low-ph (=no) and not-waterlogged.
- low_ph is okay for maku
  unless not waterlogged - in which case Sharnae or Goldie
   unless
    rainfall > 1000 and not saline - Kenya
    rainfall < 1000 and soil_depth <=30 - Shalw
  flooding >=3 then Flood.

AEL - Low-Ph is good for Maku. Flooding is bad but the soil should be waterlogged. If not, Sharnae is a good alternative, except in low-rainfall and shallow soil. With higher rainfall and non-saline soil, Kenya grass should be considered. In Non-low_ph soil that is waterlogged, Maku can be used, but other legumes may be better.

LM - Maku is suitable in Low-Ph areas that do not flood/waterlog i.e. it's better in dryer areas unless there is shallow soil. (left hand-side).
Maku may however be okay in wetter soils with high_ph if used in combination with other legumes.

SL - Ph seems to have no bearing on Maku, No conclusions based on Ph suggest to plant Maku. (*didn't understand the diagram).*

PC- Maku is useful for low-ph areas, but there are plenty of exceptions - hard to get consistent the theory.
MWK - If there is not a low ph and the land is waterlogged, then consider strawberry clover and other legumes. Otherwise, in general Maku is recommended. However, if there is flooding , don't use maku lotus. Care must also be taken if the area is not waterlogged, in which case Sharnae or Goldie should be used, unless it's not saline and it rains a lot in which case you should use Kenya Clover. Alternatively, if it doesn't rain and the soil is deep don't use Maku Lotus.

WV- Obviously if acidity is low (high_ph) we would hope for waterlogging - otherwise we would not know what to do.
Otherwise, if there's no waterlogging - go for Sharn else -----

Figure 8.21: The subjects observations from the line diagram they were asked to analyse and comment on from the LOTUS agricultural domain.

8.2.1.4 Comparing Representations - Section Four of the Survey

Six different representations of the rules which cover Cendrowska's lens data, which were presented in Section 8.1.2.2, were shown to the subjects using an overhead projector. A printed copy of each diagram was taped to the top of the table and when the subjects were moving off to lunch they were asked to circle a number between 1-5 for each diagram along four dimensions. The completed surveys were then to be left on the table. Note that only 11 subjects filled in this section of the survey and the last two aspects were missed by one subject (these questions were on the overleaf page). The six representations were:

Diagram 1 - MCRDR KBS (Figure 8:11)
Diagram2 - Formal Context (crosstable/decision table) (Figure 8:12)
Diagram 3 - Concept Matrix (Figure 8:13)
Diagram 4 - Concept Lattice - Line diagram (Figure 8:14)
Diagram 5 - Gaines' Visual Language (Figure 8:15)
Diagram 6 - Exception Directed Acyclic Graph (EDAG) (Figure 8:16)

A brief description of how to read each diagram was given when the diagram was shown on the overhead projector. The scores for each representation are shown individually by feature and an averaged score is given and shown in Tables 8.3(a-f). An overall rating is given which is the sum of the averaged scores. Using the overall score, the diagrams in order of preference are the EDAG, the MCRDR, the line diagram, Gaines and Shaw's Visual language, the crosstable and the concept matrix. Table 8.4 provides a summary of each representation by feature. It is noteworthy that the line diagram scored the highest for structure, which has been the claimed strength of the line diagram in this thesis. The poor score for the concept matrix (Diagram 3) was expected because the representation required more description to really understand, but there was insufficient time, and because the ability to reveal patterns which can be compared was not being evaluated here.

Diagram 1												
Pleasing appearance	2	3	2	4	5	5	2	2	4	4	5	3.45
Comprehensibility	4	3	4	4	5	5	3	2		4	4	3.80
Good Structure	-	3	4	4	-	4	3	3	4	5	3	3.66
Able to determine	-	4	5	4	-	5	5	2	4	5	4	4.20
Implications												15.11

Table 8.3(a): The results for the 11 subjects for MCRDR KBS by feature.

Diagram 2												
Pleasing appearance	1	2	2	2	1	3	2	3	4	2	2	2.18
Comprehensibility	2	5	3	1	1	4	3	4	-	2	1	2.60
Good Structure	-	4	3	4	-	3	3	3	4	3	2	3.20
Able to determine	-	2	2	1	-	3	3	4	4	3	1	2.55
Implications												10.53

Table 8.3(b): The results for the 11 subjects for the crosstable by feature.

Diagram 3												
Pleasing appearance	1	1	3	1	1	2	3	2	1	1	1	1.27
Comprehensibility	1	3	4	1	1	2	3	3	1	1	1	1.91
Good Structure	-	4	4	4	-	3	3	2	1	3	2	2.80
Able to determine	-	1	3	1	-	3	3	2	1	3	1	2.00
Implications												7.98

Table 8.3(c): The results for the 11 subjects for concept matrix by feature.

Diagram 4												
Pleasing appearance	4	3	3	4	4	3	4	3	3	3	4	3.45
Comprehensibility	3	3	4	4	4	3	4	2	3	2	3	3.18
Good Structure	-	4	5	4	-	4	5	3	3	4	3	3.88
Able to determine	-	4	3	4	-	3	4	3	3	4	3	3.44
Implications												13.95

Table 8.3(d): The results for the 11 subjects for line diagram by feature.

Diagram 5												
Pleasing appearance	1	3	4	1	4	3	4	4	1	3	3	2.80
Comprehensibility	1	4	4	1	2	2	4	3	1	1	2	2.27
Good Structure	-	3	4	4	-	4	3	3	1	5	1	3.10
Able to determine	-	4	3	1	-	4	4	3	1	2	2	2.66
Implications												10.83

Table 8.3(e): The results for the 11 subjects for the visual language by feature.

Diagram 6												
Pleasing appearance	5	5	5	4	5	3	4	4	3	4	2	4.00
Comprehensibility	5	5	5	4	5	5	4	4	3	3	2	4.09
Good Structure	-	3	5	4	-	4	4	4	4	5	1	3.77
Able to determine	-	5	4	2	-	4	5	4	4	4	4	4.00
Implications												15.86

Table 8.3(f): The results for the 11 subjects for the EDAG by feature.

	MCRDR	Crosstable	Concept Matrix	Line Diag	Visual Lang.	EDAG
Pleasing appearance	3.45	2.18	1.27	3.45	2.80	4:00
Comprehensibility	3.80	2.60	1.91	3.18	2.27	4:09
Good Structure	3.66	3.20	2.80	3.88	3.10	3.77
Determine to	4.20	2.55	2.00	3.44	2.66	4:00
Implications	15.11	10.45	7.98	13.95	10.83	15.86

Table 8.4: The combined averaged results for the 11 subjects.

8.2.1.5 Discussion and Conclusion

A comprehensive discussion on the use and value of visual languages and representations is offered in Kremer (1998). Kremer (1998) remarks that visual representations have been found to offer noteworthy advantages over linear textual representations. He cites Nosek and Roth (1990) who have found semantic networks to be more "computationally efficient" (comprehensible) than predicate logic. Visual representations have also been useful for visual programming, decision making, knowledge representation, database queries and more. Even the use of indentation in programs has found to be an important aid in program comprehensibility (Smith 1977). As shown in the small survey described in this chapter, reading graphical representations is not always easy and can be considered an "acquired skill" (Petre and Green 1993). In the Petre and Green study it was found that their subjects could use the textual representations more quickly but the graphical representations offered a much richer representation once the subject had passed through the learning curve. Kremer (1998) lists the arguments of Smith (1977) which describe the psychological benefits of visual languages for programs and data. These arguments include:

- abstract reasoning is pictorial and thought may involve manipulation of images,
- linear verbal language only provides one dimension but visual reasoning uses two or three dimensions;
- visual images assist "chunking" and short-term memory;
- images tend to be analogical.

Charles Sanders Peirce, who developed a visual logic language known as *existential graphs* on which Sowa (1984) based his Conceptual Graphs, states:"

> *I do not think I ever reflect in words: I employ visual diagrams, firstly because this way of thinking is my natural language of self-communion, and secondly, because I am convinced that it is the best system for the purpose {Ms 620, p. 8] (Roberts 1973, p.126).*

Kremer (1998) also makes the point that to date there is little hard evidence of the benefit of visual representations and that most interest is due to an intuitive belief that visual representations offer many benefits over textual or linear representations. The survey reported in this section has attempted to gain some evidence of the value of the line diagram representation over the linear textual rule trace. As noted by Kremer (1998) and evidenced in Petre and Greens's (1993) study, use of a visual language requires time and effort and this makes evaluation of the line diagram with novices a difficult task. Despite the difficulty, in this section domain and line diagram novices have been used to interpret and evaluate line diagrams from a number of viewpoints and the line diagrams have been compared with other representations and rule traces. The number of subjects does not warrant the application of statistical techniques to determine if the results are significant and at what level of confidence.

In summary, the survey found that 10 out of the 12 subjects were able to learn to read a line diagram within a few minutes, that the line diagram was easier and faster to use than a rule trace in answering questions about the knowledge base and that even novices could reason about the knowledge using the line diagram allowing the tool to be used for such purposes as hypothesis testing and tutoring. Given the time restriction for conducting the survey the results can be considered promising and generally supportive of the claim that the line diagram is a reasonable representation of the knowledge base and that they can also be used for explanation and teaching.

8.2.2 An evaluation by experts

In the previous section we can see that the line diagram can be used to gain some understanding of the domain. The purpose of evaluation with experts was to see if the type of knowledge that could be discovered was accurate and useful. The four domains introduced in Chapter Five have been used in this evaluation. Thus, four different experts have been involved in these evaluations. Access to all four experts has been difficult mostly due to availability and geographical separation and meant that the evaluations had to be conducted in one meeting or discussion. In each subsection the nature of the interaction is described.

8.2.2.1 The Blood Gases Domain

The 105 Blood Gases KBS was developed by a pathology expert over two years prior to this evaluative study. Since development of the KBS, the expert had moved to Perth, some 3,000 kilometres from the author. The opportunity for a meeting occurred while the author attended the AI'97 conference which was being held in Perth. This evaluation was the first performed and to a large extent the nature of the outcome and the best presentation to the expert was unclear. As preparation for the meeting 11 line diagrams were drawn based on selections the novice, myself, thought might be useful. In a real teaching system the student would be more directed and probably be given questions to pursue or have questions that had arisen from other learning situations such as reading or a lecture. I made a brief synopsis of each diagram in one to ten sentences. I then removed any reference to particular parts of the diagram to be left with one and a half pages of written description of what I had learnt about the domain. This one and a half pages was given to the expert for comment. This format was also used in the classification domains - Lotus and Sisyphus III.

The blood gases KBS I had been given did not include a description of the conclusion codes. I did not worry about this because I thought the expert would know what the codes stood for since he had created them. I also thought that since the description often includes the reasons for the conclusion I would be less biased or influenced by the description and my ability to learn would be based on the rules only. The observations I made and handed to the expert are shown below in Times New Roman single-spaced in 10-pt. The verbal comments of the pathologist are interleaved with each different observation and are shown in a different font (Arial). I have added further comment regarding the expert's response in italics. Each excerpt has been numbered for easier reference.

1. For the %MC002 conclusion there appears to be an inverse relationship between the value of BLOOD_PH and BLOOD_BIC and the two appear to be causally connected. For this conclusion when BLOOD_PH is HIGH, BLOOD_BIC is LOW. Conversely, when BLOOD_PH is LOW, BLOOD_BIC is HIGH. Also if the BLOOD_PH level is increasing then the BLOOD_BIC level must be decreasing. It also appears that if the BLOOD_PH level is NORMAL then the value of BLOOD_PC02 needs to be checked. If BLOOD_PC02 is high then the %MC002 conclusion should be given but if it is low then the value of BLOOD_BIC must also be low to receive this conclusion.

The expert could not remember what the conclusions stood for which made it difficult to say if my analysis was correct. The analysis I gave seemed unusual and he was trying to make sense of it. After a while, he thought that %MC002 was metabollic compensation. This was a weird conclusion that could occur in two quite different circumstances and he had previously decided that if he built the KBS again he would have treated this conclusion differently. This was why it was using High Blood_Ph with Low Blood_Bic or Low Blood_Ph with High Blood_Bic.

It was unexpected that the expert did not remember what the codes were for. I had assumed that he had not bothered with a description of them because he knew them so well. The comment that I had picked a weird conclusion was interesting because it shows the value of the tool for validation in detecting anomalous parts of the KBS. I had not been aware of the anomaly. The expert seemed reasonably impressed with my being able to analyse the rules and said I was a "human induction engine". The expert suggested that I was making life too hard for myself (and him probably) by not knowing the meaning of the conclusion codes. He suggested that I try again to learn using the conclusion descriptions. However, the opportunity to meet again did not seem likely so we carried on with the material available. When I later found the conclusion codes descriptions, it turned out that %MC002 stood for "Metabolic compensation.2" so that the comments by the expert was appropriate.

2. From the previous discussion, I was interested in finding out what conclusion would be given if BLOOD_PH was normal, BLOOD_PC02 was LOW and BLOOD_BIC was HIGH. I did a search on this combination but it was not contained in any rule. I then did a search on HIGH(BLOOD_BIC)=TRUE but there were no rules in addition to rule 15 which only considered LOW(BLOOD_PH)=TRUE and HIGH(BLOOD_BIC=TRUE and gave the %MC002 conclusion. This made me consider what was the effect of LOW(BLOOD_PC02)=TRUE. So I have entered that as the selection.

The expert thought that it was unusual that there would not be a rule to cover alternative situations when Blood_Bic was high.

It can also be argued that the tool was useful for determining missing knowledge.

3. It can be seen that if LOW(BLOOD_PC02)=TRUE the next most important attribute to consider is BLOOD_PH. If BLOOD_PH is abnormal it should be given the conclusion %RK002 or %RC002 depending on whether the value is HIGH or LOW, respectively. A number of rules which are part of the %MC conclusion family and include the conditions NORMAL(BLOOD_PH)=TRUE and LOW(BLOOD_BIC)= TRUE. The actual conclusion assigned depends on the actual PH values and if BLOOD_BIC is increasing.

The expert agreed with this paragraph and gave it two ticks.

Selected conclusion family %RK. The key features of the case to consider for this class of conclusions are BLOOD_PH and BLOOD-PC02. All rules are concerned with BLOOD_PH being high or having been high (over 7.45) except for rule 30 which concludes %RK002 when BLOOD_PH is NORMAL, BLOOD_BIC is LOW and BLOOD_PC02 is LOW. As was noted earlier there appears to be an inverse causal relationship between BLOOD_BIC and BLOOD_PH and the fact that BLOOD_BIC is LOW may indicate that BLOOD_PH is expected to become HIGH. If BLOOD_PH has been high at any stage then an increase in BIC and decrease in PH and BIC will give the conclusion %RK004.

The expert thought this paragraph was okay but had a problem with the last sentence. While he was seeing a patient, I looked again at the diagram and what I had written and saw that “then an increase in BIC” in the last sentence should have read “then an increase in PC02”. When I changed this the expert was satisfied but was not completely sure as he could not remember what the meaning of conclusion code %RK004 which I later found meant “Resolving respiratory alkalosis”.

The expert said that I had a misunderstanding of the relationship between Blood-Bic and Blood_PH. He commented that if you told an expert that Blood_Ph was low they would say the Blood_Bic should be low. The conclusion %MC002 for metabollic compensation resulted in the situation being reversed and unusual so it was misleading me into thinking that this is a common occurrence.

This was useful in identifying and correcting a misconception that I had. I had formed a hypothesis of an inverse causal relationship between Blood_Ph and Blood_Bic but this was generally not so and had only occurred because of an anomalous situation.

5. The conclusion family %MK was selected. The attributes to consider for these conclusions are BLOOD_PH, BLOOD_BIC and BLOOD_PC02. The two rules for %MK003 use the conditions DECR(BLOOD_PH)=TRUE and MAX(BLOOD_BIC)>=27. If BLOOD_BIC is decreasing then the conclusion %MK003 should be given otherwise the value of BLOOD_PC02 should be considered. The concept “HIGH(BLOOD_PH)=TRUE and NORMAL(BLOOD_PC02)=TRUE” may represent a higher concept that could be labelled as indicative of a particular type of patient condition.

The expert said that HIGH(BLOOD_PH)=TRUE and NORMAL(BLOOD_PC02)=TRUE did represent a higher concept which could be described as there is something wrong that needs further investigation.

6. The values used by the %RK and %MK conclusion families are similar and indicates that there is a close relationship between these two families

The expert pointed out that the use of K in the classification code meant that the PH was high. Whether the conclusion was R or M depended on a refinement of the initial rule/analysis based on Ph.

It was therefore natural that these two conclusion families were related since they were both based on a High PH. I had not realised such a naming convention. This also takes us back to point 4 where it was noted that %RK002 was the only rule that didn't require BLOOD_PH to be HIGH when all other similar rules did have this A-V pair. However, the coding used by the expert for the conclusion code has implied that this conclusion typically applies when BLOOD_PH is HIGH even though it was not present in the case that prompted the rule to be added.

7. Selected clause MIN(BLOOD_PO2). Only the rules which conclude %OX006 use this attribute and function. The key to this conclusion is a low BLOOD_PO2 level. The only other attribute that is considered is seen in rule 42 where the condition AGE<=70 is included. MIN(BLOOD_P02) <=80 is the highest value given, but the values range from 59 to 80 with 6 different values used.

The expert asked me if I was able to determine the purpose of the age attribute. I looked for a while at the diagram but not knowing what the conclusion codes were and without the tool to make various inquiries I could not determine the effect of age.

As the expert pointed out, if I had the meanings of the code I could have seen how age changed the conclusion given. After returning to Sydney I tried to see if I could answer this question. Since age was only considered in conjunction with the attribute BLOOD-PO2 I selected all the rules that used BLOOD_PO2 which gave me all of the rules for the %OX conclusion family. Looking at the description of the conclusion codes I could see that the %OX conclusion family all dealt with conditions related to oxygen and thus assumed that BLOOD_PO2 is some type of measurement of oxygen content in the blood. My observations regarding age are given in point 11 and are followed by the experts comments.

8. The %AB conclusion family first considers whether NORMAL(BLOOD_PC02)=TRUE and then whether BLOOD_PH is NORMAL or HIGH. It appears that %AB001 may indicate that everything or a certain condition is normal and %AB002 may indicated that the condition is abnormal.

The expert said this was plausible but he couldn't confirm this without the conclusion code descriptions.

When I later found the code descriptions it turned out that my assumptions were correct. %AB001 stood for "No disturbance of acid base status" and %AB002 meant "Mild respiratory alkalosis".

At a later date I sent the following email to the expert to clarify a couple of the outstanding issues raised.

This first comment concerns point four and the conclusion %RK004. The expert's comment are again interleaved.

9. If BLOOD_PH has been high at any stage then an increase in PC02 and decrease in PH and BIC will give the conclusion %RK004-""Resolving respiratory alkalosis".

Yes that seems right. Although I would qualify it that if the PH has been high AND THE PCO2 WAS LOW AT THE SAME TIME and THE BIC WAS NOT HIGH (at least this would be the usual case)

10. When I saw you, you asked me if I could see how the age affects the conclusion given. The only attribute that was affected by the age attribute was BLOOD_PO2. I therefore selected all the rules that used BLOOD_PO2 which gave me all of the rules for the %OX conclusion family. Looking at the description of the conclusion codes I could see that the %OX conclusion family all dealt with conditions related to oxygen and thus assumed that BLOOD_PO2 is some type of measurement of oxygen content in the blood.

Yes.

11. If BLOOD_P02 was normal the conclusion %OX005-"Normal Oxygenation" should be given except where MIN(BLOOD_PO2)<=75 (rule 37) or <=74 (rule 42) which results in the conclusion %OX006-"resolving hypoxaemia". However, if MIN(BLOOD_PO2) is between 74 and 80 then the age must be taken into consideration. If the age is <=70 then the conclusion %OX007-"hypoxaemia for this age" is given. It appears that even though the BLOOD_PO2 level is within the normal range it is a bit low for patients under the age of 71.

Exactly right. Would you like a job as an anaesthetist?

The comments of the expert in this domain provided some support that the line diagrams were useful for learning about the blood gases domain. Terminology, including the absence of conclusion code descriptions, appeared to be my main impediment to learning and it seems that in such a domain some initial training at least regarding the domain terms would be needed to learn further from the KBS. The value of knowledge learnt is not clear although it is obvious that the line diagrams provided a focal point for discussion which otherwise would not have existed.

8.2.2.2 The LOTUS Domain

The KBS for this domain were kindly supplied by Dr. Zvi Hochman, an agricultural scientist working for CSIRO in Toowoomba. I had had no discussions with Dr. Hochman regarding Lotus but I had met him on a couple of occasions and he agreed to review my observations. Since we were geographically separated I sent the following email message to him which is shown in Times New Roman - 10 point. His comments are shown in Arial 12 and my remarks on his comments are shown italicised in *Times New Roman 12 point*. I have included some introductory remarks for Dr Hochman since we had not had previous opportunities to discuss the background to this work. In the discussion on comparing viewpoints I have removed the names of the individual KBS and replaced them with more generic labels, that is, Lotus 1-4. Dr. Hochman had requested previously that the identities of the experts used in the four KBS remain unknown.

The following comments are observations made from line diagrams which show the hierarchical structure between concepts in a knowledge base. The key benefits of this structure are the higher level concepts and abstractions that are uncovered from the lower level rules. In the Lotus domain there are not many higher level concepts as the rules tend to be self-contained with few intersections of rule conditions with other rules. Using different views of the knowledge, however, we are able to ask questions such as: how does rainfall or PH affect the decision; what other conclusions are close to a specified conclusion and what are the differences between the different KB.

The reason for asking you to add your comment to my observations is to offer some evaluation of the claim that I can use line diagrams to assist me in learning about a domain. Obviously much of what is stated below can be found in the rule traces but these are difficult to structure and relationships between concepts are not easily found. I do not expect to have uncovered any new insights into Lotus but would like to know if my comments appear to be consistent with the domain knowledge. Put as much or as little comment as you feel necessary along the way and perhaps add at the end your comments regarding this whole exercise.

If you would like to pose a question regarding this domain which you feel I should be able to uncover in the KBS then please do so.

I begin with comments for the combined Lotus KBS with some comments regarding the individual KBS at the end. I must add that the conclusion code long descriptions provided in the combined KBS were so comprehensive I could possibly just have used them. In my analysis I did not look at these descriptions but later looked at them and generally found them to echo what I had found. However, the fact that I could get such information without the use of the diagrams is not a strong point in terms of internal validity of this evaluation.

OBSERVATIONS

1. The important feature in choosing the conclusion %SALIN- "Maku OK. Consider Strawberry Clover" is the condition "waterlog = yes" as both rules which give this conclusion share this condition. Additional considerations in choosing this conclusion are the soil must be saline OR the PH must not be LOW.

2. The importance of WATERLOG=YES prompts me to consider what other rules use this condition. Using this as my basis of selection I obtain an additional rule - rule number 21 - which gives a null conclusion if the RAINFALL>=1200 and the FERTILITY=LOW. This would suggest that strawberry clover does not do well if the ground is highly waterlogged and the ground only has low fertility.

3. The conclusion - %SHALW - Maku is unlikely to persist through a prolonged dry period" depends on two different sets of criteria as shown in rules 12 and 22. Rule 22 considers that the rainfall must be below 1000mm and the soil depth below 30mm. Rule 12 looks at a similar feature to low rainfall by specifying that the ground should not be waterlogged. However rather than considering soil depth rule 12 considers the soil PH, which should be low. As is shown in the conclusion description the key to this conclusion is the amount of moisture in the soil which is identified in both rules.

4. Next we look further at the impact of LOW-PH on the conclusion. When PH is low a number of conclusions are possible. Regardless of other features, low PH is an indicator that the soil is suitable for MAKU- with the conclusion %MAKUU -Maku lotus is recommended" being given in rule 2. Other recommendations are also offered when additional conditions exist. If the soil is not waterlogged then the conclusion "%SHARN- Maku OK. Consider Sharnae or Goldie" should be given which indicates that Sharnae or Goldie are other possibilities in dry conditions. A further exception to this is the situation where the soil is not waterlogged but there is still plenty of rain. If in addition the soil is not saline then the recommendation "%KENYA- Lotus is suitable but you may consider Kenya Clover too" may be another alternative crop to plant. From the long description offered for this conclusion - "In high rainfall and on the better soils Kenya clover may provide a higher quality pasture" it appears that if these conditions exist it would be preferable to plant Kenya clover over Sharnae, Goldie or Maku. Where the soil is not waterlogged but the rainfall is <= to 1000mm and the soil-depth is <=30 Kenya would not be recommended and a warning regarding the dry conditions is issued in the conclusion %SHALW- "Maku is unlikely to persist through a

prolonged dry period".Alternatively if the soil is waterlogged (as indicated by Flooding>=3), Maku will not survive and the conclusion %FLOOD- Lotus cannot tolerate 3 or more days of flooding in summer" will be given.

5. It would therefore appear that if the ideal conditions of low-ph, not-waterlogged, adequate-rainfall and non-saline-soil all existed Kenya clover should be planted otherwise plant Maku., Sharnae or Goldie but watch out for either too much or too little rain.

6. If the soil PH is not low and there is plenty of water the recommendation %SALIN - Maku OK. Consider Strawberry Clover" suggests that Maku and Strawberry Clover are both suitable. Alternatively, regardless of the PH, if the rainfall is adequate, but not resulting in the soil being waterlogged, it would be worth considering white clover. If the soil-fertility is low, fertilizer should be added for white clover. One might also assume that if the soil FERTILITY was not LOW then a different conclusion which also recommended to consider white clover without the need for fertiliser should exist. If the soil is waterlogged then the options offered are MAKU or Strawberry Clover if the soil is either saline or the PH is low. If in addition to being waterlogged, the rainfall is high and the soil has low-fertility then no recommendation is given.

A number of attributes such as (RYEGRASS>=15); (SOWING_DEPTH>=15); (SOW_MOISTURE<=75);
(WINTER=TRUE); (SPRING=TRUE); (SCARABS=YES); (RENOVATE=YES); (GRAZ_MANAG=NO)
(TROPIC_GRASS>=3); (SPRING=TRUE); (SUMMER=TRUE)
are used to provide a conclusion but there is no interaction between attributes and the conclusions are found by going directly from the attribute to the conclusion. These types of facts are not terribly interesting. To increase the depth of knowledge in the KBS, it may be worthwhile review the factors that affect these various conclusions to see if some factors have been missed and if relationships between the rule conditions and/or conclusions exist.

Comparing Viewpoints

The Lotus1 viewpoint appears to be the foundation used by the other three viewpoints since it is contained within all other three viewpoints.

The additional concepts found in each of the viewpoints are:

Lotus2
If Soil-type=light then %Light-Maku unsuited to light quick drying soil
If Rainfall=yes then %Maku2-Maku should persist
If Flooding>=3 then %Flood-Flood. Long description =Maku is killed by 3 or more days of flooding if water is still and high temp

Lotus3
If summer-flood >=3 then %Maku3-Maku Ok,but flooding with persistence
If Lotus-Rate>=3 then %Maku4-Maku Ok only at higher sowing rates
If scarabs=yes then %SCARA-Persistence of Maku limited by Scarabs

Lotus4
If moisture<=1250 then %Maku5-Maku at rainfalls above 1250mm
If companion-grass>-1.5 then %SUMM1-Keep grass down in summer. Long description-grass competition is excessive for Lotus - keep grass down in summer

For the purpose of dicussion the following concepts appear related and should be discussed using the cases to assist in reconciliation of differences.

CONCEPTS RELATED TO MOISTURE CONTENT
L2- If Rainfall=yes then %Maku2-Maku should persist
L2-If Flooding>=3 then %Flood-Flood. Long description =Maku is killed by 3 or more days of flooding if water is still and high temp

L3 -If summer-flood >=3 then %Maku3-Maku Ok,but flooding with persistence
L4 -If moisture<=1250 then %Maku5-Maku at rainfalls above 1250mm - This concept appears to be contradictory. The rule condition indicates that the moisture must be below 1250mm but the conclusion comment indicates the situation occurs above 1250mm. If this is an error in the rule condition then these four concepts appear to all consider the affect of rainfall on the persistence of Maku. Resolution of differences in terminology and concepts should be considered.

CONCEPTS RELATED TO COMPETITION FROM WEEDS
L3 -If scarabs=yes then %SCARA-Persistence of Maku limited by Scarabs
L4-If companion-grass>-1.5 then %SUMM1-Keep grass down in summer. Long description-grass competition is excessive for Lotus - keep grass down in summer

CONCEPTS RELATED TO SOIL TYPE
L2- If Soil-type=light then %Light-Maku unsuited to light quick drying soil

Your discussion on this domain reads as though you have some knowledge of it (which I think I am safe in assuming you did not possess before examining the KBs)! You make only one nonsensical statement. It is about a situation in which soil depth is <0. If there is no soil, then soil depth=0 for soil depth to be less than 0 it would have to antimatter or something....A less important error, but one that would make you look silly with clients, was your assumption that scarabs are weeds (they are insects). The common thread here is that weeds and scarabs have an impact on the population/survival of lotus. High densities of ryegrass and ungrazed (poorly managed) tropical grasses have a similar effect.

It was pleasing that it appeared I had some knowledge of this domain. I could not find the nonsensicaI statement mentioned when I checked back to the email I had sent. The above transcript is identical to what was sent. I can only assume that some of the line breaks had messed up the flow of text and that where I refer in point 3 to "soil depth below 30mm" has been segmented so that it reads as 0mm. My assumption that scarabs were weeds instead of insects is amusing and demonstrates my lack of expertise in this domain besides the knowledge I was provided through the KBS. As noted by the expert, at a higher conceptual level the two can be viewed similarly as weeds and pests both pose threats to the survival of Lotus.

Your observations in the last para before "Comparing Viewpoints" are valid. This is to be expected as the rules are being added as expertise is gained. Deep knowledge should follow.

The author believes that the MCRDR/FCA tool can assist in developing and uncovering this deeper knowledge but this belief is not tested here.

On the moisture issue the situation is that lotus requires reasonably wet conditions. This can be mitigated towards drier conditions if the soil has a reasonable storage capacity (depth and light soils are used as proxies for this concept in some rules). Unlike many other species (but in common with strawberry clover lotus can tolerate waterlogged conditions. However conditions must not get so wet that there is a high risk that lotus ends up flooded (still water) for more than 3 days in hot conditions. I think you got close to covering this one!

I also believe my analysis supports these statements by the expert. In fact in the first two points I have already identified the tolerance of Maku (and strawberry clover) for waterlogged soil. In point four I note that while Maku tolerates waterlogged conditions rainfall for longer than 3 days will not be tolerated particularly on hot summer days.

The results in the Lotus domain were more encouraging that those obtained in the blood gases domain. It had only been my goal to determine if I was able to gain an accurate understanding of the domain knowledge and it appears that this was achieved.

8.2.2.3 The SISYPHUS III Geology Domain

The geological experts associated with Sisyphus III KA material were not available for consultation. It was therefore necessary to find another geological expert who could comment on what I had learnt about igneous rocks from the MCRDR KBS that was built. An inquiry was made with the Head of the School of Applied Geology to find if an expert could be found. At the meeting with the geological expert a verbal introduction to the purpose of the meeting was given and the following text was shown to the expert. As in the previous evaluations the expert's comments are shown in Arial 12 and my remarks on his comments are shown italicised in *Times New Roman 12 point*. The numbering system relates to the numbers on the diagrams I had worked from and were not shown to the expert. I used them on my copy of the observations in case I needed to refer to a diagram. They are shown here for reference purposes.

1. There appears to be a typical set of values for the attributes calcium, potassium, iron and iron. When calcium=high, potassium=low,iron=rich. When calcium=medium, potassium=medium, iron=low. When calcium=low, potassium=high and iron= inter-mediate.

3. If a rock is granodiorite, granite or microgranite the COLOUR of the rock will be L_GREEN_WHITE and the value of OLIVINE will be NEVER. These two conditions provide a good starting point and assist the expert in pruning the search space for rock that can be classified as a type of granite.The three types of granite are further distinguished from one another by other attributes. Granodiorite shares the attribute GRAIN_SIZE=COARSE with granite but that they differ in their QUARTZ content, with granodiorite having QUARTZ=SATURATED and granite having QUARTZ=OVERSATURATED. Granite is similar to microgranite in the QUARTZ content but differs in GRAIN_SIZE which is MEDIUM. Granite also takes into account the FELDSPAR content and tbe rule for microgranite includes PYROXENE=POOR. Is there a relationship between PYROXENE and FELSPAR ? If there is these two rules may be considering the same feature of a rock.

8. The density of the rock is determined by the content of dark minerals.

4. Density of the rock is dense if percentage of dark minerals is from 56% to 90%. We reach a conclusion that the density of the rock is very dense if percentage of dark minerals is greater than 90%. A fuller description of the conclusion code reveals that dark minerals includes fe-ag,hornblende,labradorite and opaques.

5. We reach the conclusion that the density of the rock is very light if percentage of dark minerals is less than 30% . We reach a conclusion that the density of the rock is light or medium if percentage of dark minerals from 30% to 56%.

13. We can see the four conclusions %DE000, %DE001, %LI000 and %LI001 are mutually exclusive. As the four attribute-values cover the full range of possible values every rock that has had the DARK_MINERAL content identified can be assigned one of these four conclusions.

6. Grain size is the key factor when considering rock formation. If grain size is coarse then the form of formation of the rock is major intrusions or small slivers. If grain size of the rock is medium then its form of formation is small intrusives. If grain size of the rock is fine then its form of formation is extrusive flows (larva).

7. All rules that conclude Basic Rock-%BS000 have QUARTZ=UNDERSATURATED. The SILICA content is then the next most important feature as it is used in all of the rules with values ranging from very little to intermediate.

9. All rules that conclude Acidic Rock-%AC000 have SILICA=VERY_HIGH. The QUARTZ content is then the next most important feature as it is used in all of the rules with values oversaturated or saturated.
10. From looking at the %AC000 and %BS000 it appears that these two conclusions are mutually exclusive and based on alternate values for the same attributes. Acidic rock have a high silica content. Quartz may be oversaturated or saturated. If saturated the colour tends to be light-green-white. Basic rocks have medium to very little silica content and quartz is undersaturated. The colour may tend towards grey.

11. The closeness of Basalt to Dacite was considered. These two rocks were picked at random as an example of how two or more rocks can be compared. The student would probably be more directed and pick two rocks of particular interest that may be similar but which need to be diffentiated..

    Both rocks have GRAIN_SIZE=FINE. The other distinguishing features are unrelated except for colour. If they had been distinguished using the same attributes then the differences would be alignable and can be seen as more meaningful. The diagram prompts the question - What other rocks have a fine grain_size and may be similar to these two rocks ?

12. As a follow on from 11, the condition (GRAIN_SIZE=FINE) was selected. Firstly it can be seen that a fine grain-size is sufficient to support a number of conclusions - the rock is volcanic (%VC000), extrusive (%EX000) and (%MJ000). Five specific rocks use this conclusion and can be identified based on a number of other features. The rocks are Basalt, Dacite, Rhyolite, Andesite and Trachyte. The attributes COLOUR, OLIVINE and QUARTZ are sufficient to identify two of these rocks (Andesite and Dacite). The additional attribute SILICA classifies Basalt and Trachyte and FELDSPAR is used to classify Rhyolite.

Regarding point 8 the expert commented that it is also necessary to consider the content of light minerals.

Information about light mineral was not contained in the KBS. The identification of missing knowledge is an important outcome of using the line diagrams for learning.

The expert found the comments reasonably correct but that the knowledge only covered igneous rocks. He said that it was well accepted within the geology community that all igneous rocks can be classified based on grainsize and silica (quartz) content. It appeared to him that the KBS had simply developed rules from a firm set of rules accepted worldwide. He felt that classification of igneous rocks was a stable field of knowledge and straightforward and didn't involve too much variation.

These comments seemed somewhat inconsistent with the impression of the domain I had gathered from the KA material due to the inconsistencies between sources and from the findings of others such as Jansen, Schreiber and Wielinga (1998). However, this expert *also stated that if a particular rock was shown to four different experts there would be at least five classifications offered. It appears that much of the problem is in the perception of different attributes and their values by different experts. Since rocks are potentially comprised of many types of minerals a classification of a particular rock is difficult even though there may be consensus at a higher level of abstraction on the important features in classifying different types of igneous rocks.*

What I had learnt he considered to be what a geology student would learn in their first year. What I had told him was simply the accepted rules and classification system for igneous rocks. He commented that at this higher level if there was disagreement between experts it was because they had forgotten their first year training or were trying to change the accepted rules to fit their own research goals. Since I had not uncovered anything new (to him that is) about the domain his comment was also “so what”.

The expert appeared disappointed that he was unable to say that I had learnt something exciting. I was pleased that he considered what I had learnt was what a first year geology student would be taught. I did not expect to uncover something unknown about the domain but hoped that as a novice I would be able to gain some understanding of what was already known by experts.

The expert commented that my observations were a “Reliable reflection of accepted rules for igneous rocks”.

I thought this comment was worth quoting since I felt it provided confirmation of my ability to learn about the igneous rock domain from the line diagrams.

He asked if I had any questions and I asked him if there was a relationship between feldspar, plagioclase and orthoclase. He explained that if the mineral was calcium feldspar it belongs towards the orthoclase end and if the mineral was potassium feldspar it belongs towards the plagioclase end. This explained why some rocks were classified using one or the other.

It was interesting to clarify some of the issues that I had pondered when looking at the relationships between various concepts. It seems if the line diagrams could be used for learning in conjunction with occasional discussion with a domain expert the student could greatly increase their knowledge and that the expert would also be driven to update the knowledge as misunderstandings or missing knowledge became apparent.

Although the expert seemed somewhat unimpressed with what I had learnt I was satisfied by his comments. I think the expert was expecting that I may have uncovered something new and exciting and so repeating basic first year knowledge seemed trivial. To me this was not the case as I was hoping that what I had learnt was in keeping with accepted domain knowledge.

8.2.2.4 The Ion Chromatography Domain - a configuration task

As described in Section 5.3.4, the concepts derived by MCRDR/FCA for the ion chromatography configuration KBS revealed a flat structure with little variation. It was difficult to know if this "poor" (as opposed to rich and interesting) structure was due to the way the KE had entered the rules, the nature of the domain or the nature of the configuration task. Expert evaluation of what I had learnt from this KBS was pointless as there was little of any interest. After discussion with the supervisor (Mary Mulholland) of the PhD student who developed the KBS it appears that the structure was probably due to the way the KE developed the rules (tending to develop rules with one condition and many rules to cover a case) together with the nature of the subdomain that was covered. The set of approximately 300 cases from the full set of 4000 cases could be considered a standard subdomain of ion chromatography that showed little variation in solutions with there only being a couple of ways that various substances could be combined and was largely controlled by a few manufacturers with a limited set of possible configurations. Thus the KE's method and the nature of the cases can be seen as the reasons why the line diagrams developed are not terribly interesting. A new KBS is planned to be built to cover a wider set of cases over the next 12 months and it would be interesting to find the concepts for the new KBS. However, since this time frame is too long for this study, the question of if and how a configuration domain differs structurally to a classification domain must be left to future work.

8.2.3 Discussion of the evaluation by experts

In all of the classification domains that were evaluated by experts the results were satisfactory to excellent. In the Sisyphus III- geological and Lotus- agricultural domains the results were very encouraging and it appears from the feedback that the experts primarily agreed with my observations and felt that I had gained a reasonable understanding of the domains. The results were not so clear with Blood Gases- medical domain but these results were partly affected by the absence of conclusion code descriptions during the meeting and the terminology used was quite foreign. For example, I do not know the meaning of attributes such as BLOOD_PC02 and BLOOD_BIC describe and conclusion descriptions such as "metabollic compensation". These domain terms can be seen as basic knowledge that needs to be acquired first before a student can reason within the domain. The other two classification domains used terms that were more comprehensible such as "waterlogged=yes" and "grainsize=coarse" and the more familiar terminology may account the greater ability to learn. The results with the Ion Chromatography - configuration domain were inconclusive as it was not possible to learn much about the domain and the reasons for the lack of models or abstractions in the rules may be due to a combination of limitations associated with the cases chosen, the KE's skill or the nature of the domain. Despite the poor results in the latter domain, the overall conclusion is that the MCRDR/FCA line diagrams are useful for learning.

8.3 Chapter Summary

This chapter has sought to evaluate some of the claims made regarding MCRDR/FCA. In particular the value of the FCA line diagrams has been considered from a number of perspectives including generality for rule-based KBS, compatibility with other graphic rule-based representations, acceptability and comprehensibility by users and usefulness for explanation and learning. A number of different methods of evaluation have been offered including comparison, demonstration, a survey and expert interviews/comments. The evaluations have been subjective and have offered little opportunity for the use of statistical means of determining significance. Despite these limitations it is considered that the line diagram as used in MCRDR/FCA does satisfy the criteria of generality, compatibility, acceptability, comprehensibility and usefulness.

In the next and final chapter we return our focus to the work in this thesis as a whole considering the various contributions, limitations and possible future work.