Aimeur, E., & Ganascia, J. (1993). Reasoning with Classification in Interactive Knowledge Elicitation. 4th ASIS SIG/CR Classification Research Workshop, 19-38. doi:10.7152/acro.v4i1.12609

Reasoning with Classification in Interactive Knowledge Elicitation

Esma Aïmeur
Orstom
Laboratoire d'Informatique Appliqueé
Centre de Bondy
72 route d'Aulnay
93143 BONDY CEDEX FRANCE
aimeur@bondy.orstom.fr
or
Laforia
Université de Paris VI
Tour 46, 4 place Jussieu
75252 PARIS CEDEX 05 FRANCE
aimeur@laforia.ibp.fr

Jean Gabriel Ganascia
Laforia
Université de Paris VI
Tour 46, 4 place Jussieu
75252 PARIS CEDEX 05 FRANCE
ganascia@laforia.ibp.fr

This paper describes METIS, a system based upon a Method of Elicitation of Taxonomies used to Identify conceptual Structures. METIS is an interactive tool using a minimal core of structured descriptions, in order to acquire objects and to discriminate them. This core is progressively extended through the acquisition and recognition of a lot of objects; it represents a generalization hierarchy. The knowledge acquisition, and therefore the building of the generalization hierarchy, is guided by the taxonomy of classes associated to objects. The knowledge acquisition principle is incremental and uses the formalism of conceptual graphs.

We have applied METIS to the ichthyological field (i.e., the part of zoology dealing with fishes) as part of a research convention with U.R.2.C (i.e., research unit on environment and aquatic resources of tropical river valleys) from ORSTOM (i.e., the French cooperative research institute for the purpose of development).

1. INTRODUCTION

How do zoologists, botanists, physicians, microbiologists and pathologists proceed when they are interested in the anatomy, morphology, physiology, genetics, molecular biology, ethology and ecology of man, animals and plants? How do they gather and reassemble their results with those of their fellow researchers into a coherent set? They create a taxonomy and make use of a classification reasoning.

The classification term has two meanings which are often mingled [NAPOLI and RECHENMANN, 93]. The first one is linked to the hierarchical organization of knowledge: building of categories and organization of these categories into hierarchies. The second one is linked to the use of these hierarchical representations.

ISSN: 2324-9773

In the biological sciences, the building of a classification and therefore the description of taxa (taxonomic units: species, kinds, families), due to the diversity and the richness of the living world, has led to a great accumulation of information which is reproduced in catalogues and expressed in a language of textual description which is varied and full of different meanings [PAUGY, 86] [LEVEQUE *& al.*, 91].Taking into account these numerous data, the question is to know which is the best method for reading out, characterizing and making use of pertinent knowledge.

Moreover, one hopes that in the long term it will be possible to keep catalogues through hypermedia in order to be able on the one hand to associate the picture and the text more easily and on the other hand to get access more rapidly to the information which is of interest to us. The use of these hypermedia requires that the knowledge should be indexed by non-exhaustive structural descriptions where the redundant, ambiguous and irrelevant information will be removed. One must realize that taking into account the whole vocabulary makes it very hard to automatically compare the description of species and it makes the development of an interrogation language adapted to this vocabulary too long and too detailed.

The acquisition method chosen in this work (see § 3) consists in dividing our knowledge into two parts: one part is operational knowledge from which we build a structural model of the object. We then use this model to index the second part (supplementary knowledge). This part remains unworked on and is stored in a hypertext interface. To do this, we choose, in agreement with the expert, a minimal core of structured descriptions which are progressively refined to build the structural model and the taxonomy of classes; we proceed by means of partial elicitation because the specific taxa inherit the characteristics of the taxa directly above them in the hierarchy.

Our approach builds upon some aspects of the techniques of knowledge elicitation [BOOSE, 84], [BOOSE, 86] and of the learning tools of discriminating features [QUINLAN, 86], [GANASCIA, 87]. In fact METIS is an interactive tool which allows the vocabulary of the field to emerge gradually. Through a given classification METIS, either by learning or by asking the expert, acquires the features incrementally which discriminate between classes [AÏMEUR, 93], [AÏMEUR and GANASCIA, 93].

Work done in "Conceptual Clustering" [STEPP and MICHALSKI, 86] or in "Concept Formation" [FISHER, 87] aims at finding a natural grouping for a set of objects. Our approach, on the contrary, considers that classes are already known; it means that the position of classes in taxonomy is not questioned, but only their characterization. Then the question is to use the existing classification as a guide for the process of identifying discriminating features in order to gradually build a knowledge base where descriptions will be reliable, homogeneous and coherent and where taxonomy will be better characterized.

In fact, in METIS the classification reasoning will consist on working on a generalization hierarchy in order to recharacterize classes and to recognize afterwards an individual belonging to one of these classes. Although taxonomy is predefined, its incremental building is used as a guide to the acquisition process, and the hierarchy of the discriminating features is constantly being constructed.

ISSN: 2324-9773

Aimeur, E., & Ganascia, J. (1993). Reasoning with Classification in Interactive Knowledge Elicitation. 4th ASIS SIG/CR Classification Research Workshop, 19-38. doi:10.7152/acro.v4i1.12609

The layout of our article is the following: in section 2, the structure of the knowledge base will be explained with indications of the underlying representation formalism; following which, section 3 will describe the acquisition principle, illustrated with three examples. Section 4 will make the comparison between METIS and other systems. Finally, as a conclusion in section 5, strong points and constraints of this language acquisition method will be given.

2. KNOWLEDGE REPRESENTATION IN METIS.

In order to represent structured knowledge, we used the formalism of conceptual graphs [SOWA, 84], since it gives a proper visualization of the description of the objects.

2.1. Conceptual Graphs

A conceptual graph [NOGIER, 91] is defined as a graph with two types of nodes:

- the *concepts* represented by boxes or between brackets [...] correspond to contents of the thought and comply with the following notion: [<type>: <referent>]. The referent of the concept specifies its meaning. It gives the level of quantification of the concept, whether it is quantified explicitly by a number or qualitatively, for instance: [Person: 'Max']. In the case of a generic concept, for instance: [Person: *] or [Person], the referent is not instanced and can assume all the possible values in keeping with the type of concept;

- the *relations* represented by circles with an incoming and outgoing arrow or in parentheses:
  -> (...) -> symbolize the relations existing between the graph concepts.

Example:

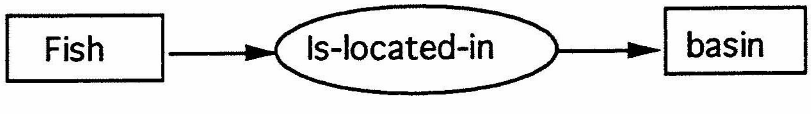

Figure 1. Example of Conceptual Graph in Ichthyology

ISSN: 2324-9773

Aimeur, E., & Ganascia, J. (1993). Reasoning with Classification in Interactive Knowledge Elicitation. 4th ASIS SIG/CR Classification Research Workshop, 19-38. doi:10.7152/acro.v4i1.12609

Limitation:
Searching for similarities between structured objects is directly bound to the application of a graph matching technique. Now our examples (fish descriptions) are totally matched; therefore there is only one possible generalization and the general basic structure of a fish is always the same. We limit ourselves to SOWA's graphs where each concept only appears once and where relations are binary [MINEAU & *al.*, 90], [GEY, 91]. This restriction corresponds exactly to the case of objects completely matched. Moreover, they are simple shapes making processing easier and enabling a clear formalization of initial restrictions without any ambiguity. In this case, generalizing two graphs just means finding out their common edges.

2.2. Structure of Knowledge Base

As described in the introduction, the acquisition method of METIS is based on the structural model, and therefore one of the constituents is the *generalization hierarchy*. The *taxonomy of classes* associated with objects is used as a framework to guide the acquisition, and the incremental building, of the generalization hierarchy.

2.2.1. A Few Basic Notions

In order to define the generalization hierarchy let us mention the following notions: *Canonical graphs, Canonical basis,* and *Lattice of concept types.*

Canonical graphs

These are meaningful graphs which represent real or possible situations in the external world. In our case, a canonical graph will be at least a triplet (concept, relation, concept) stored in a **canonical basis**. We also store sub-graphs which will be used several times in fish descriptions.

Lattice of concept types

This is a highly significant notion because it hierarchically arranges types of concepts. Therefore concepts are arranged according to the degree of generality by a partial order relation: “<“.

Example: Dorsal-fin < Part-of-Natatory-System
This means that Dorsal-fin *is a kind of* Part-of-Natatory-System

2.2.2. Structural Model

In order to define the structural model, let us explain the following notions: *Canon, Primitive Graph* (figure 2), *Generalization Hierarchy* (figure 3).

The *Canon* contains the information necessary for deriving a set of canonical graphs. Moreover, in our case it has five components because we explicitly add the set of relations which gives the following list:
- A type lattice,
- A set of individual markers,
- A set of relations,

ISSN: 2324-9773

Aimeur, E., & Ganascia, J. (1993). Reasoning with Classification in Interactive Knowledge Elicitation. 4th ASIS SIG/CR Classification Research Workshop, 19-38. doi:10.7152/acro.v4i1.12609

- A conformity relation,
- A canonical basis.

It is the canon built incrementally, which defines our description language.

The *Primitive Graph* G_0 is associated with the Primitive type in the type lattice, and is called basic structural model (see figure 2). It is common to all specialized graphs and is used as a descriptive framework for the expert's knowledge. It is essentially composed of constant structures (existing in all fishes), concepts of classes ([Family], [Kind], [Species]) and some *descriptive concepts* (see § 2.2.3) that are not discriminating on the semantic level although they are on the syntactic level (examples: [Descriptor] and [Year-of-description]). Some descriptive concepts are very relevant (example: [Basins]) for checking coherence, even if they are of no use for discrimination.

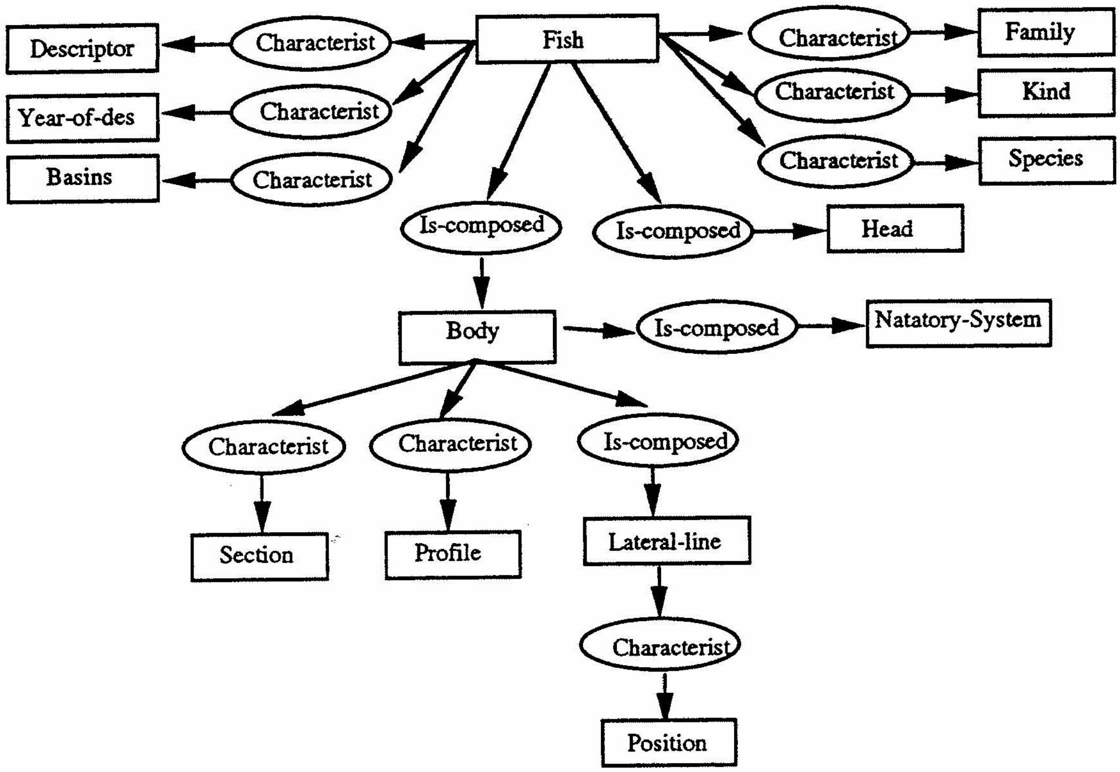

Figure 2. Primitive Graph G_0

The *Generalization Hierarchy* (GH) (figure 3) is the data structure where the graphs associated to the taxa are stored. Thus, they form a partial order structure [ELLIS, 91]. We will show that this generalization hierarchy whose construction is guided by the field taxonomy corresponds in fact to what SOWA calls *"Aristotelian hierarchy"* [SOWA, 84]. Bearing in mind the idea of Sowa let us define the following notions:

ISSN: 2324-9773

Aimeur, E., & Ganascia, J. (1993). Reasoning with Classification in Interactive Knowledge Elicitation. 4th ASIS SIG/CR Classification Research Workshop, 19-38. doi:10.7152/acro.v4i1.12609

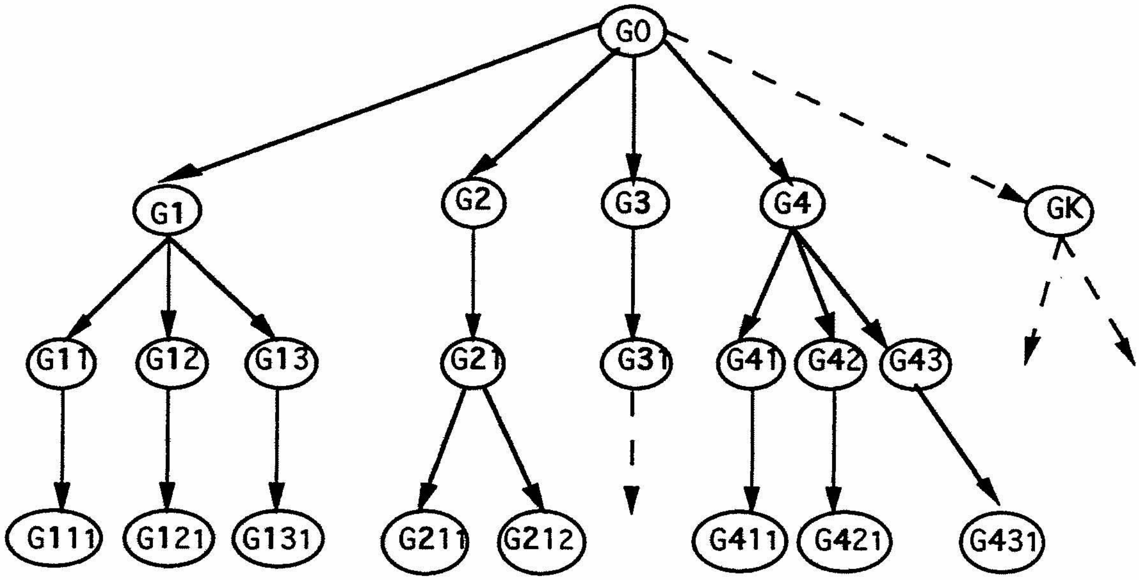

Figure 3. The Generalization Hierarchy

Definition 1. [SOWA, 84]:
"An *n-adic* abstraction, $\lambda a_1, \ldots, a_n u$ consists of a canonical graph u, called the *body*, together with a list of generic concepts $a_1, \ldots, a_n$ in u, called the *formal parameters*. The *parameter* list following λ distinguishes the formal parameters from the other concepts in u."

Thus, the abstraction (inspired by the lambda calculus) corresponds to a process in a programming language. The letter λ introduces the formal parameters and in the body u, each concept used as a formal parameter contains one of the variables in its referent field.

Example: λx,y [To-Fish]-
(Agent) -> [Fisherman: *x]
(Object) -> [Fish: *y] -> (Colour) -> [Blue]

In this example λx,y identifies [Fisherman: *x], [Fish: *y] as formal parameters. The concepts [Fish] and [Blue] are local variables in a process.

ISSN: 2324-9773

Aimeur, E., & Ganascia, J. (1993). Reasoning with Classification in Interactive Knowledge Elicitation. 4th ASIS SIG/CR Classification Research Workshop, 19-38. doi:10.7152/acro.v4i1.12609

Definition 2. [SOWA, 84]:
"A *type definition* declares that a type label *t* is defined by a monadic abstraction λau. It is written, type *t* (a) is u. The body u is called the *differentia* of t, and *type* (a) is called the *genus* of *t*. The abstraction λau may be written in the type field of any concept where the type label *t* may be written."

Definition 3. [SOWA, 84]:
"A type hierarchy *T* is said to be *Aristotelian* if every type label *t* that is a proper subtype of another type label is defined by an abstraction *t* = λau."

In our case, the generalization hierarchy is an Aristotelian hierarchy where the root corresponds to the Primitive type which is associated to the Primitive graph G_0 (common description to all fishes). Each node is a type defined by an abstraction which involves the type which is higher in the type hierarchy. Thus the Citharinus-kind type (see figure 7) is defined by an abstraction which involves the Citharinidae-family type (see figure 6), which in turn is defined on the basis of the Primitive type (see figure 2).

A type definition is composed of two parts and is associated to each type of concept: the first one is used only in the discriminating process (see § 2.2.3), the second one is involved in the building of the generalization hierarchy. The latter includes the former, as well as other structures which describe the concept, but which are not used in the discriminating process. Both parts of the definition are not fixed: sometimes in the acquisition process some structures, which are not discriminating at a given step, become discriminating afterwards, and vice versa.

The notion of type definition will be used in two operations: the *type contraction* (which consists in replacing a subgraph corresponding to a type of concept by the concept of this type) and the *type expansion* (which allows a conceptual graph to spread by using the knowledge implicitly contained in the concept to expand). Both operations will make processing easier in the acquisition process. The operation of type contraction and type expansion are symmetrical due to the aforementioned Aristotelian hierarchy given by SOWA.

Examples:
A general type common to all fish is defined through an abstraction operation on the Fish *genus*. The graph corresponding to this type is a Primitive graph:

type [Primitive (x)] *is*
[Fish: *x] -> (Characterist) -> [Family]
[Fish: *x] -> (Characterist) -> [Kind]
[Fish: *x] -> (Characterist) -> [Species]
[Fish: *x] -> (Characterist) -> [Descriptor]
[Fish: *x] -> (Characterist) -> [Year-of-description]
[Fish: *x] -> (Characterist) -> [Basins]
[Fish: *x] -> (Is-composed) -> [Head]
[Fish: *x] -> (Is-composed) -> [Body]
[Body] -> (Characterist) -> [Section]
[Body] -> (Characterist) -> [Profile]
[Body] -> (Is-composed) -> [Natatory-System]
[Body] -> (Is-composed) -> [Lateral-line]-> (Characterist) ->[Position]

ISSN: 2324-9773

Aimeur, E., & Ganascia, J. (1993). Reasoning with Classification in Interactive Knowledge Elicitation. 4th ASIS SIG/CR Classification Research Workshop, 19-38. doi:10.7152/acro.v4i1.12609

After building the graph G_1 corresponding to the Citharinidae family (see figure 6) the following type is associated:

type [Citharinidae-family (x)] *is*
[Fish: *x] -> (Characterist) -> [Family: Citharinidae]
[Fish: *x] -> (Characterist) -> [Kind]
[Fish: *x] -> (Characterist) -> [Species]
[Fish: *x] -> (Characterist) -> [Descriptor]
[Fish: *x] -> (Characterist) -> [Year-of-description]
[Fish: *x] -> (Characterist) -> [Basins]
[Fish: *x] -> (Is-composed) -> [Head]
[Fish: *x] -> (Is-composed) -> [Body]
[Body] -> (Characterist) -> [Section: Compressed]
[Body] -> (Characterist) -> [Profile: High]
[Body] -> (Is-composed) -> [Natatory-System]
[Body] -> (Is-composed) ->[Lateral-line] -> (Characterist) ->[Position: In-The-Middle]

After making an abstraction which involves the Primitive type, this type can be written after contraction as follows:

type [Citharinidae-family (x)] *is*
[Primitive: *x] -> (Characterist) -> [Family: Citharinidae]
[Body] -> (Characterist) -> [Section: Compressed]
[Body] -> (Characterist) -> [Profile: High]
[Lateral-line]-> (Characterist) ->[Position: In-The-Middle]

Similarly, the type corresponding to the Citharinus kind (see figure 7), belonging to the Citharinidae family, is:

type [Citharinus-kind (x)] *is*
[Citharinidae-family: *x] -> (Characterist) -> [Kind: Citharinus]
[Body] -> (Is-composed) -> [Scales]
[Scales] -> (Characterist) -> [Shape: Cycloid]

Finally, the type corresponding to the Citharidium kind, belonging to the Citharinidae family, is:

type [Citharidium-kind (x)] *is*
[Citharinidae-family: *x] -> (Characterist) -> [Kind: Citharidium]
[Body] -> (Is-composed) -> [Scales]
[Scales] -> (Characterist) -> [Shape: Ctenoid]

If the expansion of the Citharinus-kind type should be made, the associated graph G_{11} should be found again (see figure 7).

2.2.3. Class Taxonomy

Our aim is to get a discriminating relation for each node of the taxonomy of classes. This relation will be called a *discriminating feature*.

A discriminating feature is a conceptual graph in the form of a triplet ([Concept-*origin*] -> (relation) -> [Concept-*destination*]), where both concepts are generic.

ISSN: 2324-9773

Aimeur, E., & Ganascia, J. (1993). Reasoning with Classification in Interactive Knowledge Elicitation. 4th ASIS SIG/CR Classification Research Workshop, 19-38. doi:10.7152/acro.v4i1.12609

Example: the feature [Scales] -> (Characterist) -> [Shape] is discriminating because the shape of the scales allows us to distinguish the Citharinus kind from the Citharidium kind (see example in § 2.2.2).

The taxonomy of classes, or the *discriminating structure*, can be extracted from the generalization hierarchy. To each node is associated the discriminating part (set of features) of the definition of the type concerned by the node.

In fact, if the generalization hierarchy also allows the discrimination of taxa, the discriminating structure permits it more rapidly and more cheaply, since only discriminating and, therefore minimal parts of the type definition, are associated to its nodes.

Before dealing with the acquisition principle, we should also mention the notion of structural concepts and *descriptive concepts*. In biology, and more particularly in ichthyology, besides the fact that the object description is expressed in natural language, it is rather structured. On the one hand, we distinguish the structural concepts and relations which define the object components and the links binding them; on the other hand, the descriptive concepts and relations which describe the attributes concerning these components.

Example: [Body] -> (Is-composed) -> [Scales] (structural knowledge).
[Scales] -> (Characterist) -> [Colour: green] (descriptive knowledge).

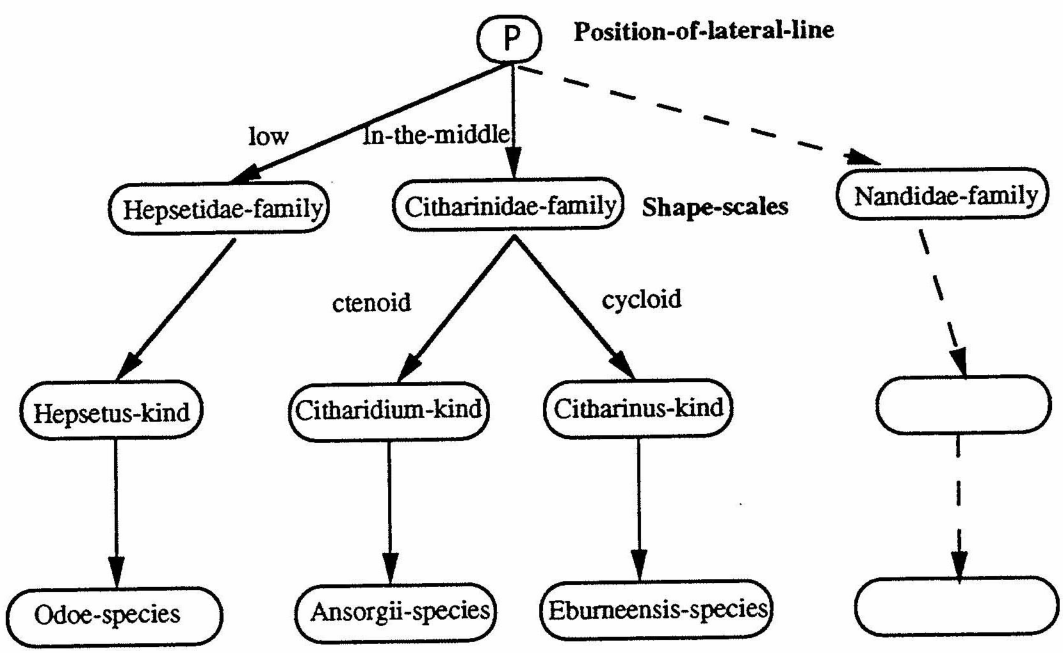

Figure 4. Simplified Discrimination Structure

ISSN: 2324-9773

Aimeur, E., & Ganascia, J. (1993). Reasoning with Classification in Interactive Knowledge Elicitation. 4th ASIS SIG/CR Classification Research Workshop, 19-38. doi:10.7152/acro.v4i1.12609

3. THE ACQUISITION PRINCIPLE OF METIS

In this section, we will deal with the acquisition principle by first giving three examples before showing the general principle (see § 3.2).

3.1. Examples

The acquisition process is illustrated in three examples. The acquisition of a species description amounts to acquiring partial descriptions which characterize the family, the kind and the species. We proceed taxon by taxon.

Initializing stage:

The first description to be acquired is: the *Citharinus eburneensis* species belonging to the *Citharinus* kind of the *Citharinidae* family. METIS gets acquainted with the species to be acquired and with its position in the taxonomy. It builds the discrimination structure (figure 5) and then asks the expert for the characteristic features of the family, of the kind and of the species, which leads METIS to build an initial generalization hierarchy, composed of G_0, G_1, G_{11} and G_{111}, which correspond respectively to the type definitions of concepts [Primitive], [Citharinidae-family], [Citharinus-kind] and [Eburneensis-species] (see § 2.2.2 figure 3). At this stage, no discrimination has to be made, because there are no discriminating features at all, i.e. for there is only one species.

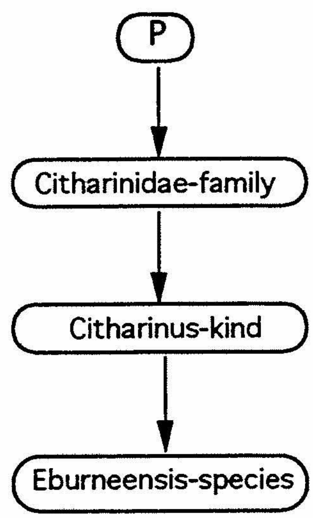

Figure 5. Discrimination Structure 1

ISSN: 2324-9773

Aimeur, E., & Ganascia, J. (1993). Reasoning with Classification in Interactive Knowledge Elicitation. 4th ASIS SIG/CR Classification Research Workshop, 19-38. doi:10.7152/acro.v4i1.12609

Remarks: in the following graphs, grey areas correspond to restrictions of concepts by specialization of their referents. Bold-faced structures are specific features of the graph being considered. They correspond to the *"differentia"* as mentioned in the type definition by abstraction according to the Aristotelian approach, while the remaining parts of the graph are inherited.

First step:

METIS asks the expert questions through the type definition of the Primitive concept in order to acquire the information about the family. If the core is not sufficient to characterize the family, METIS asks the expert to give a characteristic feature likely to discriminate it from all the other families. However, given that the name of the descriptor, the year of description, as well as the names of basins are elements of information concerning the species identification (given in the form of metaknowledge when the generic model is designed), METIS will wait for the step of the species description before asking questions. The Citharinidae-family type is defined through an abstraction which involves the Primitive type (see § 2.2.2).

Second step:

At this step, METIS asks the expert for a feature which characterizes the *"Citharinus"* kind. For, according to the expert the family description inherited from the previous step is not sufficient to characterize it, and therefore the family graph must be extended. The Citharinus-kind type is defined through an abstraction which involves the Citharinidae-family type.

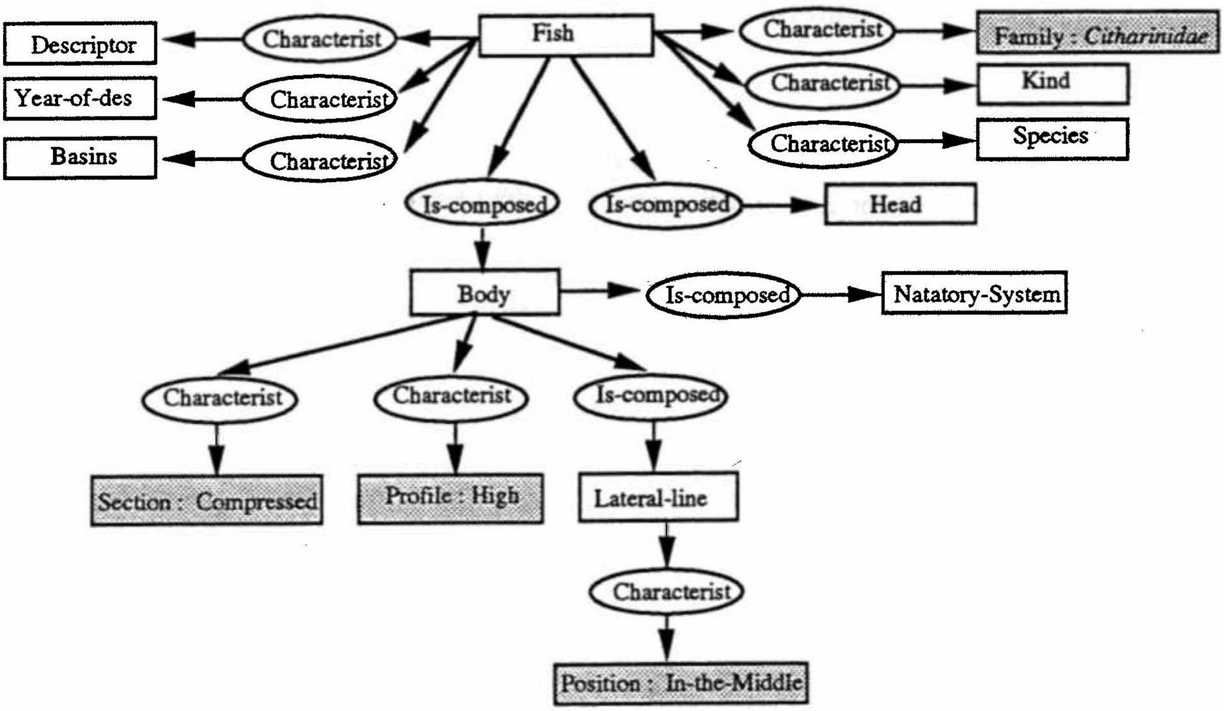

Figure 6. Graph of the Citharinidae Family (G_1)

ISSN: 2324-9773

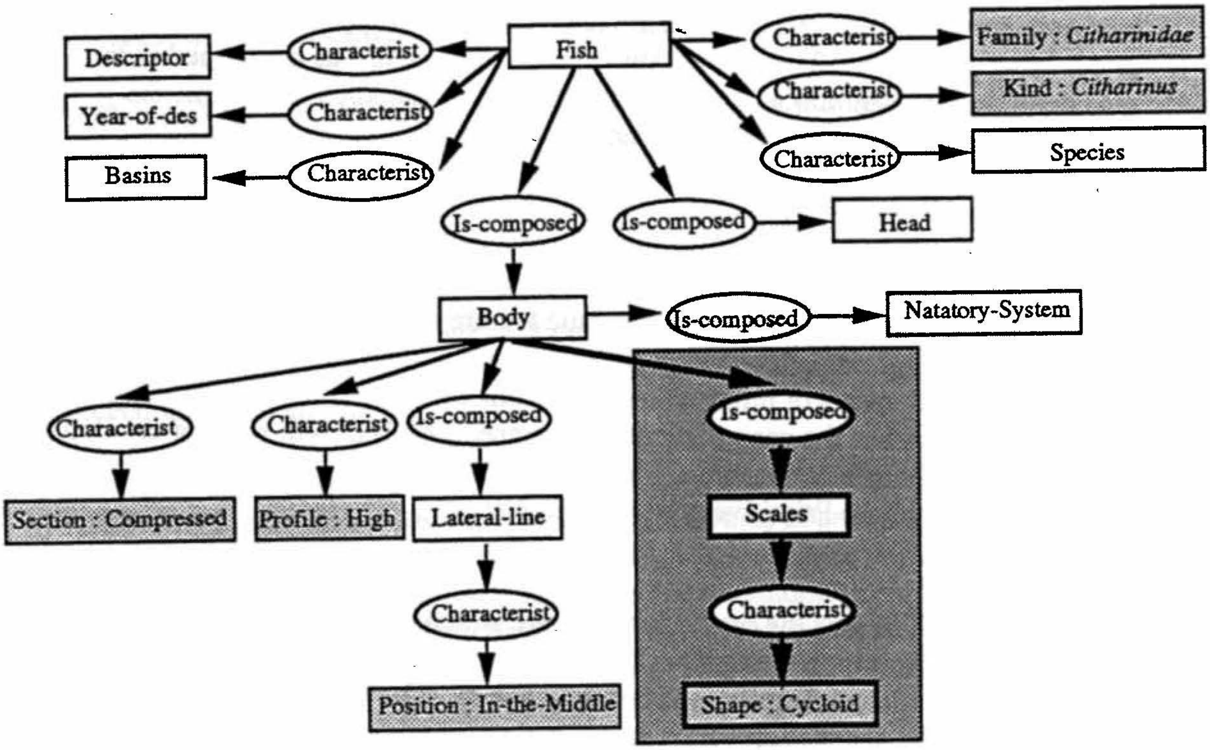

Figure 7. Graph of the Citharinus Kind (G_{11})

Third step:

In this last step, METIS will first improve the type definition of the Primitive concept by asking the expert for all the identification concepts: the name of the descriptor, the year of description and the names of basins where this species is found, and then it will acquire what characterizes this species and is likely to distinguish it from all the other species, according to the expert. Once more, an abstraction will be made to define the Eburneensis-species type.

ISSN: 2324-9773

Aimeur, E., & Ganascia, J. (1993). Reasoning with Classification in Interactive Knowledge Elicitation. 4th ASIS SIG/CR Classification Research Workshop, 19-38. doi:10.7152/acro.v4i1.12609

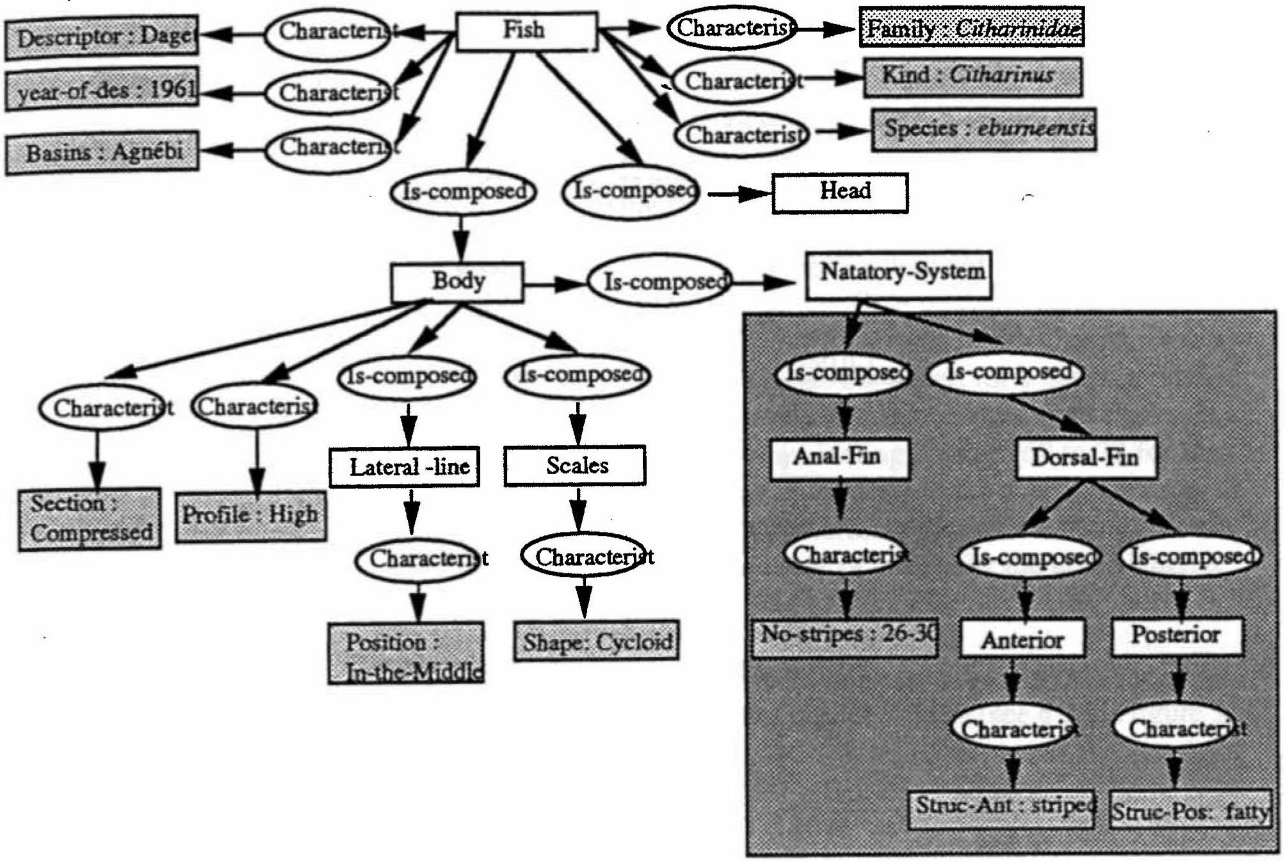

Figure 8. Graph of the Citharinus Eburneensis Species (G111)

Building stage of discriminating features:

The second description to be acquired is that of the *Citharidium ansorgii* species belonging to the *Citharidium* kind of the *Citharinidae* family.

Since the description of the family is common to that of the *Citharinus eburneensis* species, METIS will directly seek information about the shape of scales of *Citharidium ansorgii* species. The expert replies immediately that it is no longer cycloid but ctenoid. Then METIS decides that the shape of scales is a discriminating feature and it submits this conclusion to the expert. The latter gives his agreement because the feature proposed is really relevant to distinguish the two kinds (see figure 9). METIS updates the discrimination structure by initializing the discriminating part of the Citharinidae-family type definition and moves up the structure to the family graph (see figure 10). Thus, according to the referent associated to the shape concept, one advances either in the graph of the *Citharinus* kind or in that of the *Citharidium* kind (see figure 11).

The structural model is extended further when the expert introduces the body colour in the species description. Finally, METIS increases the generalization hierarchy with G12, G121 which correspond respectively to the definition of the Citharidium-kind type and the Ansorgii-species type (see § 2.2.2 figure 3).

ISSN: 2324-9773

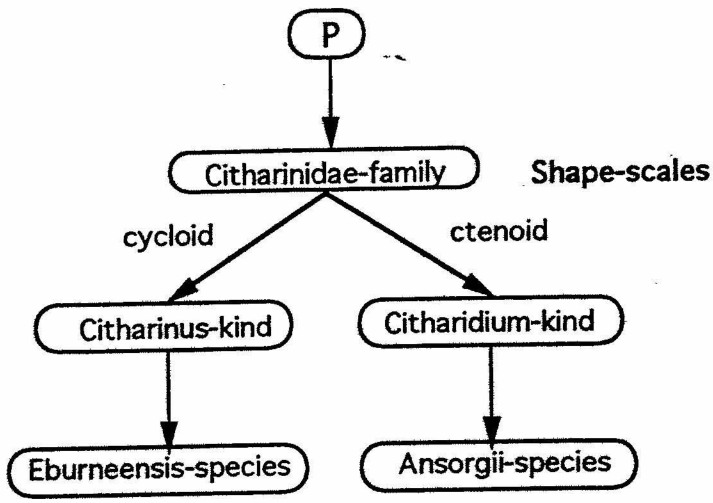

Figure 9. Discrimination Structure 2

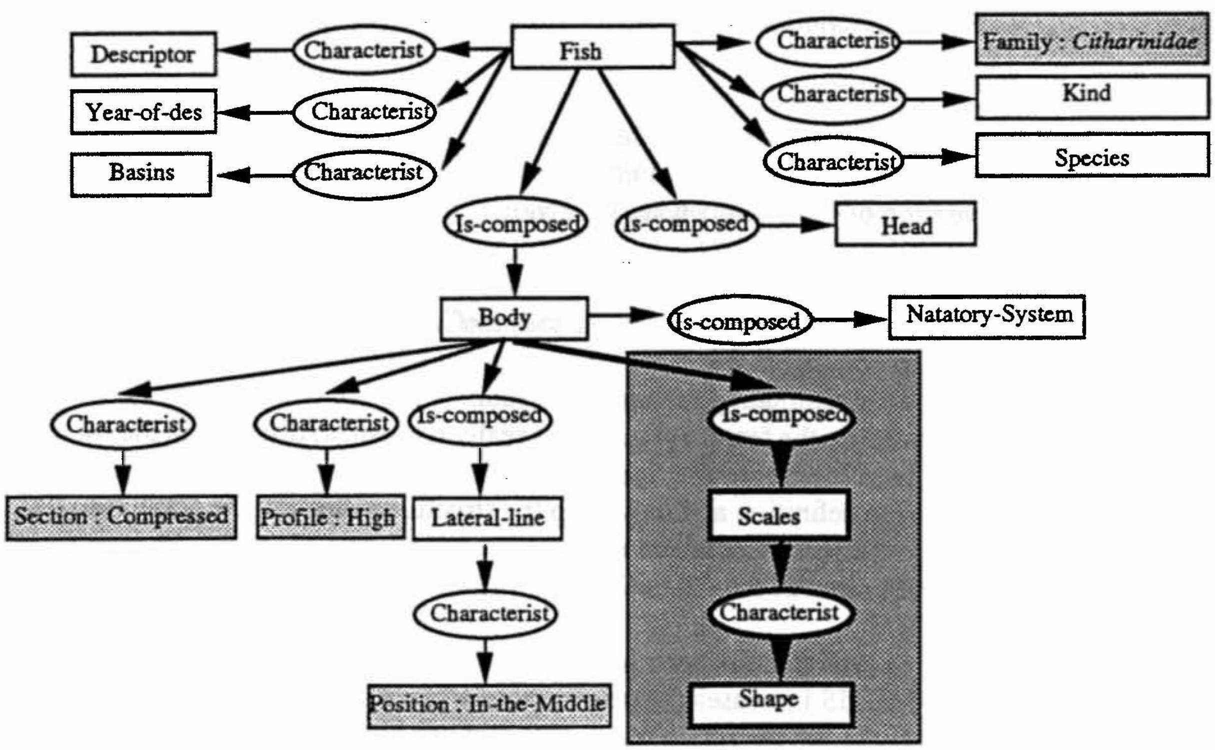

Figure 10. Modified Graph of the Citharinidae Family (G_1)

ISSN: 2324-9773

Aimeur, E., & Ganascia, J. (1993). Reasoning with Classification in Interactive Knowledge Elicitation. 4th ASIS SIG/CR Classification Research Workshop, 19-38. doi:10.7152/acro.v4i1.12609

Classification stage:

The third description to be acquired is that of the *Citharinops distichodoïdes* species, which belongs to the *Citharinops* kind of the *Citharinidae* family.

Once more, since the family description is common, METIS will try to classify the new example using the discrimination structure by asking the expert for the scales shape of the *Citharinops distichodoïdes* species (discriminating part of the Citharinidae-family type definition). If the answer is cycloid, then the discriminating feature is no longer sufficient since the Citharinus-kind type definition has the same feature. For the moment, METIS asks the expert for a new feature due to the fact that it does not have enough elements to discover this by itself. If the expert says that it is the number of scales, the latter then becomes the new discriminating feature (68 to 73 for the *Citharinus* kind and 50 to 56 for the *Citharinops* kind). In order to increase the discrimination part of the Citharidium-kind type definition, and possibly the characterizing part of the Citharidium-kind type definition, METIS seeks information about the number of scales for the latter. Then it extends the information acquired, to the species corresponding to each kind.

The generalization hierarchy is increased with G_{13}, G_{131}, which corresponds respectively to the Citharinops-kind type definition and to distichodoïdes-species type definition (see § 2.2.2 figure 3), and the new discrimination structure is as follows:

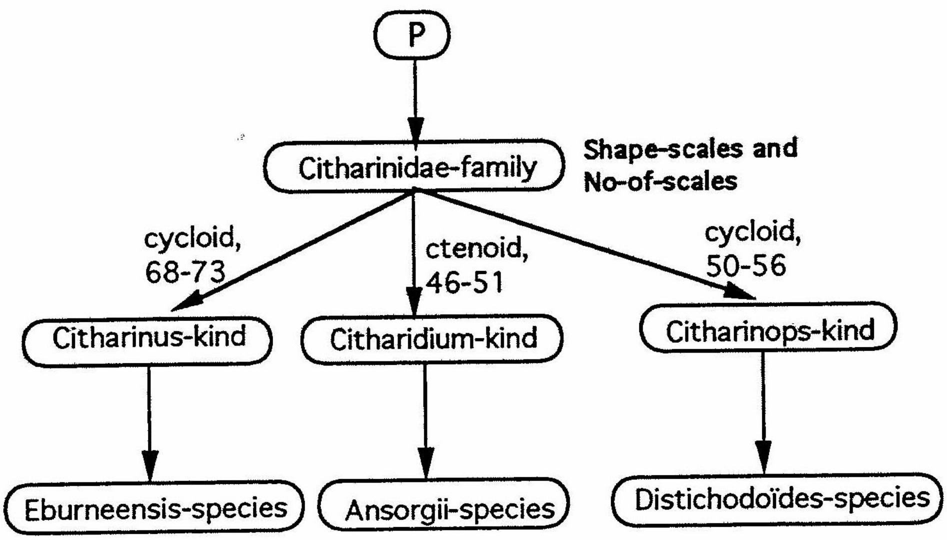

Figure 11. Discrimination Structure 3

ISSN: 2324-9773

Aimeur, E., & Ganascia, J. (1993). Reasoning with Classification in Interactive Knowledge Elicitation. 4th ASIS SIG/CR Classification Research Workshop, 19-38. doi:10.7152/acro.v4i1.12609

3.2. General Principle

We have seen above that the acquisition principle of METIS is based upon the use of the generalization hierarchy (by working on the characterizing parts of the type definitions) and upon the use of a discriminating structure, (by working on discriminating parts of the type definitions). Generally, once the *initialization stage* is achieved, METIS will first try the *classification stage* and then, in case of failure during discrimination, it will go through the *building stage of discriminating features*.

3.2.1. Classification stage

The process is interactive. METIS begins by asking the expert questions about the species to be acquired and about its position in the taxonomy (family, kind, species), then it tries to classify the new object through the discrimination structure. As each stage of this graph is successfully achieved (one is going from family level to kind level, for example, by validating the discriminating feature), the corresponding level in the generalization hierarchy is brought up to date if the expert wishes to increase the descriptions (modification of the characterizing part of type definitions through abstraction). Then the process comes back to the discrimination structure and so on, down to the lowest level. At the lowest level, METIS acquires the other pieces of information, which are in the Primitive graph and which were of no use during discrimination (example: [basins: Agnébi]). When an inherited discriminating feature is not valid, an incoherence occurs. Then, METIS must either increase the field of values of the discriminating feature with the new value and make the necessary updates, or request additional input from the expert.

Updates.

The canon becomes richer at each new acquisition (updating the type lattice, the set of individual markers, the set of relations, the conformity relation, and finally the canonical basis).

3.2.2. Building stage of discriminating features

If at a given level (kind level, for example), there is a failure in the classification stage, that is to say that the discriminating feature proves to be insufficient and METIS did not succeed in discriminating the new taxon from the other ones, it must find at least one feature which will allow it to distinguish it from the other taxa, by default by asking the expert. In order to discover new features, METIS compares, by means of a matching technique which will not be specified in this paper, some descriptions of taxa that are close in the generalization hierarchy (in our example, descriptions of other kinds of the same family) by considering only the *differentia*. Thus, it identifies analogies and, in case it does not discover any characteristic feature, it asks the expert directly and explicitly.

METIS discovers a feature

When METIS discovers a feature enabling it to distinguish two taxa, the expert can reject it, for although this feature is discriminating on the syntactic level, it does not add anything on the semantic level: the example ([Fish] -> (Characterist) -> [Year-of-description]) is obviously not discriminating. The expert is the only one able to assess the relevance of discriminating features: therefore it is interesting to get an interactive system which takes into account the expert's view on a punctual basis.

ISSN: 2324-9773

Aimeur, E., & Ganascia, J. (1993). Reasoning with Classification in Interactive Knowledge Elicitation. 4th ASIS SIG/CR Classification Research Workshop, 19-38. doi:10.7152/acro.v4i1.12609

The expert gives the discriminating feature

The acquisition of a discriminating feature with the help of the expert necessitates going through the structural concepts (see § 2.2.3) in a previously defined order. METIS can dynamically modify this order in agreement with the expert as new structural knowledge is acquired. Therefore, if the discriminating feature is [Scales] -> (Characterist) -> [Shape] and the scales concept is not known by METIS, the latter will ask the following questions: "Is the scales concept related to the head?". If the expert answers "no", it will go on by asking whether the scales concept is related to the body. If the expert answers "yes", METIS will ask what is the structural relation that links the body concept to the scales concept, and then METIS will relate the structure ([Body] -> (Is-composed) -> [Scales] and [Scales] -> (Characterist) -> [Shape]) to the body concept. When the expert gives the discriminating feature, METIS has to check whether or not the feature does belong to descendants of the sibling node in the generalization hierarchy. If such is the case, we have to come back to the level of parent node (see figure 10).

Data acquisition

When the first descendant of a given node is created (see in 3.1. initializing stage, second step), it is necessary for the expert to describe what characterizes it. This is due to the fact that at this stage METIS does not have any sibling nodes to be compared with in order to help its acquisition. On the other hand, if the taxon has other siblings, the processing is guided and METIS will ask questions about the feature which characterizes its other siblings in order to improve the description of the taxon considered.

Updates

Each update requires a return to the description of sibling node (horizontal updating) to acquire the referents associated with the *destination concept* (Cf. § 2.2.3) of the discriminating feature (see in 3.1. classification stage) or to correct false ones. It also requires a study of the descendants of the node in question (vertical update), thus aiming at keeping the inheritance in the generalization hierarchy, and therefore in the discrimination structure. Finally, once the discriminating feature is found, the structural model is extended with a possible restructuring of the generalization hierarchy (modification of the characterizing parts), and therefore of the discriminating structure (modification of the discriminating parts).

4. COMPARISON WITH OTHER SYSTEMS

In this section, we will try to compare METIS with two other works also based on knowledge elicitation, although they do not use the formalism of conceptual graphs in the knowledge representation.

The ETS system designed by BOOSE [BOOSE, 84] proposes some automatic interview techniques in order to build a description language. For this purpose, he makes use of the repertory grid techniques drawn from the psychotherapy methods developed by KELLY [KELLY, 55].

The AQUINAS system [BOOSE, 86] is a generalization and an improvement of ETS. AQUINAS includes several sub-systems intended for different tasks and combines several ways of solving

ISSN: 2324-9773

Aimeur, E., & Ganascia, J. (1993). Reasoning with Classification in Interactive Knowledge Elicitation. 4th ASIS SIG/CR Classification Research Workshop, 19-38. doi:10.7152/acro.v4i1.12609

problems. Among these tasks are: the decomposition of problems, the processing of unreliable data, the use of several sources of expertise, the integration of different types of data etc.

METIS like ETS and AQUINAS, is a knowledge elicitation system, which builds a description language on an interactive and gradual basis and which is well adapted to solve classification problems. In the three systems, the knowledge engineer is of no use, there is a predefined number of objects to be characterized (classes in the case of METIS) and the object and their features are revealed.

One of the differences between the METIS system and the ETS and AQUINAS systems lies in the fact that the objects to be characterized in METIS are not individual objects (instances) but classes of object organized into a taxonomy.

In ETS, everything is on the same level: the language is described by a sole repertory grid close to the propositional language. On the contrary, in AQUINAS, hierarchies are taken into account. Therefore, when building the evaluation grid, four types of hierarchy are used [BOOSE, 86]: the hierarchy of features, the hierarchy of solutions, the hierarchy of experts and finally the hierarchy of cases. In METIS, the features and the solutions are not separated but are organized in the same hierarchy: the generalization hierarchy, which is copied from the class taxonomy and where each node is itself described hierarchically in the form of SOWA's graph. Thus knowledge is directly structured and easily exploited.

To show how the hierarchies are taken into account, let us take the example of a taxonomy of fruits from a vitamin angle. Using ETS, one would try to discriminate oranges apples, apricots, prunes etc., without taking into account the hierarchy which exists. When using METIS and AQUINAS, the fresh fruits would first be discriminated from the dried fruits; then with the descendants of fresh fruits the citrus fruit would be discriminated from the stone fruits and the pip fruits; and, finally, with the descendants of citrus fruits, oranges would be discriminated from lemons, mandarins etc.

Using ETS and AQUINAS, the expert is not given help in finding discriminating features. As soon as the system detects that two objects have similar "profiles," the expert is obliged to provide the system with a feature which will allow them to be distinguished. Such is not the case for the METIS system where one refers to the structural properties of the description language in order to deduce a set of discriminating features. This set is then proposed to the expert who is the only one able to assess the pertinence of the features. When METIS does not succeed in finding by itself a set of potentially discriminating features, it poses the question clearly to the expert.

5. CONCLUSION

In this paper we have described the METIS system as an interactive system which elicits knowledge, gradually builds a description language and which is well adapted to solve the classification problems. In METIS, the objects to be characterized are not individual objects (instances), they are classes of objects organized into a taxonomy. In order to discriminate them, a generalization hierarchy is used in which each node is described hierarchically in the form of

ISSN: 2324-9773

Sowa's graphs. The knowledge is thus directly structured and can therefore be easily used [AÏMEUR and GANASCIA, 93].

The selection of a basic generic model was a significant stage in the cognitive process, which permits us to go from observation to description, and which is a starting point for the process of incremental acquisition.

The progressive building of a generalization hierarchy, copied from the taxonomy of classes, occurs thanks to a continuous dialogue with the expert, who is then faced with his own knowledge. This interactive validation of knowledge sometimes raises questions about the structural model. It must then be updated as well as the proper taxon description.

Moreover, the structural model can be used as a framework for a much larger information base including texts and images. A natural link would be established between the base of structured knowledge and hypertext interface. The resultant interaction between the user and METIS is likely to be quite ergonomic and user-friendly [AÏMEUR & al., 93].

We have defined in this article a formal framework where all the operations can receive precise definition. The choice of heuristic methods will be dealt with in more detail at another opportunity.

In [AÏMEUR, 93] we have compared METIS with other works applied to the biological sciences [VIGNES, 91], [CONRUYT and MANAGO, 92]. However, METIS can be applied to other fields such as archaeology and mineralogy where there are numerous objects which must be characterized and identified. These objects profit directly from a structural similarity and belong to an already predefined taxonomy. METIS is currently being studied in the context of ichthyology. At present, we are studying five families, corresponding to one hundred and fifty species.

6. REFERENCES

AÏMEUR, E. Acquisition d'Objets Structurés par Caractérisation. *JAVA: Journées Françaises d'Acquisition de Validation et d'Apprentissage,* Saint-Raphael, France 1993.

AÏMEUR, E.; KIEU, Q.; & GANASCIA, J.G. Acquisition of Structured Descriptions with Sowa's graphs. *ICO: 4th International Conference on Cognitive and Computer Sciences for Organizations,* pp 263-271, Montréal, Québec, Canada, 1993.

AÏMEUR, E. & GANASCIA, J.G. Elicitation of Taxonomies Based on the Use of Conceptual Graph Operators. *ICCS: First International Conference on Conceptual Structures,* Québec City, Canada, 1993.

BOOSE, J. Personal Theory and the Transfer of Human Expertise. *Proc 3th AAAI,* 1984.

BOOSE, J. Expertise Transfer and Complex Problems: using AQUINAS as a Knowledge Acquisition Workbench for Expert Systems. *Colloque sur l'acquisition des connaissances,* Banff, Canada, 1986.

CONRUYT, N. & MANAGO, M. Modélisation, Formalisation et Analyse d'Objets biologiques en vue de leur identification. *3èmes journées "Symbolique-Numérique",* pp. 343-358, Paris, France, 1992

ISSN: 2324-9773

Aimeur, E., & Ganascia, J. (1993). Reasoning with Classification in Interactive Knowledge Elicitation. 4th ASIS SIG/CR Classification Research Workshop, 19-38. doi:10.7152/acro.v4i1.12609

ELLIS, G. Compiled Hierarchical Retrieval. *Proceedings of the 6th Annual workshop on Conceptual Graphs*. pp 187-207, Binghamton, NY, USA, 1991.

FISHER, D.H. Knowledge Acquisition via Conceptual Clustering. *Machine Learning*, vol 2, pp 139-172, 1987.

GANASCIA. J.G. *AGAPE & CHARADE: deux techniques d'apprentissage symbolique appliquées à la construction de base de connaissances*. Thèse d'état de l'université de Paris Sud, France, 1987.

GEY, O. COCLUSH: Un Générateur de Classifications d'Objets Structurés, Suivant Différents Points de Vue. *Actes des 6èimes Journées Françaises de l'apprentissage*, pp 165-184, Sete, 1991.

KELLY, G.A. *The psychology of personal constructs*. New York: Norton, 1955.

LEVEQUE,. C.; PAUGY, D.; & TEUGELS, G.G. *Faune des poissons d'eaux douces et saumâtres de l'Afrique de l'Ouest*. ORSTOM / MRAC Editions, France, 1991.

MINEAU, G.; GECSEI, J.; & GODIN, R. La Classification Symbolique: Une Approche Non-Subjective. *Actes des 5ièmes Journées Françaises de l'apprentissage*, pp 169-189, Lannion, 1990.

NAPOLI, A. & RECHENMANN, F. Introduction à la journée d'étude sur la classification Représentation par objets, *La Grande Motte*, pp 211-213, 1993.

NOGIER, J.F. *Génération automatique de langage et graphes conceptuels*. Editions HERMES, pp. 49-74, Paris, 1991.

PAUGY, D. *Révision Systématique des ALESTES et BRYCINUS Africains*. *Collection Etudes et Thèses*. Editions ORSTOM, France, 1986.

QUINLAN, J.R. Induction of Decision Trees. *Machine Learning* 1, pp 81-132, 1986.

SOWA, J.F. *Conceptual Structures: Information Processing in Mind and Machine*. Addison-Wesley Publishing Co, 1984.

STEPP, R.E. & MICHALSKI, R.S. Learning from Observation: Conceptual Clustering: Inventing Goal Oriented Classifications of Structured Objects. *Machine Learning: An Artificial Intelligence Approach*, vol 2, pp 471-498. Morgan-Kaufman Publishing Co, 1986.

VIGNES, R. *Caracérisation automatique de groupes biologiques*. Thèse de l'université de Paris 6, France, 1991.

ISSN: 2324-9773