**1/1** (2003), 125–135
tmcs@math.klte.hu
http://tmcs.math.klte.hu

**Teaching**
**Mathematics and**
**Computer Science**

# Aspects théoriques de la classification à base de treillis

Katalin Bognár

*Résumé.* La classification est une notion cruciale dans les systèmes orientés objets et se fait de plus en plus présente en représentation de connaissances. Elle permet principalement de trouver des regularités dans un grand tableau de nombres. Dans ce sens général, il s'agit donc d'une méthode qui joue un rôle important dans différents domaines scientifiques où les connaissances sont à organiser selon certaines hiérarchies (biologie, chimie, etc.). En informatique nous parlons aussi de langages de classes sans mentionner les aspects mathématiques de la classification. Dans cet article l'auteur a pour but de proposer une introduction à la classification à travers la notion de treillis. Nous sommes persuadés que l'étude de la classification permet aux étudiants de familiariser leurs connaissances sur la modélisation et la programmation orientée objet.

*Mots-clés :* représentation de connaissances, modélisation orientée objet, treillis.

Theoretical aspects of classification based on lattices

*Abstract.* The classification is a crucial notion in the object oriented systems and more and more appears in the knowledge representation. It allows us to find the regularities in a huge table of numbers. In this general sense the classification plays an important role in various domains of science, where knowledge has to be organized into hierarchy (biology, chemistry, etc.) In the computer science the languages of classes are often studied without mathematical aspects of the classification. In this paper the author has the goal to propose an introduction to the classification through the notion of lattices. We are convinced that the study of classification allows students to enlarge their knowledge on the object oriented modelling and programming.

*Key words and phrases:* knowledge representation, object oriented modelling, lattices.

*ZDM Subject Classification:* D80, H50.
*ACM Comp. Class. System:* I.2.4.

# 1. Introduction

Effectuer une classification, c'est mettre en évidence des relations entre des objets et entre ces objets et leurs paramètres. A partir de mesures de proximités ou de dissemblances, il s'agit de construire une partition de l'ensemble des objets en un ensemble de classes les plus homogènes possibles. Les systèmes classificatoires sont à la croisée de nombreux chemins :

- logique (systèmes de représentation),
- réseaux sémantiques et frames,
- languages de classes,
- logique de descriptions.

Les sytèmes classificatoires s'appuient sur une hiérarchie de classes – qui possèdent des instances – sur lesquelles opère le raisonnement par classification. La classification de classes recouvre les processus de construction de classes en hiérarchies et l'insertion d'une classe dans une hiérarchie. La classification d'instances recouvre le processus de reconnaissance de la classe d'un individu (ou un objet individuel). Dans cet article nous tentons de présenter une approche possible pour introduire la notion de classification au cours de l'enseignement de systèmes à base de connaissances définissant la hiérarchie à base de treillis (voir [1]). L'article est organisé comme suit : d'abord la définition et des propriétés de treillis sont présentées, puis la hiérarchie conceptuelle et le processus de la classification sont introduits. Enfin, après avoir présenté un exemple simple, je proposerai quelques pistes de discussion.

# 2. Arrière-plan mathématique

## 2.1. La notion de treillis

*Définition 1.* L'ensemble $S$ est partiellement ordonné, si une relation $R$ reflexive, antisymétrique et transitive est définie sur certains couples d'éléments de $S$.

Dans un ensemble $S$ partiellement ordonné, la propriété suivante est vraie : pour tous les $a, b \in S$, $R(a, b)$ ou $R(b, a)$ ou $a$ et $b$ sont non-comparables.

*Définition 2.* Soit $S$ un ensemble partiellement ordonné. $S$ est ordonné si tous les $a, b \in S$ sont comparables.

*Définition 3.* Soit $S$ un ensemble partiellement ordonné et $a, b, x \in S$. L'élément $x$ est une borne supérieure (borne inférieure) des éléments $a$ et $b$ si $R(a, x)$ et $R(b, x)$ ($R(x, a)$ et $R(x, b)$ pour l'inférieure).

*Définition 4.* Soit $S$ un ensemble partiellement ordonné et $a, b, c \in S$. L'élément $c$ est la borne supérieure minimale (borne inférieure maximale) des éléments $a$ et $b$ si

- $c$ est une borne supérieure (borne inférieure) des éléments $a$ et $b$, et
- pour $\forall x \in S$, si $x$ est une borne supérieure (borne inférieure) des éléments $a$ et $b$, alors $R(c, x)$ ($R(x, c)$ pour inférieure maximale).

THÉORIE 1. *Soit $S$ un ensemble partiellement ordonné et $a, b \in S$. Si la borne supérieure (borne inférieure) des éléments $a$ et $b$ existe, alors elle est unique.*

*Définition 5.* Soit $S$ un ensemble partiellement ordonné par la relation $R$. Le couple $(S, R)$ est un treillis, si chaque couple d'éléments $x$, $y$ de S possède une borne supérieure minimale (noté $x \wedge y$) et une borne inférieure maximale (noté $x \vee y$). Usuel, pour désigner un treillis, on ne mentionne pas la relation $R$.

*Définition 6.* Soit $P$ un treillis et $e$, $O \in P$.
L'élément $e$ est un élément unité si pour pour tout $a \in P$, il est vrai que $R(a, e)$.
L'élément $O$ est un zéro élément si pour tout $a \in P$, il est vrai que $R(O, a)$.

Dans un treillis l'élément unité et le zéro élément n'existent pas forcément.

Exemples de treillis

- Considérons les ensembles suivantes $\{a, b, c\}$, $\{a\}$, $\{b\}$, $\{c\}$, $\{b, c\}$, $\{\emptyset\}$ et la relation $\subseteq$. Voire la figure 1, qui montre le treillis.
- L'ensemble $P(X)$ de toutes les parties d'un ensemble $X$, où la relation est l'inclusion.

## 2.2. Propriétés des treillis

Dans cette section, soit $P$ un treillis, la relation $R$ définit l'ordre partiel sur $P$ et $a, b, c \in P$.

Il est évident par la conséquence de la Définition 4. que si $R(a, b)$, alors la borne supérieure minimale est $b$ ($a \wedge b = b$) et la borne inférieure maximale des éléments $a$ et $b$ est $a$ ($a \vee b = a$).

Dans un treillis $P$ les bornes supérieure minimale et inférieure maximale possèdent les propriétés suivantes :

1. $a \wedge a = a$ $\quad a \vee a = a$
2. $a \wedge b = b \wedge a$ $\quad a \vee b = b \vee a$
3. $a \wedge (b \wedge c) = (a \wedge b) \wedge c$ $\quad a \vee (b \vee c) = (a \vee b) \vee c$
4. $(a \wedge b) \vee a = a$ $\quad (a \vee b) \wedge a = a$

Ces propriétés sont habituellement appellées axiomes des treillis.

Si la cardinalité du treillis $P$ est finie, alors l'élément unité et le zéro élément existent. Soit $P = \{a_1, a_2, \ldots, a_n\}$. Alors, $e = a_1 \wedge a_2 \wedge \ldots \wedge a_n$ et $O = a_1 \vee a_2 \vee \ldots \vee a_n$. Il est évident que, si l'élément unité et le zéro élément existent, alors ils sont uniques. De plus, ils possèdent les propriétés suivantes :

1. $e \wedge a = e$ $\quad e \vee a = a$
2. $O \wedge a = a$ $\quad O \vee a = O$

Soit $S$ un ensemble ordonné. $S$ est un treillis, tel que $a \wedge b = \max(a, b)$ et $a \vee b = \min(a, b)$.

## 3. La hiérarchie conceptuelle

Une classe représentant un concept du monde réel est une entité générique qui regroupe un ensemble d'éléments et qui peut posséder une description qui lui est propre. Donc, une classe $C$ possède une identité et un ensemble de propriétés qui sont caractéristiques de l'état et du comportement du concept représenté, et se peut être décrit par la conjonction $C = (a_1, s_1) \sqcap (a_2, s_2) \ldots \sqcap (a_n, s_n)$ où les $a_k$ sont des attributs et les $s_k$ des spécifications attachées à l'attribut précisant le type, le domaine et la cardinalités des valeurs de l'attribut (les $a_k$ sont différents deux à deux). La classification de classes est une opération qui consiste à découvrir des régularités pour regrouper des entités individuelles en classes. Dans le cadre des systèmes de représentation de connaissances par objets (RCO), deux approches principalement sont à distinguer. La première est relative à la conception de hiérarchies d'héritage. Dans la deuxième approche, les objets de base sont traités globalement ; ils sont regroupés en des classes et l'ensemble de ces classes est organisé en une hiérarchie.

La subsomption est une relation générale qui permet d'organiser les classes en hiérarchies. La définition formelle précise est donnée dans le cadre de la logique de description (voir [2]). Intuitivement une classe $C$ est subsumée par une classe $D$ (ce qui est noté $C \sqsubseteq D$), si

- tous les attributs de $D$ sont possédés par $C$,
- toutes les spécifications d'état des attributs de $D$ sont vérifiés par les attributs de $C$.

*Définition 7.* Une hiérarchie conceptuelle $\mathcal{H}$ est un treillis $(\chi, \top, \sqsubseteq)$, où $\chi$ est un ensemble fini de classes, $\sqsubseteq$ est une relation d'ordre partiel sur les classes, appelée subsomption et $\top$ est l'élément d'unité de $\chi$ suivant $\sqsubseteq$. $\top$ est appelée la racine de la hiérarchie.

Un système de RCO s'appuie sur une telle hiérarchie.

Dans le diagramme de treillis de $\chi$ on représente le fait qu'une classe $D$ est subsumée par $C$ en traçant l'arc $\vec{DC}$.

## 4. Le processus de classification

Le processus de classification cherche à mettre en évidence les dépendences implicites qui existent entre les objets de $\mathcal{H}$, les classes entre elles, les classes et les instances. La classification recouvre les processus de reconnaissance de la classe d'un objet, et l'insertion éventuelle d'une classe dans une hiérarchie. Ce mode de raisonnement permet de reconnaître un objet en identifiant ses caractéristiques, relativement à la hiérarchie étudiée. La classification fait intervenir un processus de décision d'appartenance. Le processus de classification qui permet d'insérer l'objet x dans la hiérarchie $\mathcal{H}$, se schématise comme suit :

$$(\chi, \top, \sqsubseteq) \times \{x\} \rightarrow (\chi \cup \{x\}, \top, \sqsubseteq).$$

Le processus de classification s'appuie sur le caractère nécessaire et suffisant de l'ensemble des propriétés qui caractérisent l'état d'une classe.

- *Condition nécessaire*
  Soit $C$ une classe et $i$ une instance de la classe $C$. Alors $i$ vérifie toutes les propriétés de $C$ et possède tous les attributs de $C$.
- *Condition suffisante*
  Soit $x$ un objet possèdant toutes les propriétés de $C$. Dans ce cas, $x$ peut être classifié comme une instance de $C$.

## 4.1. Un algorithme de classification

L'opération de classification qui permet de placer l'objet $x$ dans la hiérarchie $\mathcal{H}$, se décompose en trois étapes :

(1) La recherche des subsumants les plus spécifiques de $x$(SPS).

(2) La recherche des subsumés les plus généraux de $x$ (SPG).

(3) La mise en place de nouvelles relations entre l'objet à classer $x$ et ses subsumants et ses subsumés.

### 4.1.1. La recherche des subsumants les plus spécifiques

Le principe est de parcourir le graphe des classes en profondeur en partant de la racine jusqu'à trouver une classe qui ne corresponde pas aux caractéristiques de l'objet à classer.

### 4.1.2. La recherche des subsumés les plus généraux

Il suffit de ne considérer que l'ensemble des descendants des SPS qui possèdent les mêmes propriétés que l'objet à classer. Si celui-ci subsume un descendant, alors il est un SPG et sa descendance est ignorée, sinon leurs descendants sont testés à leur tour, jusqu'à ce qu'un SPG soit trouvé ou bien jusqu'à ce qu'il n'y ait plus de descendant à tester.

### 4.1.3. La mise en place de nouvelles relations entre l'objet à classer $x$ et ses subsumants et ses subsumés

Lorsque les SPS et les SPG de classe $x$ ont été découverts, de nouveaux liens sont mis en place entre la classe $x$, ses SPS et ses SPG. Si la classe $x$ est equivalente à une classe déjà présente dans la hierarchie, alors la classe $x$ et la classe présente sont identifiés.

## 4.2. Exemples

### 4.2.1. Exemples 1.

Soit la relation de subsomption suivante : $x$ subsume $y$ si $x$ divise $y$ dans $\mathcal{N}$. La figure 2 représente la hiérachie des nombres suivants : 1, 2, 3, 5, 6, 7, 14, 21, 28, 126, 210, 252, 280. La figure 3 montre la hiérachie après l'insertion du nombre 42. Les SPS de 42 sont 6, 14 et 21 et les SPG de 42 sont 126 et 210. Les arcs du nombre 210 vers 6 et du nombre 126 vers 6, 14 et 21 sont supprimés.

4.2.2. Exemples 2.

La figure 3 montre la hiérachie des mots de l'ensemble $\{\Lambda, a, ab, abc, b, bab, bcd, c, d, dd, ddd\}$ où $\Lambda$ qui désigne le mot vide, est la racine de la hiérachie. Une classe détermine un ensemble de mots, qui contiennent chacun un motif défini sur l'alphabet $\{a, b, c, d\}$. Le nom de la classe est donné par le motif qui la caractérise. Ainsi $ab = (\text{motif}, *ab*)$ où motif est un nom d'attribut et $*$ dénote une chaîne quelconque. Définissons la relation subsomption entre deux classes $x$ et $y$ telle que : $x \sqsubseteq y$ si le motif $y$ est un sous-mot du motif $x$, c'est-à-dire, si $x$ s'écrit $mym'$ où $m$ et $m'$ sont deux mots. La figure 5 montre l'insertion du mot $bc$ dans la hiérachie. Les SPS de $bc$ sont $b$ et $c$, et les SPG de $bc$ sont $abc$ et $bcd$. Les nouveaux arcs sont en gras. Les arcs du mot $abc$ vers $c$ et du mot $bcd$ vers $b$ sont supprimés.

## 5. Discussion

Le raisonnement dans un système de RCO consiste à exploiter les propriétés d'une hiérarchie représentant des connaissances relatives à un certain domaine de référence. Les opérations principales qui sont à la base du raisonnement sont les suivants :

- Le test de subsomption consiste à vérifier qu'une classe $C$ subsume une classe $D$.
- La classification de classes consiste à placer une nouvelle classe $X$ dans l'ordre associé à une hiérarchie $\mathcal{H}$.
- La classification d'instances consiste à déterminer les classes dont un objet $x$ donné peut être une instance.
- La recherche de propriétés consiste à trouver les propriétés détenues par une classe ou une instance, les restrictions associées à ces propriétés et/ou leurs valeurs.

Comme tout système à base de connaissance, un système de RCO peut se voir comme un système logique. Dans l'approche syntaxique le processus de construction de classes est une procédure d'inférence qui repose sur la subsomption. Le processus de construction de classes peut être réalisé par la sémantique attachée à une interprétation donnée suivant la subsomption actuelle.

## 6. Conclusion

Dans cet article nous nous sommes intéressés à la représentation de connaissances par objets, en essayant d'introduire la notion de classification. Nous avons montré que la notion de classification peut être définie à travers de treillis des classes, où la relation de treillis assure la classification. Le point de vue que nous avons adopté possède entre autres avantages d'être formellement défini. Nous sommes persuadés qu'il est très important pour les étudiants de connaître la base théorique. Une suite de cette approche concerne l'étude précise du raisonnement à partir de cas (voir [3]).

## Figures

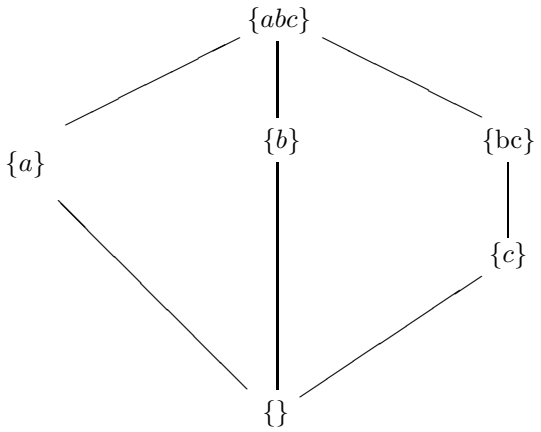


*Figure 1*

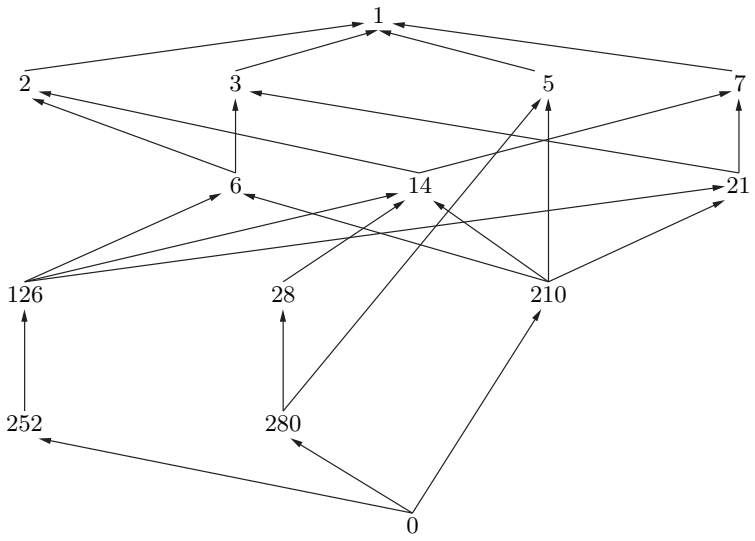


*Figure 2*

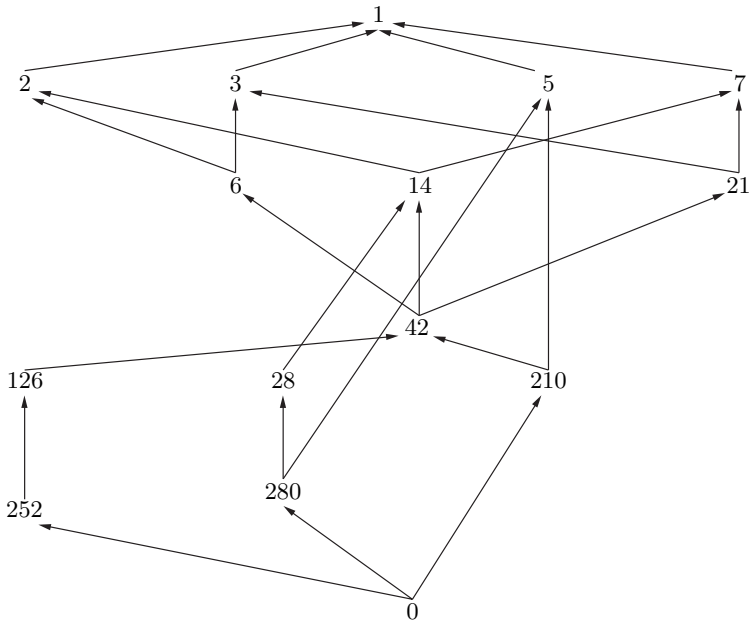


*Figure 3*

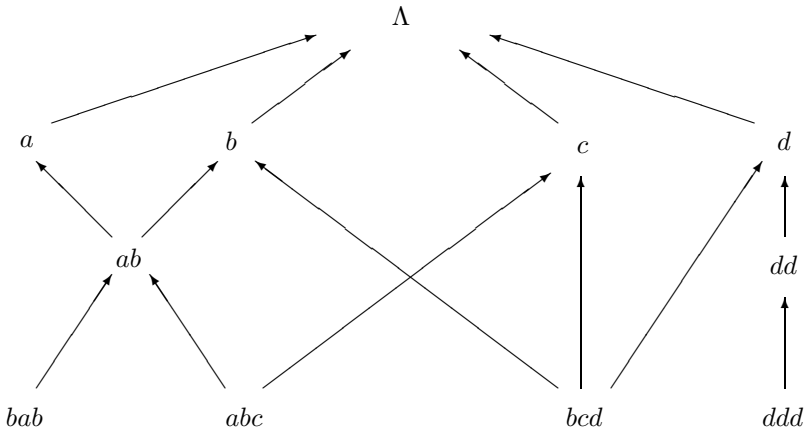


*Figure 4*

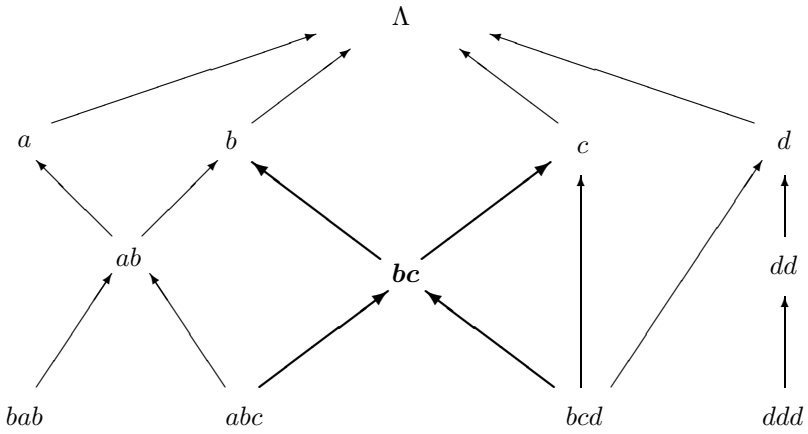


*Figure 5*

## Références

[1] R. Godin, H. Mili, G. W. Mineau and R. Missaoui, Design of class hierarchies based on concept (Galois) lattices, *Theory and Application of Object Systems* **4**(2) (1999), 117–134.

[2] K. Bognár, Leíró logikák az ismeretábrázolásban (Description logics in knowledge representation), *Alkalmazott Matematikai Lapok, Budapest* **20** (2000), 183–193 (in Hungarian).

[3] A. Napoli, *Catégorisation, raisonnement par classification et raisonnement à partir de cas*, JAVA'94, Strasbourg, 1994, E1–E14.

KATALIN BOGNÁR
UNIVERSITÉ DE DEBRECEN
INSTITUT DE MATHÉMATIQUES ET INFORMATIQUE
DEBRECEN H-4010, P.O. BOX 12
HONGRIE

*E-mail:* bognar@math.klte.hu

*(Received January 15, 2003)*