BULLETIN OF THE POLISH
ACADEMY OF SCIENCES
MATHEMATICS
Vol. 36, No. 7-8, 1988

THEORETICAL COMPUTER SCIENCE

Independence of Attributes

by
Miroslav NOVOTNÝ and Zdzisław PAWLAK

Presented by Z. PAWLAK on March 30, 1988

Summary. If K is a congruence on the semilattice $(B(A), \cup)$ where $B(A)$ denotes the system of all subsets of A, the set of all elements in $B(A)$ is investigated such that any of them is minimal in some block of K; such elements are said to be independent. For any subset P of A its reduct is defined to be a minimal element in the block of K containing P; furthermore, a subreduct of P is a maximal element in the set of all indepedent elements that are included in P. Some results concerning the intersection of all reducts and the union of all subreducts of P are presented. This investigation is motivated by dependency problems concerning attributes in an information system.

1. Introduction. Let (X, A, V, f) be an information system (see [1, 5]). This means that X, A, V are finite nonempty sets and f is a mapping of $X \times A$ into V. Elements of X are interpreted to be objects, elements of A are said to be attributes, elements in V are called values of attributes; the equality $v = f(x, a)$ is interpreted to express that the attribute a has the value v for the object x.

For any set $P \subseteq A$ of attributes we put

$$\tilde{P} = \{(x, y) \in X \times X; f(x, a) = f(y, a) \quad \text{for any } a \in P\}.$$

Clearly, $\tilde{P}$ is an equivalence on the set X. Thus, a set P of attributes defines an equivalence $\tilde{P}$ on the set X of objects; objects of X that are in a block of $\tilde{P}$ are indiscernible by means of attributes of the set P.

Two sets $P \subseteq A$ and $Q \subseteq A$ of attributes are said to be equivalent if $\tilde{P} = \tilde{Q}$, i.e. if they define the same indiscernibility relation. A set $P \subseteq A$ of attributes is said to be independent if $\tilde{P} \neq \tilde{Q}$ for any $Q \subseteq P$ with $Q \neq P$.

For any set of attributes $P \subseteq A$, there exists at least one independent set $Q \subseteq P$ such that $\tilde{Q} = \tilde{P}$. It is said to be a reduct of P. Reducts of a set of attributes are of practical use; any of them define the same indiscernibility relation as the original set but can be essentially smaller.

1.1. EXAMPLE. Let (X, A, V, f) be an information system where X is the set of patients in a hospital, A is the set of medical symptoms (e.g. temperature, blood pressure, etc.), V the set of values of these symptoms. For a set $P \subseteq A$, a block of $\tilde{P}$ is a set of patients with the same values of all symptoms in P; we suppose that a diagnosis can be assigned to any block of $\tilde{P}$. It can happen that the same blocks are obtained on the basis of some set $Q \subseteq P$, $Q \neq P$, i.e., that some symptoms in P are superfluous. □

For this reason, reducts are investigated in the present paper. The intersection of all reducts of a set P of attributes is proved to be the set of all the so-called indispensable elements in P.

If $P \subseteq A$ is a set of attributes, then any reduct Q of P has the following property:

(p) $Q \subseteq P$, Q is independent and for any independent set R with $Q \subseteq R \subseteq P$ the equality $Q = R$ holds.

We shall construct some examples of sets Q, P with the property (p) where Q is not a reduct of P. If Q, P are sets of attributes with the property (p), then Q will be called a subreduct of P. We prove that the union of all subreducts of P and of a suitable set of attributes equivalent with Ø equal to P.

2. Equivalence of sets of attributes. Let $S = (X, A, V, f)$ be an information system and $\boldsymbol{B}(A)$ denote the set of all subsets of A. We put

$$K_S = \{(P, Q) \in \boldsymbol{B}(A) \times \boldsymbol{B}(A);\ \tilde{P} = \tilde{Q}\}.$$

This means that K_S is the equivalence of attribute sets as it was mentioned in Section 1. We have

2.1. THEOREM. (i) *For any information system $S = (X, A, V, f)$ the equivalence K_S is a congruence on the semilattice $(\boldsymbol{B}(A), \cup)$.*

(ii) *For any finite set $A \neq \emptyset$ and any congruence K on the semilattice $(\boldsymbol{B}(A), \cup)$ there exists an information system $S = (X, A, V, f)$ such that $K = K_S$.*

For the proof, see 2.3 and 2.4 of [2] and 2.1 of [3]. □

Thus, instead of equivalences of attribute sets we may investigate congruences on semilattices of the form $(\boldsymbol{B}(A), \cup)$. This is a simplification from the formal point of view. The following results concerning congruences will be useful.

2.2. LEMMA. *Let K be a congruence on $(\boldsymbol{B}(A), \cup)$. If P, Q, R are in $\boldsymbol{B}(A)$, and $P \subseteq Q \subseteq R$, $(P, R) \in K$ hold, then $(P, Q) \in K$, $(Q, R) \in K$ hold, too.*

Proof. Clearly, $(P \cup Q, R \cup Q) \in K$ which means $(Q, R) \in K$. Transitivity and symmetry of K imply $(P, Q) \in K$. □

2.3. LEMMA. *Let K be a congruence on $(\boldsymbol{B}(A), \cup)$ where A is finite. If $P \in \boldsymbol{B}(A)$ and $\emptyset \neq \boldsymbol{M} \subseteq \boldsymbol{B}(A)$ are such that $(P, M) \in K$ for any $M \in \boldsymbol{M}$, then $(P, \bigcup \boldsymbol{M}) \in K$ where $\bigcup \boldsymbol{M}$ means $\bigcup \{M;\ M \in \boldsymbol{M}\}$.*

Proof. The finiteness of A implies that of $\boldsymbol{B}(A)$ and, therefore, of $\boldsymbol{M}$. Thus, $\boldsymbol{M} = \{M_1, \ldots, M_p\}$ for some $p \geqslant 1$, and $(P, M_i) \in K$ for any i with $1 \leqslant i \leqslant p$ implies that $(P, \bigcup \boldsymbol{M}) = (\underbrace{P \cup \ldots \cup P}_{p \text{ times}}, M_1 \cup \ldots \cup M_p) \in K$. □

3. Independent elements. Let $A \neq \emptyset$ be a finite set, K a congruence on the semilattice $(\boldsymbol{B}(A), \cup)$. The set $\boldsymbol{B}(A)$ is (partially) ordered by inclusion. An element $P \in \boldsymbol{B}(A)$ is said to be *K-independent* if there exists a block $C \in \boldsymbol{B}(A)/K$ such that P is minimal in $\boldsymbol{C}$. We denote by $\mathbf{IND}_K$ the set of all K-independent elements in $\boldsymbol{B}(A)$.

3.1. Lemma. *If $P \in \mathbf{IND}_K$ and $Q \subseteq P$, then $Q \in \mathbf{IND}_K$.*

Proof. Suppose $Q \subseteq P$, $Q \neq P$, and $Q \notin \mathrm{IND}_K$. There exists $C \in \boldsymbol{B}(A)/K$ such that $Q \in \boldsymbol{C}$; clearly, Q is not minimal in $\boldsymbol{C}$. Hence, there exists $Q_0 \subseteq Q$, $Q_0 \neq Q$ such that Q_0 is minimal in $\boldsymbol{C}$. Furthermore, $(Q_0, Q) \in K$ which implies that $(Q_0 \cup (P-Q), P) \in K$, $Q_0 \cup (P-Q) \subseteq P$, $Q_0 \cup (P-Q) \neq P$. Thus, P is not minimal in its block of K. Hence, $P \notin \mathrm{IND}_K$ which is a contradiction. □

3.2. Lemma. *The following two assertions are equivalent*:

(i) $P \in \mathbf{IND}_K$.

(ii) $(P, P-\{p\}) \notin K$ *for any* $p \in P$.

Proof. Clearly, (i) implies (ii). Suppose that (ii) holds and that $P \notin \mathrm{IND}_K$. Thus, there exists $Q \subseteq P$ such that $Q \neq P$ and $(Q, P) \in K$. Let $p \in P-Q$ be arbitrary. Then $Q \subseteq P-\{p\} \subseteq P$ and, hence, $(P-\{p\}, P) \in K$ by 2.2 which is a contradiction. □

4. Reducts. Let $A \neq \emptyset$ be a finite set, K a congruence on the semilattice $(\boldsymbol{B}(A), \cup)$. For any set $P \in \boldsymbol{B}(A)$, the set $Q \in \boldsymbol{B}(A)$ is said to be a *K-reduct of P* if $Q \in \mathrm{IND}_K$, $Q \subseteq P$, $(P, Q) \in K$ hold. The set of all K-reducts of P is denoted by $\mathbf{RED}_K(P)$.

4.1. Theorem. *For any $X \in \boldsymbol{B}(A)$, the set $\mathbf{RED}_K(X)$ is nonempty.*

It is a consequence of the fact that the set $\{Y \in \boldsymbol{B}(A);\ Y \subseteq X,\ (Y, X) \in K\}$ is finite and nonempty. Thus, the set of its minimal elements with respect to inclusion is nonempty. □

Let P be a subset of A.

An element $p \in P$ is said to be *K-dispensable for P* if $(P, P-\{p\}) \in K$. An element $p \in P$ is said to be *K-indispensable for P* if it is not K-dispensable for P, i.e., if $(P, P-\{p\}) \notin K$. The set of all K-indispensable elements for P is said to be the *K-core of P* and is denoted by $\mathrm{CORE}_K(P)$.

4.2. Theorem. *$CORE_K(P) = \bigcap\{Q;\ Q \in \mathbf{RED}_K(P)\}$ for any $P \in \boldsymbol{B}(A)$.*

Proof. If $Q \in \mathbf{RED}_K(P)$ and $p \in P-Q$, then $(P, Q) \in K$, $Q \subseteq P-\{p\} \subseteq P$ which implies that $(P-\{p\}, P) \in K$ by Lemma 2.2. Thus, $p \notin \mathrm{CORE}_K(P)$ and, therefore, we have $\mathrm{CORE}_K(P) \subseteq \bigcap\{Q;\ Q \in \mathbf{RED}_K(P)\}$.

On the other hand, if $p \in Q$ for any $Q \in \mathbf{RED}_K(P)$ and $(P, P-\{p\}) \in K$, then there exists a minimal set Q_0 in the set $\{Q \in \boldsymbol{B}(A);\ Q \subseteq P-\{p\},\ (P, Q) \in K\}$. Then $Q_0 \in \mathbf{RED}_K(P)$. Furthermore, $p \notin Q_0$ which is a contradiction. Thus, $\bigcap\{Q;\ Q \in \mathbf{RED}_K(P)\} = \mathrm{CORE}_K(P)$. □

5. Subreducts. Let $A \neq \emptyset$ be a finite set, K a congruence on the semilattice $(\boldsymbol{B}(A), \cup)$. For any set $P \in \boldsymbol{B}(A)$, we set $\mathbf{IND}_K(P) = \{Q \in \mathbf{IND}_K;\ Q \subseteq P\}$; let $\mathbf{SRED}_K(P)$ be the set of all elements in $\mathbf{IND}_K(P)$ that are maximal with respect to the inclusion. Any set in $\mathbf{SRED}_K(P)$ is called a *K-subreduct of P*. Clearly

5.1. LEMMA. *For any $P \in \boldsymbol{B}(A)$, the inclusion $\boldsymbol{RED}_K(P) \subseteq \boldsymbol{SRED}_K(P)$ holds.* □

An element $P \in \boldsymbol{B}(A)$ is said to be an almost K-zero element if $(P, \emptyset) \in K$. Furthermore, an element $P \in \boldsymbol{B}(A)$ is called *K-accessible* if there exists a set $\boldsymbol{M} \subseteq \mathbf{IND}_K$ and an almost K-zero element Q such that $P = Q \cup \bigcup \boldsymbol{M}$. Clearly

5.2. LEMMA. *Any almost K-zero element is K-accessible.* □

5.3. LEMMA. *For any element $X \in \boldsymbol{B}(A)$ that is not almost K-zero element, there exists an element $Y \in \boldsymbol{B}(A)$ such that $X = Y \cup \bigcup \boldsymbol{RED}_K(X)$, $Y \subseteq X$, $(Y, X) \notin K$.*

Proof. Since $\mathbf{RED}_K(X) \neq \emptyset$ by 4.1 and $(Q, X) \in K$ for any $Q \in \mathbf{RED}_K(X)$, we have $(\bigcup \mathbf{RED}_K(X), X) \in K$ by Lemma 2.3. If $\bigcup \mathbf{RED}_K(X) = X$, we put $Y = \emptyset$ and the assertion holds.

Suppose $\bigcup \mathbf{RED}_K(X) \neq X$. Since $(\bigcup \mathbf{RED}_K(X), X) \in K$ and X is not almost zero element, we have $\bigcup \mathbf{RED}_K(X) \neq \emptyset$. Thus, $\emptyset \neq X - \bigcup \mathbf{RED}_K(X) \neq X$. We put $Y = X - \bigcup \mathbf{RED}_K(X)$. If $(X, Y) \in K$, there exists a set $Y_0 \in \mathbf{RED}_K(Y)$ by 4.1. Thus $Y_0 \in \mathbf{IND}_K$ and $(Y_0, X) \in K$, $Y_0 \subseteq Y \subseteq X$. This implies that $Y_0 \in \mathbf{RED}_K(X)$ and, therefore, $Y_0 \subseteq \bigcup \mathbf{RED}_K(X)$. Hence $Y_0 \subseteq Y \subseteq X - \bigcup \mathbf{RED}_K(X)$ which entails that $Y_0 = \emptyset$ and, consequently, X is an almost K-zero element contrary to our hypothesis. Thus, $(X, Y) \notin K$, $Y \subseteq X$, and $X = Y \cup \bigcup \mathbf{RED}_K(X)$. □

5.4. LEMMA. *If $X \in \boldsymbol{B}(A)$ is such that any $Z \in \boldsymbol{B}(A)$ with $Z \subseteq X$, $Z \neq X$ is K-accessible, then X is K-accessible.*

Proof. If X is an almost K-zero element, then it is K-accessible by 5.2. If X is not almost K-zero element, there exists $Y \in \boldsymbol{B}(A)$ such that $X = Y \cup \bigcup \mathbf{RED}_K(X)$, $Y \subseteq X$, $Y \neq X$, $(Y, X) \notin K$ by 5.3. By hypothesis, Y is K-accessible, i.e. there exists a set $\boldsymbol{N} \subseteq \mathbf{IND}_K$ and an almost K-zero element Q such that $Y = Q \cup \bigcup \boldsymbol{N}$. It follows that $X = Q \cup \bigcup (\boldsymbol{N} \cup \mathbf{RED}_K(X))$. Clearly, $\boldsymbol{N} \cup \mathbf{RED}_K(X) \subseteq \mathbf{IND}_K$. □

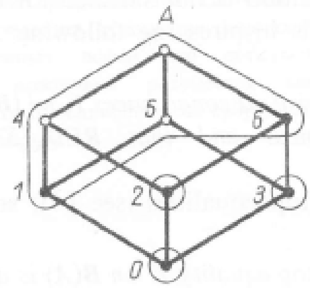
5.5. COROLLARY. *Any element in $\boldsymbol{B}(A)$ is K-accessible.* □

5.6. THEOREM. *For any $X \in \boldsymbol{B}(A)$ there exists an almost K-zero element Y such that $X = Y \cup \bigcup \boldsymbol{SRED}_K(X)$.*

Proof. By 5.1 there exists $\boldsymbol{M} \subseteq \mathbf{IND}_K$ and an almost K-zero element Y such that $X = Y \cup \bigcup \boldsymbol{M}$. Clearly $\boldsymbol{M} \subseteq \mathbf{IND}_K(X)$ which implies that $X = Y \cup \bigcup \boldsymbol{M} \subseteq Y \cup \bigcup \mathbf{IND}_K(X)$. Since $\mathbf{SRED}_K(X)$ is the set of all maximal elements in $\mathbf{IND}_K(X)$, for any $Q \in \mathbf{IND}_K(X)$ there exists $P \in \mathbf{SRED}_K(X)$ such that $Q \subseteq P$; it follows that $X \subseteq Y \cup \bigcup \mathbf{IND}_K(X) \subseteq Y \cup \bigcup \mathbf{SRED}_K(X) \subseteq X$ which implies that $X = Y \cup \bigcup \mathbf{SRED}_K(X)$. □

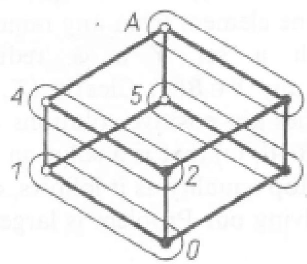
6. Examples. In the following examples, we investigate congruences on the semilattice $(\boldsymbol{B}(A), \cup)$ where $A = \{a, b, c\}$. For the sake of brevity, we put $0 = \emptyset$, $1 = \{a\}$, $2 = \{b\}$, $3 = \{c\}$, $4 = \{a, b\}$, $5 = \{a, c\}$, $6 = \{b, c\}$. In the following diagrams, the blocks of K and their minimal elements are marked.

6.1. EXAMPLE. The blocks of K are: $\{0\}$, $\{2\}$, $\{3\}$, $\{1, 4, 5, 6, A\}$. Then $\mathbf{IND}_K = \{0, 1, 2, 3, 6\}$. We have $\mathbf{RED}_K(5) = \{1\}$, $\mathbf{SRED}_K(5) = \{1, 3\}$,



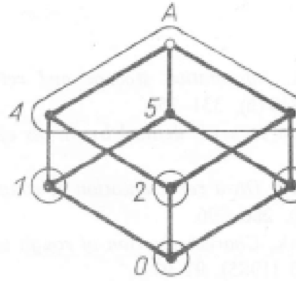
$\mathbf{CORE}_K(5) = 1$, $\bigcup \mathbf{SRED}_K(5) = 5$. Furthermore, $\mathbf{RED}_K(A) = \{1, 6\} = \mathbf{SRED}_K(A)$, $\mathbf{CORE}_K(A) = 0$, $\bigcup \mathbf{SRED}_K(A) = A$. □

6.2. EXAMPLE. The blocks of K are $\{0, 1\}$, $\{2, 4\}$, $\{3, 5\}$, $\{6, A\}$. Then $\mathbf{IND}_K = \{0, 2, 3, 6\}$. We have $\mathbf{RED}_K(5) = \{3\} = \mathbf{SRED}_K(5)$, $\mathbf{CORE}_K(5) = 3$,



Furthermore, 1 is an almost K-zero element and $5 = 1 \cup 3 = 1 \cup \bigcup \mathbf{SRED}_K(5)$. □

6.3. EXAMPLE. The blocks of K are $\{0\}$, $\{1\}$, $\{2\}$, $\{3\}$, $\{4, 5, 6, A\}$. Then $\mathbf{IND}_K = \{0, 1, 2, 3, 4, 5, 6\}$. Clearly, $\mathbf{RED}_K(A) = \{4, 5, 6\} = \mathbf{SRED}_K(A)$,



$\mathbf{RED}_K(X) = X = \mathbf{SRED}_K(X)$ for any $X \in \mathbf{B}(A) - \{A\}$. Furthermore, 0 is the only almost K-zero element. □

7. Problem. In 6.3, we have constructed a congruence K on $(\boldsymbol{B}(A), \cup)$ such that $\mathbf{RED}_K(X) = \mathrm{SRED}_K(X)$ for any $X \in \boldsymbol{B}(A)$ and that $\emptyset$ is the only K-zero element in $\boldsymbol{B}(A)$. This condition is not satisfied for congruences described in Examples 6.1 and 6.2. This inspires the following

7.1. PROBLEM. *Characterize all congruences K on $(\boldsymbol{B}(A), \cup)$ such that $\emptyset$ is the only almost K-zero element and that $\boldsymbol{RED}_K(X) = \boldsymbol{SRED}_K(X)$ for any $X \in \boldsymbol{B}(A)$.* □

We prove that rough top equalities (see [4]) solve our Problem.

7.2. THEOREM. *A rough top equality K on $\boldsymbol{B}(A)$ is a congruence on $(\boldsymbol{B}(A), \cup)$ such that $\emptyset$ is the only almost K-zero element and that $\boldsymbol{RED}_K(X) = \boldsymbol{SRED}_K(X)$ for any $X \in \boldsymbol{B}(A)$.*

Proof. There exists an equivalence R on A such that for any $X \in \boldsymbol{B}(A)$ and any $Y \in \boldsymbol{B}(A)$ the condiotion $(X, Y) \in K$ means that either $X \cap C \neq \emptyset$, $Y \cap C \neq \emptyset$ or $X \cap C = \emptyset$, $Y \cap C = \emptyset$ for any $C \in A/R$.

It follows that $X \in \mathbf{IND}_K$ if and only if $X \cap C$ has at most one element for any $C \in A/R$. If $X \in \boldsymbol{B}(A)$ is arbitrary, then $Y \in \boldsymbol{B}(A)$ is a subreduct of X if and only if Y contains exactly one element from any nonempty intersection $X \cap C$ where $C \in A/R$; but such a set Y is a reduct of X, too. Thus, $\mathbf{RED}_K(X) = \mathrm{SRED}_K(X)$ for any $X \in \boldsymbol{B}(A)$. Clearly, $(X, \emptyset) \in K$ implies $X = \emptyset$. □

Thus, rough top equalities are positive solutions of Problem. Example 6.3 presents a congruence on $(\boldsymbol{B}(A), \cup)$ that is a solution of our Problem, too; this congruence is not a rought top equality as it follows, e.g. from 3.8 of [4]. Thus, the class of congruences solving our Problem is larger than the class of rough top equalities.

INSTITUTE OF MATHEMATICS ČSAV, MENDLOVO NÁM. 1, 60300 BRNO (CZECHOSLOVAKIA)
DEPARTMENT OF COMPLEX CONTROL SYSTEMS, POLISH ACADEMY OF SCIENCES, BAŁTYCKA 5, 44-100 GLIWICE
(ZAKŁAD SYSTEMÓW AUTOMATYKI KOMPLEKSOWEJ PAN)

REFERENCES

[1] W. Marek, Z. Pawlak, *Information storage and retrieval systems – mathematical foundations*, Theor. Comp. Sci., 1 (1976), 331–354.

[2] M. Novotný, *Remarks on sequents defined by means of information systems*, Fundam. Inform., 6 (1983), 71–79.

[3] M. Novotný, Z. Pawlak, *On a representation of rough sets by means of information systems*, Fundam. Inform., 6 (1983), 286–296.

[4] M. Novotný, Z. Pawlak, *Characterization of rough top equalities and rough bottom equalities*, Bull. Pol. Ac.: Math., 33 (1985), 91–97.

[5] Z. Pawlak, *Information systems – theoretical foundations*, Inform. Syst., 6 (1981), 205–218.

М. Новотны, З. Павляк, **Независимость признаков**

Каждое множество признаков информационной системы определяет отношение эквивалентности на множестве предметов этой системы. Множество признаков называется независимым, если оно определяет отношение эквивалентности отличное от отношений эквивалентности определенных всеми его собственными подмножествами. Под субредуктом множества признаков разумеется максимальное его независимое подмножество. Доказано, что объединение всех субредуктов любого множества признаков „почти равно" этому множеству.