**Teoria zbiorów przybliżonych** (Zdzisław Pawlak, 1982)

Dana jest następująca **tablica decyzyjna**:

| **obiekt** | **C1** | **C2** | **C3** | **D** |
|---|---|---|---|---|
| o1 | a | 1 | + | B |
| o2 | a | 3 | - | A |
| o3 | a | 2 | + | A |
| o4 | b | 1 | - | B |
| o5 | a | 2 | + | A |
| o6 | b | 3 | + | B |
| o7 | a | 1 | + | A |

gdzie C1-C3 to **atrybuty warunkowe**, a D to **atrybut decyzyjny**, którego wartość określa przynależność do jednej z dwóch **klas decyzyjnych** (A lub B).

W ogólności, $C$ oznacza zbiór wszystkich atrybutów warunkowych, a $P \subseteq C$ jest rozważanym zbiorem atrybutów warunkowych.

Podstawowe pojęcia teorii zbiorów przybliżonych:

- **funkcja informacyjna** $f(x, a_i)$ – zwraca wartość obiektu $x \in U$ na atrybucie $a_i$.
- **relacja $\boldsymbol{P}$-nierozróżnialności** $I_P = \{(x, y) \in U \times U: f(x, a_i) = f(y, a_i)\ \forall a_i \in P\}$.
- **Zbiór $\boldsymbol{P}$-nierozróżnialny** (granula, atom) dla obiektu $x \in U$: $I_P(x) = \{y \in U: xI_Py\}$.
- **$\boldsymbol{P}$-dolne przybliżenie** zbioru $X$: $\underline{P}(X) = \{x \in U: I_P(x) \subseteq X\}$.
- **$\boldsymbol{P}$-górne przybliżenie** zbioru $X$: $\overline{P}(X) = \{x \in U: I_P(x) \cap X \neq \emptyset\}$.
- **$\boldsymbol{P}$-brzeg** zbioru $X$: $Bn_P(X) = \overline{P}(X) - \underline{P}(X)$.
- **Trafność przybliżenia** (ang. accuracy of approximation) zbioru $X$: $\alpha_P(X) = \frac{|\underline{P}(X)|}{|\overline{P}(X)|}$.
- Niech dany będzie zbiór klas decyzyjnych $\boldsymbol{Cl} = \{X_1, X_2, \dots, X_n\}$.

  **Jakość** (**przybliżenia**) **klasyfikacji** (ang. quality of (approximation of) classification) $\boldsymbol{Cl}$: $\gamma_P(\boldsymbol{Cl}) = \frac{\sum_{i=1}^{n}|\underline{P}(X_i)|}{|U|}$.
- **Redukt** zbioru atrybutów $P$ – każdy minimalny zbiór $R \subseteq P$: $\gamma_R(\boldsymbol{Cl}) = \gamma_P(\boldsymbol{Cl})$.
- **Rdzeń** (ang. core) zbioru atrybutów – część wspólna wszystkich reduktów.

Podstawowa własność zawierania (ang. rough inclusion property):

$\underline{P}(X) \subseteq X \subseteq \overline{P}(X)$.

Przykład.

Rozważmy zbiór atrybutów $P$ = {C1,C2,C3}.

$P$-dolne przybliżenie klasy decyzyjnej A: $\underline{P}(A) = \{o2, o3, o5\}$.

$P$-górne przybliżenie klasy decyzyjnej A: $\overline{P}(A) = \{o1, o2, o3, o5, o7\}$.

$P$-brzeg klasy A: $Bn_P(A) = \{o1, o7\}$.

Trafność przybliżenia (ang. accuracy of approximation) klasy A: $\frac{|\underline{P}(A)|}{|\overline{P}(A)|} = \frac{3}{5} = 0{,}6$.

$P$-dolne przybliżenie klasy decyzyjnej B: $\underline{P}(B) = \{o4, o6\}$.

$P$-górne przybliżenie klasy decyzyjnej B: $\overline{P}(B) = \{o1, o4, o6, o7\}$.

$P$-brzeg klasy B: $Bn_P(B) = \{o1, o7\}$.

Trafność przybliżenia (ang. accuracy of approximation) klasy B: $\frac{|\underline{P}(B)|}{|\overline{P}(B)|} = \frac{2}{4} = 0{,}5$.

Jakość klasyfikacji $\boldsymbol{Cl} = \{\mathrm{A}, \mathrm{B}\}$: $\gamma_P(\boldsymbol{Cl}) = (3+2)/7 = 5/7$ .

Redukty: {C1,C2}, {C2,C3}.

Rdzeń: {C2}.

Minimalny zbiór minimalnych reguł decyzyjnych (pewnych i przybliżonych):

| | |
|---|---|
| (C2=2) → (D=A) | pokrywa obiekty o3 i o5 |
| (C1=a) & (C2=3) → (D=A) | pokrywa obiekt o2 |
| (C1=b) → (D=B) | pokrywa obiekty o4 i o6 |
| (C1=a) & (C2=1) → (D=A) or (D=B) | pokrywa niespójne obiekty o1 i o7 |