

Krzysztof CZAJKOWSKI, Krzysztof BOBOWSKI
Politechnika Krakowska, Instytut Teleinformatyki, Wydział Fizyki,
Matematyki i Informatyki

ANALIZA DANYCH Z WYKORZYSTANIEM ZBIORÓW PRZYBLIŻONYCH W OCHRONIE ZABYTKÓW ARCHITEKTURY

Streszczenie. Artykuł omawia zagadnienia dotyczące zastosowania systemu analizy danych w zabezpieczaniu i konserwacji zabytków architektury. Z uwagi na bardzo dużą liczbę obiektów zabytkowych o silnie zróżnicowanym stanie zachowania, wykonanych z materiałów o różnej podatności na degradację, istotne jest opracowanie systemu wspierającego prace konserwatorskie. W prezentowanym artykule omówiono wykorzystanie teorii zbiorów przybliżonych do analizy danych na temat czynników środowiskowych, mających wpływ na obiekty zabytkowe. Zaprezentowano elementy systemu realizującego taką analizę oraz wyniki przeprowadzonych eksperymentów.

Słowa kluczowe: zabytek architektury, analiza danych, zbiory przybliżone, klasyfikacja

DATA ANALYSIS USING ROUGH SETS IN THE PROTECTION OF ARCHITECTURAL MONUMENTS

Summary. The article discusses issues concerning the application of data analysis system in the preservation and conservation of architectural monuments. Due to the extremely large number of historic buildings with highly differentiated state of preservation, made of materials with different susceptibility to degradation, it is important to develop a system to support the work of restoration. In the presented paper it has been discussed the use of rough sets theory to the analysis of data on environmental factors affecting historic buildings. The elements of the system performing this analysis and the results of experiments have been presented.

Keywords: monument of architecture, data analysis, rough sets, classification

1. Wstęp

Problem ochrony zabytków jest zagadnieniem bardzo złożonym, z uwagi na różnorodność obiektów klasyfikowanych jako zabytek, jak również konieczność stosowania różnych podejść, technik i rozwiązań w zależności od potrzeb. Zabytkiem, zgodnie z [16], nazywamy nieruchomość lub rzecz ruchomą, ich część lub zespoły, będące dziełem człowieka lub związane z jego działalnością i stanowiące świadectwo minionej epoki bądź zdarzenia, którego zachowanie leży w interesie społecznym ze względu na posiadaną wartość historyczną, artystyczną lub naukową.

Zabytki w ogólności można podzielić na trzy grupy:

- Zabytki nieruchome.
- Zabytki ruchome.
- Zabytki archeologiczne.

Zabytki nieruchome można z kolei podzielić na: krajobrazy kulturowe, układy urbanistyczne, zespoły budowlane, dzieła architektury i budownictwa, dzieła budownictwa obronnego, obiekty techniki (huty, kopalnie, elektrownie, inne zakłady przemysłowe), cmentarze, parki, ogrody oraz inne formy zaprojektowanej zieleni, miejsca upamiętniające wydarzenia historyczne lub działalność wybitnych osobistości i instytucji.

Na szczególną uwagę, pod kontem ochrony, zasługują właśnie zabytki klasyfikowane jako nieruchome. Jest to spowodowane faktem, że te obiekty, w przeciwieństwie do zabytków ruchomych, nie mogą podlegać ochronie stałej (umieszczeniu w muzeum, zadaszeniu itp.). Sama liczba tych obiektów powoduje, że problem ich monitoringu i ochrony jest szczególnie trudny i wymaga wsparcia ze strony nowoczesnych technologii. W Polsce, jako zabytki nieruchome zarejestrowanych jest ponad 64 673 obiektów, z czego 4 948 w samym tylko województwie małopolskim (stan na 04.10.2010) [13]. Ich liczba jednak w dalszym ciągu rośnie, czego przykładem może być 114 decyzji o wpisaniu do rejestru zabytków w samym tylko pierwszym kwartale 2009 roku. W związku ze złożonością problemu oraz obowiązkami, stojącymi przed podmiotami zarządzającymi zabytkami, konieczne jest wsparcie ze strony systemów informatycznych. Wsparcie takie możliwe jest na wielu poziomach, począwszy od systemów katalogujących, poprzez systemy informacji przestrzennej, aż po systemy wsparcia podejmowania decyzji odnośnie działań zabezpieczających i konserwatorskich. W ostatnim z wymienionych obszarów użyteczne mogą być aplikacje wykorzystujące elementy sztucznej inteligencji.

W artykule zaproponowano rozwiązanie wykorzystujące teorię zbiorów przybliżonych, zintegrowaną z systemami bazodanowymi w celu efektywnego klasyfikowania obiektów wymagających różnego typu prac konserwatorskich. W większości rozwiązań wykorzystują-

nych zbiory przybliżone, operacje przetwarzania danych wykonywane są z użyciem zwykłych plików „płaskich” bez wykorzystania, o wiele bardziej wydajnych w tym przypadku, systemów bazodanowych. Dodatkową wadą podejść tradycyjnych (wykorzystujących dedykowane aplikacje, tworzone w celu przeprowadzania operacji, związanych z teorią zbiorów przybliżonych) jest konieczność przenoszenia danych (często dużej ich ilości) z miejsca ich przechowywania (czyli właśnie bazy danych) do osobnych aplikacji. Najkorzystniejszym rozwiązaniem wydaje się być wykonywanie operacji wewnątrz systemu bazodanowego, a więc zaimplementowanie rozwiązania korzystającego ze zbiorów przybliżonych nie jako osobnej (wydzielonej) aplikacji, ale właśnie w ramach serwera bazodanowego. Do implementacji wykorzystano serwer firmy Oracle, z uwagi na fakt jego dużej popularności, a co za tym idzie dużej szansy na to, że właśnie w takich systemach gromadzone mogą być dane na temat obiektów zabytkowych. Systemy te posiadają również silne wsparcie dla rozwiązań typu GIS oraz przechowywania i przetwarzania multimedialnych.

2. Trwałość budowli

Degradacja budowli rozpoczyna się z momentem jej wzniesienia. Jej trwałość mierzona jest w dziesiątkach, a nawet setkach lat i zależy głównie od projektu, który powinien uwzględniać wymagania trwałości. Trwałość budowli jest rozumiana jako zdolność do spełnienia funkcji, jakie są od niej wymagane w danych warunkach użytkowania przez określony czas. Bardzo ważne zatem jest, aby obiekt był w taki sposób wykonany oraz użytkowany, żeby jego utrzymanie nie wiązało się z nadmiernymi kosztami. W praktyce niezbędne okazują się jednak kontrole postępu degradacji, które prowadzą do ustalenia odpowiednich metod konserwacji, szczególnie ważnych w przypadku budowli zabytkowych, których okres istnienia jest bardzo długi i bardzo często przekracza planowany (w momencie budowy) okres użytkowy. Należy wziąć też pod uwagę, iż w przypadku wymagań dotyczących trwałości znaczenie ma również estetyka budowli.

Głównymi czynnikami, które mają wpływ na stan budynków, są:

- przeznaczenie i sposób użytkowania,
- jakość materiałów i wzajemne ich oddziaływanie,
- korozja: biotyczna (działanie mikroorganizmów i pleśni), klimatyczna (czynniki środowiskowe związane z klimatem), elektrochemiczna (działanie prądów błędzących), chemiczna (działanie cieczy agresywnych i roztworów), mechaniczna (naprężenia zewnętrzne, zmęczeniowymi, ścieranie i uderzenie).

Budynki (w tym budowle zabytkowe) wznoszone są z elementów o bardzo różnorodnym składzie materiałowym. Dlatego, konieczne jest zwrócenie uwagi na zróżnicowaną podatność poszczególnych materiałów, a co za tym idzie poszczególnych elementów konstrukcji (i dekoracji) na niszczenie, na skutek działania różnych warunków środowiskowych. Najistotniejszymi materiałami stanowiącymi budulec poszczególnych elementów obiektów zabytkowych są materiały: kamienne, ceramiczne, metalowe oraz drewno.

3. Czynniki wpływające na niszczenie zabytków

Obiekty zabytkowe funkcjonują w określonym środowisku, które za pomocą odpowiednich norm podlega klasyfikacji, ze względu na intensywność agresywności. Wyróżnia się trzy stopnie agresywności środowiska: słaby, średni oraz silny. Stan fizyczny środowiska dzieli go na trzy rodzaje: gazowe, ciekłe oraz stałe. Natomiast oddziaływanie środowiska, według odpowiednich norm, może być chemiczne, biologiczne lub inne.

Tabela 1

Lista czynników atmosferycznych (w kolejności alfabetycznej)

Czynnik	Czynnik	Czynnik
Amoniak (NH ₃)	Kwas azotowy V (HNO ₃)	Sadze
Amoniakalne sole miedzi	Kwas chromowy (H ₂ CrO ₄)	Siarkowodór (H ₂ S)
Chlorek amonu (NH ₄ Cl)	Kwas fluorowodorowy (HF)	Sole siarczanowe (np. Mg ₂ SO ₄)
Chlorek cynkowy (ZnCl ₂)	Kwas humusowy	Thuszcze
Chlorek magnezu (MgCl ₂)	Kwas mlekowy (CH ₃ H ₆ O ₃)	Temperatura
Chlorek sodu (NaCl)	Kwas octowy (C ₂ H ₄ O ₂)	Tlen (O)
Chlorowodór (HCl)	Kwas siarkowy VI (H ₂ SO ₄)	Trójtlenek siarki (SO ₃)
Ciśnienie atmosferyczne	Ług potasowy (KOH)	Twardość wody
Ciśnienie wody	Ług sodowy (NaOH)	Wahania temperatur
Cykliczne zamarzanie	Mgła	Wiatr
Dwutlenek azotu (NO ₂)	Odczyn środowiska	Wielkość opadów
Dwutlenek siarki (SO ₂)	Oleje roślinne	Wilgotność powietrza
Dwutlenek węgla (CO ₂)	Podciąganie kapilarne	Wodór (H)
Jon amonowy (NH ₄)	Promieniowanie UV	Woda morska
Jon magnezu Mg ²⁺	Pył węglowy	

Biorąc pod uwagę informacje dotyczące niszczenia materiałów budowlanych oraz ich odporności, możliwe jest utworzenie listy czynników, które (w różnym stopniu) wpływają na budowle i są również pomocne przy określaniu ich stanu. W ramach artykułu lista ta posłuży do zbudowania tablicy decyzyjnej, która umożliwi klasyfikację obiektów pod kontem stopnia ich degradacji. Czynniki atmosferyczne przedstawione zostały w tabeli 1.

Pozostałe czynniki dotyczące obiektów zabytkowych (w porządku alfabetycznym) to:

- Istotność poszczególnych materiałów (Waga) – zastosowana w celu określenia stopnia, w jakim dany budulec wpływa na walory użytkowe oraz estetyczne budowli.
- Okres użytkowy – okres (w latach), w czasie którego obiekt był użytkowany zgodnie ze swoim przeznaczeniem.
- Procentowa zawartość poszczególnych materiałów – pozwala określić, jak duża powierzchnia ulegnie zniszczeniu przy korozji jednego z materiałów.
- Przeznaczenie – z przeznaczenia wynika przewidywany okres trwałości budynku.
- Rodzaj atmosfery – dostarcza szacunkowych informacji o stężeniu niektórych związków chemicznych.
- Rodzaj zastosowanych materiałów – różne rodzaje danego materiału wykazują odmienną odporność na poszczególne czynniki.
- Rodzaje czynników, przed którymi materiał został zabezpieczony.
- Sposoby impregnacji materiałów.
- Stopień występowania obwodów prądu – prąd ma wpływ na korozję metali.
- Stopień występowania roślinności – w przypadku gdy w środowisku znajduje się duża ilość roślinności i drzew może to wywierać wpływ na budynki poprzez rozrost korzeni i niszczenie fundamentów, a także sprzyjać rozwojowi bakterii oraz utrzymaniu wilgoci przez materiały budowlane.
- Wiek budynku.
- Zastosowane spoiwa – spoiwa mogą reagować z innymi materiałami budowlanymi obniżając ich odporność na korozję.

4. Zbiory przybliżone w środowisku bazodanowym

W teorii zbiorów przybliżonych dane można zaprezentować w postaci tablicy decyzyjnej, w której wiersze odpowiadają obiektom, a kolumny atrybutom tych obiektów. Formalnie tablicą decyzyjną nazywamy uporządkowaną piątkę [8]:

$$DT=(U,C,D,V,f), \quad (1)$$

gdzie: $C, D \subset A$; $A \subset U$; $C \neq \emptyset, D \neq \emptyset$; $C \cup D = A$; $C \cap D = \emptyset$, U jest niepustym skończonym zbiorem zwanym uniwersum tablicy decyzyjnej, elementy tego zbioru nazywamy obiektami, C zbiór atrybutów warunkowych, D zbiór atrybutów decyzyjnych, f funkcją decyzyjną, $V = \bigcup_{a \in A} V_a$, przy czym V_a nazywamy dziedziną atrybutu $a \in A$.

Zakładając, że $B \subseteq C$, zbiór atrybutów $Q (\subseteq B)$ nazywamy reduktem zbioru atrybutów B względem atrybutu decyzyjnego d (przy założeniu, że zbiór atrybutów decyzyjnych jest jed-

noelementowy), gdy zbiór atrybutów Q jest niezależny oraz $IND(B,d)=IND(Q,d)$. $IND(B,d)$, to relacja nierozróżnialności względem decyzji d generowana jest przez zbiór atrybutów B . Redukt jest najmniejszym zbiorem atrybutów, przy którym zostaje zachowana dotychczasowa klasyfikacja (rozróżnialność) obiektów. W tabeli decyzyjnej może występować więcej niż jeden redukt. Rdzeniem, oznaczanym jako $CORE(B,d)$, nazywamy zbiór wszystkich atrybutów niezbędnych (nieusuwalnych) w zbiorze B , ze względu na atrybut decyzyjny d . Rdzeń stanowi część wspólną wszystkich reduktów.

W prezentowanym podejściu podstawowe pojęcia z teorii zbiorów przybliżonych zostały zdefiniowane z wykorzystaniem pojęć z teorii baz danych. Celem takiego rozwiązania jest zastosowanie mechanizmów istniejących w bazach danych, dla uzyskania większej wydajności w wyznaczaniu atrybutów nieusuwalnych oraz reduktów względnych.

Wszystkie atrybuty nieusuwalne są niezbędnymi elementami każdego reduktu [1], tak więc kluczowym problemem jest możliwość efektywnego wyszukiwania takich atrybutów. W podejściu tradycyjnym, popularną metodą jest konstruowanie macierzy decyzyjnej, a następnie przeszukiwanie wszystkich pozycji w takiej macierzy, celem znalezienia pozycji, które mają tylko jeden atrybut [5]. Metoda ta cechuje się niesatysfakcjonującą wydajnością, szczególnie w przypadku systemów gromadzących bardzo duże ilości danych. Z kolei inne metody, niewykorzystujące macierzy decyzyjnej, np. [6], mają złożoność obliczeniową $O(mn \log n)$ (gdzie n to liczba wierszy, a m jest liczbą atrybutów). Prezentowane podejście ma złożoność $O(mn)$ [3] i jest realizowane bez potrzeby wyznaczania dolnego i górnego przybliżenia.

Oznaczając operacje relacyjną zliczania (Count) poprzez Card, operację projekcji (Projection) jako Π , a zbiór atrybutów decyzyjnych jako D , możemy zapisać, że $C_j \in C$ jest atrybutem nieusuwalnym, jeżeli zachodzi (2).

$$Card(\Pi(C - C_j + D)) \neq Card(\Pi(C - C_j)) \quad (2)$$

Za pomocą podstawowych operacji języka SQL można ustalić, czy atrybut jest atrybutem nieusuwalnym. Niezbędne jest wykonanie tylko dwóch projekcji: jednej na atrybutach $C - C_j + D$, a drugiej na atrybutach $C - C_j$. Jeżeli liczba wierszy w obu projekcjach jest różna, to dany atrybut jest atrybutem nieusuwalnym, w przeciwnym przypadku jest on zbędny (nie ma utraty informacji, po usunięciu atrybutu).

W celu wyznaczenia reduktów zbioru atrybutów warunkowych, można zdefiniować stopień zależności pomiędzy reduktem a zbiorem atrybutów decyzyjnych – jak w (3).

$$K(REDU, D) = \frac{Card(\Pi(REDU + D))}{Card(\Pi(C + D))} \quad (3)$$

Podzbiór zbioru atrybutów warunkowych $REDU(\subseteq C)$ jest reduktem zbioru C ze względu na zbiór atrybutów decyzyjnych D , jeżeli jest minimalnym zbiorem atrybutów, które mają taką samą zdolność klasyfikowania, jak cały zbiór atrybutów warunkowych (4).

$$K(REDU, D) = K(C, D) \quad \text{oraz} \quad K(REDU, D) \neq K(R', D), \forall R' \subset REDU \quad (4)$$

Jeżeli współczynnik $K(REDU, D)$ jest równy 1, to zbiór atrybutów decyzyjnych D całkowicie zależy od zbioru atrybutów warunkowych C . Przy $K(REDU, D) < 1$, mówimy o zależności częściowej (o stopniu $K(REDU, D)$) [2].

Idea redukcji atrybutów może być uogólniona przez wprowadzenie pojęcia istotności atrybutu. Nie jest stosowana dwuwartościowa skala: znaczący i nieznaczący. Zamiast tego, konkretnym atrybutom przypisane są wartości z przedziału $[0,1]$, wyrażające istotność tych atrybutów dla tablicy decyzyjnej. Istotność atrybutu może zostać określona przez efekt wywołany jego usunięciem z tablicy decyzyjnej [2].

Współczynnik istotności dla danego atrybutu C_j ze zbioru atrybutów C definiowano jako:

$$Merit(C_j, C, D) = 1 - \frac{Card(\Pi(C - C_j + D))}{Card(\Pi(C + D))} \quad (5)$$

Wartość tego współczynnika reprezentuje wkład wnoszony przez atrybut C_j do zależności pomiędzy C (atrybutami warunkowymi) a D (atrybutami decyzyjnymi). Większa wartość tego współczynnika oznacza, że jest on istotniejszy w rozpatrywanej tablicy decyzyjnej.

5. Atrybuty nieusuwalne i redukty względne

Przedstawiany algorytm [3] opiera się na operacjach istniejących w systemie bazodanowym, bez konieczności obliczania dolnego i górnego przybliżenia i cechuje się większą wydajnością od podejścia tradycyjnego.

Algorytm 1: Core Attribute Algorithm

Wejście: tablica decyzyjna $T(C, D)$

Wyjście: Core - zbiór atrybutów nieusuwalnych (rdzeń) dla tablicy T

1. Core = \emptyset
2. For each $A \in C$ {
 - If $Card(\Pi(C - A + D)) \neq Card(\Pi(C - A))$
 - Then Core = Core \cup A

Algorytm 1 można zapisać jako operację: $Card(\Pi(X))$, gdzie X może być C , $C-A$, $C-A+D$. Wykorzystując język SQL operacja ta może zostać wyrażona za pomocą polecenia SELECT:

```
SELECT COUNT(*) FROM (SELECT DISTINCT X FROM T);
```

Jak udowodniono w [3], algorytm 1 może być zaimplementowany przy złożoności obliczeniowej $O(mn)$, gdzie m to liczba atrybutów, natomiast n jest liczbą wierszy (zakładając wykorzystanie indeksów). Prawidłowość działania algorytmu potwierdzają wyniki uzyskane w systemach Rosetta [9] i RSES [10].

W powyższych rozważaniach zakładamy niewystępowanie danych niespójnych. W celu wykrycia, czy w zbiorze danych istnieją niespójności wystarczy sprawdzić, czy zachodzi równość $Card(\Pi(C)) = Card(\Pi(C+D))$. Jeżeli równość nie jest spełniona, oznacza to, że istnieją dane niespójne. Do ustalenia, które wiersze są niespójne wystarczy zapytanie:

```
SELECT * FROM T U WHERE EXISTS (SELECT 1 FROM T V WHERE (U.C=V.C) AND
(U.D<>V.D));
```

W ten sposób niespójności w danych mogą być wyeliminowane w czasie $O(n)$, przy n równym liczbie wierszy w tabeli.

W tabeli decyzyjnej może występować więcej niż jeden redukt. Znalezienie wszystkich reduktów w tabeli decyzyjnej jest zadaniem NP-trudnym [7]. Istnieją jednak zastosowania wystarczy znalezienie jednego z nich. Częścią wspólną wszystkich reduktów jest zbiór atrybutów nieusuwalnych (rdzeń).

$$CORE(C) = \bigcap REDU(C) \quad (6)$$

Każdy element rdzenia stanowi część każdego reduktu, dlatego też zbiór atrybutów nieusuwalnych jest najważniejszym podzbiorem zbioru atrybutów warunkowych [1]. Algorytm 2 (zachłanny) wyznacza redukt w procesach selekcji i eliminacji.

Algorytm 2: Wyznaczanie minimalnego podzbioru atrybutów (reduktu)

Wejście: Tabela decyzyjna $T(C, D)$

Wyjście: Minimalny podzbiór zbioru atrybutów (REDU)

- Uruchom Algorytm 1 w celu wyznaczenia atrybutów kluczowych CORE
- REDU = CORE
- AR = C - CORE
{Forward selection}
- WHILE $K(REDU, D) \neq K(C, D)$
 - For each $A \in AR$ oblicz współczynnik istotności (merit)
 - Posortuj malejąco atrybuty w AR w oparciu o wartość merit
 - Wybierz atrybut C_j z największą wartością merit (gdy kilka atrybutów ma taką samą wartość merit, wybierz ten z mniejszą liczbą kombinacji z atrybutami w REDU)
 - REDU = REDU \cup C_j , AR = AR - C_j
- ENDWILE
{Backward elimination}
- N = ||REDU||
- FOR j=0 to N-1 DO
 - IF $a_j \notin CORE$ THEN oblicz $K(REDU - a_j, D)$
 - IF $K(REDU - a_j, D) = K(REDU, D)$ THEN REDU = REDU - a_j
- ENDFOR

Pierwszym etapem działania Algorytmu 2 jest wykonanie operacji Algorytmu 1 (wyznaczanie elementów rdzenia) i jest on operacją mniej złożoną niż ustalenie atrybutów REDU.

Prezentowany algorytm cechuje się większą wydajnością niż inne podejścia, ponieważ przetwarzanie danych i wszystkie działania służące ustalaniu niezbędności elementu w zbiorze atrybutów decyzyjnych (z punktu widzenia możliwości klasyfikacyjnych) oraz wyliczanie współczynników istotności dla poszczególnych atrybutów odbywa się za pomocą operacji, pod kontrolą systemu zarządzania relacyjnymi bazami danych. Dodatkowo, wszystkie operacje wykonywane są w miejscu przechowywania danych, bez konieczności przenoszenia danych do innych środowisk.

6. Wyznaczanie reguł decyzyjnych

Na podstawie reduktu względnego możliwe jest wygenerowanie tabeli reduktu (reduct table). Następnie na jej podstawie możliwe jest zbudowanie klasyfikatora składającego się z reguł decyzyjnych. Reguła decyzyjna jest kombinacją wartości wybranych atrybutów warunkowych, takich, dla których zbiór wszystkich przypadków (obiektów), który ma takie wartości należy do tej samej klasy (ma taką samą wartość atrybutu decyzyjnego). Możemy zapisać regułę jako implikację:

$$r : G_{i1} = V_{i1} \cap G_{i2} = V_{i2} \cap \dots \cap G_{ik} = V_{ik} \rightarrow D = d_i \quad (7)$$

Celem procesu jest uzyskanie reguł, które reprezentują najbardziej ogólne wzorce występujące w zbiorze danych. W tym celu należy z danej reguły usunąć tak wiele atrybutów warunkowych, jak to możliwe, bez utraty poprawności reguły.

Kolejność rozpatrywania poszczególnych atrybutów warunkowych w danej regule decyzyjnej (pod kątem ich ewentualnego usunięcia) determinuje możliwość uzyskania możliwie najbardziej ogólnej reguły. Dla każdej reguły, mającej n atrybutów warunkowych, liczba możliwych podzbiorów takiego zbioru wynosi 2^n , tak więc rozpatrywanie wszystkich możliwych kombinacji jest niepraktyczne. Dla uproszczenia algorytmu możliwe jest określenie dodatkowego parametru SIG dla każdego atrybutu. Parametr ten będzie określał znaczenie danego atrybutu [4]. Proces usuwania atrybutów będzie przeprowadzany w kolejności rosnących wartości tego parametru. Parametr SIG dla danej wartości atrybutu warunkowego $C_{ik}=V_{ik}$ można zapisać: $SIG(C_{ik}=V_{ik})=P(C_{ik}=V_{ik}) \cdot (P(D=d_i|C_{ik}=V_{ik})-P(D=d_i))$, gdzie: $P(C_{ik}=V_{ik})$ to prawdopodobieństwo wystąpienia wartości V_{ik} dla atrybutu C_{ik} , a $P(D=d_i|C_{ik}=V_{ik})$ to prawdopodobieństwo warunkowe wystąpienia klasy $D=d_i$ pod warunkiem zaistnienia wartości atrybutu warunkowego $C_{ik}=V_{ik}$.

W celu uniknięcia generowania zbyt wyspecyfikowanych reguł możliwe jest wykorzystanie testu Laplace'a. Umożliwia to eliminację wielu reguł pokrywających niewielką liczbę przypadków. Preferowane są bardziej ogólne reguły.

$$Laplace = (n_c + 1) / (n_{total} + k), \quad (8)$$

gdzie: k to liczba klas, n_c to liczba krotek przewidywanej klasy c pokrywanych przez tę regułę, n_{total} to całkowita liczba krotek pokrywanych przez tę regułę.

Poniżej zaprezentowany jest algorytm generujący zestaw możliwie najbardziej ogólnych reguł na podstawie tablicy decyzyjnej oraz reduktu (wyznaczonego algorytmem 2) [4].

Algorytm 3: Wyznaczanie zestawu najbardziej ogólnych reguł

Wejście: Tablica decyzyjna $T(C,D)$, redukt REDU

Wyjście: Zestaw najbardziej ogólnych reguł MG_Rules

- Wygeneruj tabelę reduktu REULES za pomocą projekcji reduktu REDU oraz atrybutu decyzyjnego D na tablicy decyzyjnej $T(C,D)$
- Zbiór reguł MG_Rules = \emptyset
- For each $C_{ik}=V_{ik} \in RULES$ oblicz wartość parametru znaczenia SIG
- For each $r_i \in RULES$ uprość każdą krotkę:
 - o Posortuj zbiór atrybutów warunkowych w regule r_i bazując na SIG
 - o For each $C_{ik}=V$
 - Usuń wartość z r_i
 - Jeżeli r_i jest niespójne z dowolną inną regułą wówczas przywróć usuniętą wartość
 - o Usuń wszystkie reguły z MG_Rules, które logicznie zawierają się w regule r_i
 - o MG_Rules = MG_Rules \cup r_i
- For each $r_i \in MG_Rules$ oblicz wartość Laplace(r_i)
- For each $r_i \in MG_Rules$ If Laplace(r_i) < threshold Then MG_Rules = MG_Rules - r_i

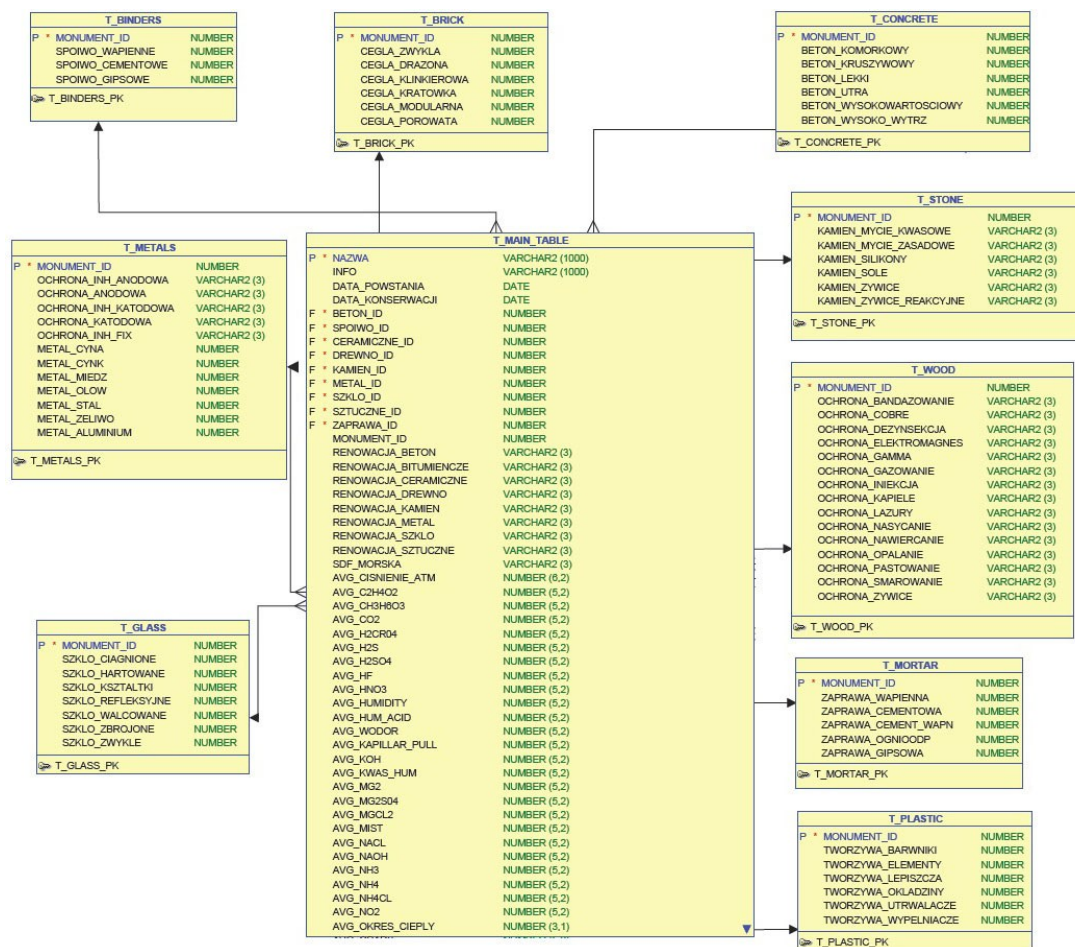
7. Aplikacja

W ramach prowadzonych badań opracowywany jest system RS Analyser For Historical Monuments, którego zadaniem będzie analiza wpływu czynników atmosferycznych na budowlę zabytkowe, na podstawie teorii zbiorów przybliżonych. Ma on umożliwiać przeprowadzanie analizy aktualnego stanu oraz stopnia zniszczenia zabytków zapisanych w bazie danych i informować o zagrożeniach dotyczących poszczególnych obiektów. Dzięki zastosowaniu teorii zbiorów przybliżonych, system poza procesem wprowadzania danych nie wymaga dodatkowych czynności ze strony użytkownika, zapewniając przy tym stabilność i niezawodność w analizowaniu danych. System zaimplementowany został z wykorzystaniem serwera Oracle. Dane niezbędne do analizy gromadzone są w bazie danych, której fragment schematu przedstawiony został na rys. 1, a opis tabel zamieszczono poniżej.

Dane dotyczące zabytku oraz środowiska, w jakim się znajduje zapisywane są do tabeli T_MAIN_TABLE, natomiast dane dotyczące poszczególnych materiałów, z jakich składa się dany zabytek, do tabel do tego przeznaczonych. Dla materiałów drewnianych i drewnopochodnych będzie to tabela T_WOOD, dla materiałów betonowych T_CONCRETE itd. Na każdej z tych tabel znajdują się wyzwalacze (np. TRG_MAIN_TABLE, TRG_WOOD, TRG_CONCRETE), których zadaniem jest zamiana wartości rzeczywistych lub tekstowych

na liczby naturalne. Odbywa się to na podstawie tabeli T_DICTIONARY, która przechowuje informacje o sposobie zamiany odpowiednich wartości. Następnie dane te zapisywane są do tabel, pełniących funkcję tablic decyzyjnych (np. T_STONE_DEC, T_WOOD_DEC itd.).

Tabele te posiadają również swoje wyzwalacze, których zadaniem jest klasyfikacja otrzymanych danych na podstawie reguł decyzyjnych. W przypadku gdy reguły takie nie są jeszcze wygenerowane (sytuacja ta występuje tylko przy pierwszym uruchomieniu systemu, gdy do bazy nie zostały załadowane dane), wyzwalacze te nie istnieją.



Rys. 1. Schemat bazy danych
Fig. 1. Database schema

Uruchomienie generowania reguł decyzyjnych, służących do klasyfikacji stopnia zniszczenia obiektów, powoduje włączenie czterech funkcji z zaimplementowanymi operacjami na zbiorach przybliżonych. Pierwsza z nich usunie z tablic decyzyjnych niespójności, następnie zostanie wyznaczony rdzeń, potem redukt i ostatecznie reguły decyzyjne. Po zakończeniu działania, w celu optymalizacji szybkości działania aplikacji, a także, aby umożliwić użytkownikowi późniejszy wgląd do procesu generacji, rdzenie, reduktu oraz reguły decyzyjne zostaną zapisane do odpowiednich tabel: T_CORES, T_REDUCTS, T_RULES. Ponadto, funkcja F_RULES uruchomi, dla każdej z tablic decyzyjnych, procedurę składowaną zaim-

plementowaną w języku Java, która na podstawie reguł decyzyjnych wygeneruje wyzwalacze, które klasyfikują obiekty do nich wstawiane. Od tej chwili, każdy nowy obiekt wprowadzony do bazy, a także rodzaje materiałów, z których się składa będą klasyfikowane za pomocą tych reguł pod względem stopnia zniszczenia. Warto zauważyć, że klasyfikacja odbywa się w określonym porządku. Najpierw przeprowadzana jest ona dla poszczególnych materiałów budowlanych, a następnie klasyfikacje te przenoszone są do tablicy decyzyjnej, dotyczącej budynku jako całości. Dopiero wówczas następuje klasyfikacja stopnia zniszczenia całego zabytku (rys. 2).



Numer zabytku	Nazwa	Informacje	Zniszczenie
1	Kościół w Anny	Kraków, Ul. w Anny 12	Bardzo duże
16	Muzeum Techniczne - Piżemysłowe	Kraków, Ul. Smoleńsk	Bardzo duże
5	Folterzy kopcu Kościuszki	Kraków	Bardzo duże
10	Kamienica, Ul. Szeroka 13	Kraków	Duże
11	Nowy Ratusz dawnego miasta Podgórze	Kraków, Rynek Podgórski 1/ul. Limanowski...	Duże
4	Synagoga wysoka	Kraków, Ul. Józefa 38	Duże
16	Pałac Pugetów	Kraków, Ul. Starowiślna 13	Duże
17	Kamienica Arcybractwa Miłosierdzia	Kraków, Ul. Stolarska 1a	Duże
19	Kramy Dominikańskie	Kraków, Ul. Stolarska 8-10	Średnie
6	Kamienica, ul. Bonerowska 6	Kraków,	Średnie
14	Hotel Saski	Kraków, Ul. Sławkowska 3	Średnie
7	Kamienica, ul. Bracka	Kraków,	Średnie
18	Pasaz Bielaka	Kraków, Ul. Stolarska 5; zespół zabudowy	Średnie
13	Grand Hotel	Kraków, Ul. Sławkowska 5-7	Małe
9	Kamienica Jordanów	Kraków, Ul. Szeroka 2/Szczeпаńska 11	Małe

Rys. 2. Okno aplikacji

Fig. 2. Application window

Możliwa jest weryfikacja wszystkich kolejnych etapów całego procesu klasyfikacji, w tym m.in. generowanych reduktów (rys. 3) i reguł decyzyjnych.



Redukt	Tabela
C2H4O2,AVG_CO2,AVG_H2S,AVG_HNO3,AVG_WODOR,AVG_KOH,AVG_KWAS_HUM,AVG_MG2,AVG_MG2SO4,AVG_NACL,AVG_NH3...	T_MAIN_TABLE_DEC
WO_WAPIENNE,SPOIWO_CEMENTOWE,SPOIWO_GIPSOWE,AVG_H2S,AVG_HF,AVG_HUMIDITY,AVG_WODOR,AVG_MG2SO4,AVG_M...	T_BINDERS_DEC
EN_MYCIE_KWASOWE,KAMIEN_MYCIE_ZASADOWE,KAMIEN_SILIKONY,KAMIEN_ZYWICE,KAMIEN_ZYWICE_REAKCYJNE,AVG_CH3...	T_STONE_DEC
LA_ZWYKLA,CEGLA_DRAZONA,CEGLA_KRATOWKA,CEGLA_MODALARNA,AVG_H2CR04,AVG_H2S,AVG_HF,AVG_HNO3,AVG_HUM...	T_BRICK_DEC
DN_KRUSZYWOWY,BETON_LEKKI,BETON_UTRA,BETON_WYSOKO_WYTRZ,AVG_C2H4O2,AVG_H2CR04,AVG_H2SO4,AVG_HF,AVG...	T_CONCRETE_DEC
O_CIAGNIONE,SZKLO_HARTOWANE,SZKLO_KSZTALTKI,SZKLO_REFLEKSYJNE,SZKLO_ZBROJONE,SZKLO_ZWYKLE,AVG_C2H4O...	T_GLASS_DEC
RONA_ANODOWA,OCHRONA_KATODOWA,OCHORNA_INH_FIX,METAL_CYNA,METAL_CYNYK,METAL_OLOW,METAL_ALUMINIUM,AVG...	T_METAL_DEC

Rys. 3. Redukty

Fig. 3. Reducts

8. Podsumowanie

W artykule zaprezentowano propozycję rozwiązania mającego na celu wsparcie prac zabezpieczających oraz konserwatorskich, dotyczących budowli zabytkowych. Podejście wykorzystuje teorię zbiorów przybliżonych do tworzenia klasyfikatorów umożliwiających klasyfikację stanu poszczególnych obiektów. Brane są pod uwagę różne czynniki środowiskowe, przede wszystkim atmosferyczne oraz podatność poszczególnych materiałów budowlanych, z jakich składają się konkretne budynki, na degradację. Na podstawie tablicy decyzyjnej, opisującej grupy obiektów, wyznaczane są atrybuty nieusuwalne (rdzeń) oraz redukty. Bazując na wyznaczonym redukcje tworzone są reguły decyzyjne dokonujące klasyfikacji obiektów

w systemie. Całość obliczeń wykonywana jest w ramach bazy danych, wykorzystując ich mechanizmy, bez potrzeby przenoszenia danych do odrębnego systemu, w ten sposób zwiększając wydajność całego procesu. Proponowane rozwiązanie może w znaczący sposób usprawnić i częściowo zautomatyzować prace mające na celu ochronę zabytków architektury. Zbieranie danych w sposób cykliczny, na podstawie instancji monitorujących stan zanieczyszczeń oraz czynniki atmosferyczne, umożliwić może w przyszłości wyznaczanie kolejności prac konserwatorskich oraz określanie priorytetów w konkretnych zadaniach zabezpieczających.

BIBLIOGRAFIA

1. Pawlak Z.: Some Issues on Rough Sets. Transactions on Rough Sets I, LNCS, Springer, 2004, s. 1÷58.
2. Komorowski J., Pawlak Z., Polkowski L., Skowron A.: A Rough Set Perspective on Data and Knowledge. [in:] Pal S. K., Skowron A. (eds.): Rough Fuzzy Hybridization Springer-Verlag, 1999, s. 107÷121.
3. Hu X., Lin T., Han J.: A New Rough Sets Model Based on Database Systems. Fundamenta Informatica. 59, IOS Press 2004, s. 135÷152.
4. Hu X.: Using rough sets theory and database operations to construct a good ensemble of classifiers for data mining applications. Proceedings IEEE International Conference ICDM 2001, s. 233÷240.
5. Cercone N., Ziarko W., Hu X.: Rule discovery from databases: A Decision matrix approach. Proc. ISMIS'96, Zakopane 1996, s. 653÷662.
6. Nguyen S. H., Nguyen H. S.: Some efficient algorithms for rough set methods. Proceedings of the Sixth International Conference Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU-96), Granada, Spain 1996, s. 1541÷1457.
7. Skowron A., Rauszer C.: The Discernibility Matrices and Functions in Information Systems. K. Slowinski (ed.): Intelligent Decision Support – Handbook of Applications and Advances of the Rough Sets Theory. Kluwer, Dordrecht 1992, s. 331÷362.
8. Mrózek A., Płonka L.: Analiza danych metodą zbiorów przybliżonych – Zastosowania w ekonomii, medycynie i sterowaniu. Akad. Oficyna Wyd. PLJ, Warszawa 1999.
9. A Rough Set Toolkit for Analysis of Data – <http://www.lcb.uu.se/tools/rosetta/>
10. Rough Set Exploration System – <http://logic.mimuw.edu.pl/~rses/>
11. Czajkowski K., Drabowski M.: Wybrane zagadnienia integracji zbiorów przybliżonych i baz danych. Studia Informatica, Vol. 30, No. 2A(83), Wyd. Pol. Śląskiej, Gliwice 2009, s. 355÷372.

12. Czajkowski K., Drabowski M.: Relational database and core, relative reducts in rough sets. Proceedings of the IASTED International Conference on Artificial Intelligence and Soft Computing, Palma de Mallorca, ACTA Press, Anaheim, USA 2009, s. 14÷20.
13. Krajowy Ośrodek Badań i Dokumentacji Zabytków <http://www.kobidz.pl/>
14. Doehne E., Price C. A.: Stone Conservation. An Overview of Current Research. The Getty Conservation Institute, Los Angeles 2010.
15. Sidraba I.: New Materials for Conservation of Stone Monuments In Latvia. Centre for conservation and Restoration of Stone Materials, Institute of Silicate Materials, Riga Technical University, 2003.
16. Ustawa z dnia 23 lipca 2003 r. o ochronie zabytków i opiece nad zabytkami. DzU 2003 nr 162 poz. 1568, <http://isap.sejm.gov.pl/DetailsServlet?id=WDU20031621568>

Recenzenci: Dr hab. inż. Adam Pelikant, prof. Pol. Łódzkiej
Prof. dr hab. inż. Bolesław Pochopień

Wpłynęło do Redakcji 31 stycznia 2011 r.

Abstract

The problem of protection of monuments is very complex, given the diversity of objects classified as monuments, as well as the need for different approaches, techniques and solutions depending on the needs. Monuments classified as immovable deserve special attention as far as protection is taken into consideration. This is due to the fact that these objects, as opposed to the movable monuments can not be protected solidly (placed in the museum, roofing, etc.). The sheer number of these objects makes that the problem of their monitoring and protection is particularly challenging and requires the assistance of modern technology. In Poland, in 2010, still registered as historical monuments is 64 673 objects, 4 948 of which in the Malopolska province.

The paper presents a proposal for a solution to support the work of protection and preservation of historical buildings. The approach uses the theory of rough sets to build classifiers to allow classification of individual objects. Various environmental factors are taken into account, primarily weather, and susceptibility of various construction materials, which concrete buildings are consisted of, to degradation. On the basis of decision table describing the group of objects, core attributes and reducts are determined. Basing on the designated reduct decision rules are created that make the classification of objects in the system. The whole calcula-

tion is performed within the database, using their mechanisms, without the need to transfer data to a separate system, thereby increasing the efficiency of the whole process.

The proposed solution can significantly streamline and partially automate the work to protect architectural monuments. Collecting data in a cyclical manner, based on the stances monitoring the state of the pollution and the weathering, may allow the determination of sequencing of conservation works and the prioritization of specific protection tasks in future.

Adresy

Krzysztof CZAJKOWSKI: Politechnika Krakowska, Wydział Fizyki, Matematyki i Informatyki, Instytut Teleinformatyki, ul. Warszawska 24, 31-155 Kraków, Polska, kc@pk.edu.pl.

Krzysztof BOBOWSKI: Politechnika Krakowska, Wydział Fizyki, Matematyki i Informatyki, Instytut Teleinformatyki, ul. Warszawska 24, 31-155 Kraków, Polska, krzysztofbobowski@gmail.com.