



Artificial Intelligence Laboratory

ZINTEGROWANY PAKIET SZTUCZNEJ INTELIGENCJI SPHINX[®] 4.0

KRZYSZTOF MICHALIK

DETREEX 4.0

dla Windows 9x/NT/2000

**INDUKCYJNY SYSTEM
POZYSKIWANIA WIEDZY**

PODRĘCZNIK UŻYTKOWNIKA

KATOWICE 2003



Artificial Intelligence Laboratory

ul. Kossutha 7, 40-844 KATOWICE
tel.: (0-32) 254-41-01 w. 374, tel./fax: w. 350
tel. kom. 0 502-99-27-28
e-mail: aitech@aitech.com.pl
WWW: <http://www.aitech.com.pl>

Copyright ©1990-2003 AITECH & Krzysztof Michalik

AITECH, Sphinx, CAKE oraz **Neuronix**
są prawnie zastrzeżonymi znakami towarowymi firmy
AITECH, ARTIFICIAL INTELLIGENCE LABORATORY

SPIS TREŚCI

Rozdział 1. WPROWADZENIE.....	1-1
CO TO JEST „UCZENIE MASZYNOWE”?	1-3
CO TO JEST POZYSKIWANIE WIEDZY?	1-3
METODY POZYSKIWANIA WIEDZY	1-5
INDUKCJA DRZEW DECYZYJNYCH.....	1-6
Rozdział 2. CZYM JEST DETREEX?	2-1
ZASTOSOWANIA DETREEXA	2-3
UŻYTKOWNICY DETREEXA	2-3
Rozdział 3. PRZYKŁAD DRZEWY DECYZYJNEGO	3-1
KREDYTY KONSUMPCYJNE	3-3
Rozdział 4. OBSŁUGA SYSTEMU DETREEX	4-1
PIERWSZE URUCHOMIENIE	4-3
MENU GŁÓWNE DETREEXA	4-3
Menu „Plik”	4-4
Menu „Drzewo decyzyjne”	4-4
Menu „Pomoc”	4-5
FORMAT PLIKU UCZĄCEGO I TESTOWEGO	4-5
UŻYTKOWANIE DETREEXA	4-8
Wskazanie pliku uczącego.....	4-8
Generowanie drzewa decyzyjnego	4-9
Wizualizacja i zapis drzewa decyzyjnego	4-9
Wizualizacja dziedziny problemu.....	4-11
Testowanie drzewa decyzyjnego	4-12
Zapis reguł z drzewa decyzyjnego.....	4-12
Uruchamianie innych systemów pakietu SPHINX	4-13
Opcje generowania drzewa decyzyjnego.....	4-13
WSPÓŁPRACA SYSTEMU DETREEX Z BAZAMI DANYCH.....	4-15
Import przykładów uczących z bazy danych.....	4-15
Rozdział 5. PRZYKŁADY DZIEDZINOWE	5-1
CHOROBA SERCA	5-3
KWIATY IRYSA.....	5-5
ZATRUDNIENIE	5-6
RODZAJE SZKŁA.....	5-7
Rozdział 6. LITERATURA.....	6-1

WPROWADZENIE

W rozdziale:

- Co to jest „uczenie maszynowe”?
- Co to jest pozyskiwanie wiedzy?
- Metody pozyskiwania wiedzy
- Indukcja drzew decyzyjnych

CO TO JEST „UCZENIE MASZYNOWE”?

Jednym z atrybutów inteligencji jest możliwość uczenia się. Dlatego dość wcześnie (w latach 60-tych XX wieku) rozpoczęły się prace naukowców nad automatyzacją tego procesu oraz czyniono pierwsze próby „nauczenia” komputera myślenia. Naukowcy opracowali kilka sposobów „maszynowego uczenia” (ang. *machine learning*), które w dużej mierze wzorują się na podobnych mechanizmach u ludzi.

Do zadań systemów komputerowych wyposażonych w umiejętność uczenia się należy zaliczyć:

- formułowanie nowych pojęć,
- wykrywanie nieznanych dotychczas prawidłowości w danych,
- tworzenie drzew i reguł decyzyjnych,
- przyswajanie nowych pojęć i struktur drogą uogólniania przykładów i analogii,
- modyfikowanie, uogólnianie i precyzowanie danych,
- zdobywanie wiedzy drogą konwersacji z ludźmi,
- generowanie wiedzy i wyjaśnień zrozumiałych dla użytkownika.

Badania w dziedzinie „uczenia maszynowego” odnoszą się do tworzenia programów komputerowych zdolnych pozyskiwać nową wiedzę i/lub ulepszać wiedzę już pozyskaną na podstawie pewnych informacji wejściowych. Informacje wejściowe takie jak przykłady, fakty, opisy itp. są zwykle wprowadzane na wejście systemu przez użytkownika.

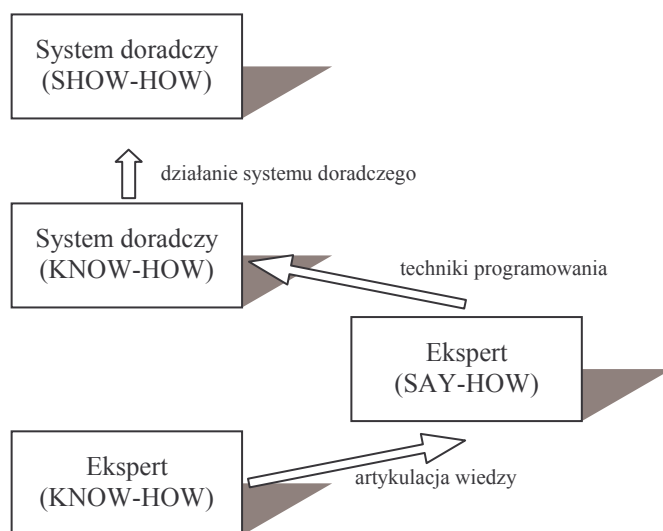
CO TO JEST POZYSKIWANIE WIEDZY?

Pozyskiwanie wiedzy (ang. *knowledge acquisition*) jest ściśle związane z uczeniem maszynowym i najogólniej rozumiane jest jako uczenie (pozyskiwanie informacji symbolicznej) połączone z nabyciem umiejętności efektywnego wykorzystania wiedzy [Cholewa, Pedrycz, 1987].

Wraz z rozwojem badań w zakresie budowy systemów doradczych pojawiły się granice tego rozwoju wynikające w szczególności z konieczności tworzenia coraz większych baz wiedzy. Konieczne okazało się opracowanie takich metod pozyskiwania wiedzy, które umożliwiłyby skrócenie czasu budowy i weryfikacji baz wiedzy pod względem ich zupełności, niesprzeczności oraz eliminacji nadmiarowości informacji.

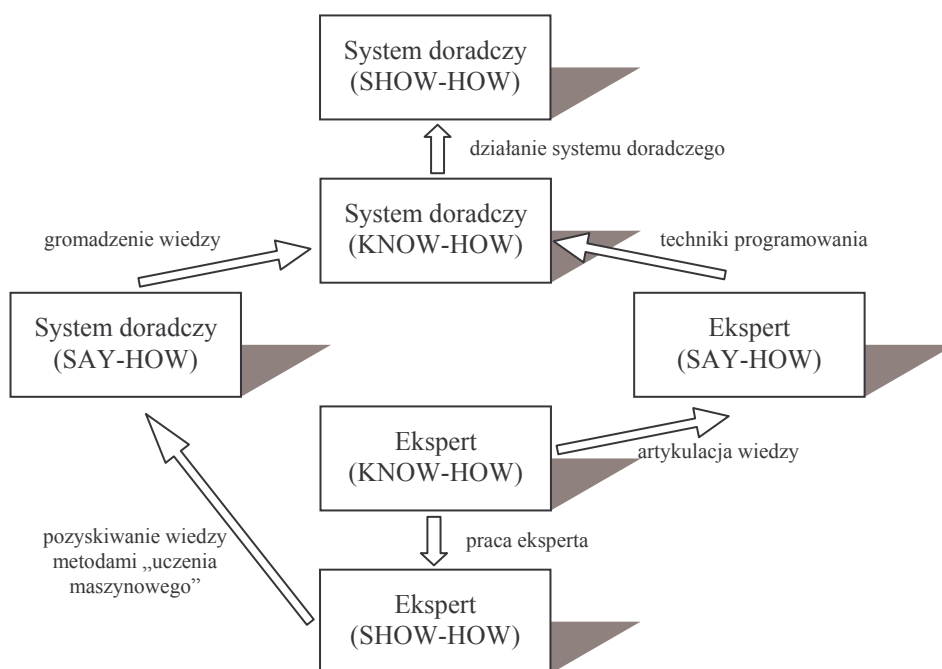
Rys. 1 przedstawia „tradycyjny” proces pozyskiwania wiedzy od eksperta, który utrudnia budowę, weryfikację i kontrolę nad złożonymi bazami wiedzy ze względu na niejednokrotne problemy w artykulacji wiedzy przez eksperta, błędy w zapisie wiedzy, itp. Najczęściej stosowane techniki pozyskiwania wiedzy od ekspertów to:

- konsultacje z ekspertem,
- zapis wiedzy przez eksperta np. w specjalnym formularzu („papierowym” i/lub „elektronicznym”),
- obserwacja i analiza pracy eksperta.



Rys. 1. „Tradycyjny” proces pozyskiwania wiedzy (na podst. [Michie, 1987])

W celu wspomagania procesu pozyskiwania wiedzy od ekspertów opracowano wiele metod umożliwiających „automatyczne-komputerowe” pozyskiwanie wiedzy, w których to metodach, ekspert nie bierze już bezpośredniego udziału. Zdobyta nowymi metodami wiedza pozyskana jest na podstawie wyników pracy eksperta(ów) oraz zgromadzonych wcześniej danych. Rys. 2 pokazuje proces pozyskiwania wiedzy od eksperta wspomagany metodami „uczenia maszynowego”.

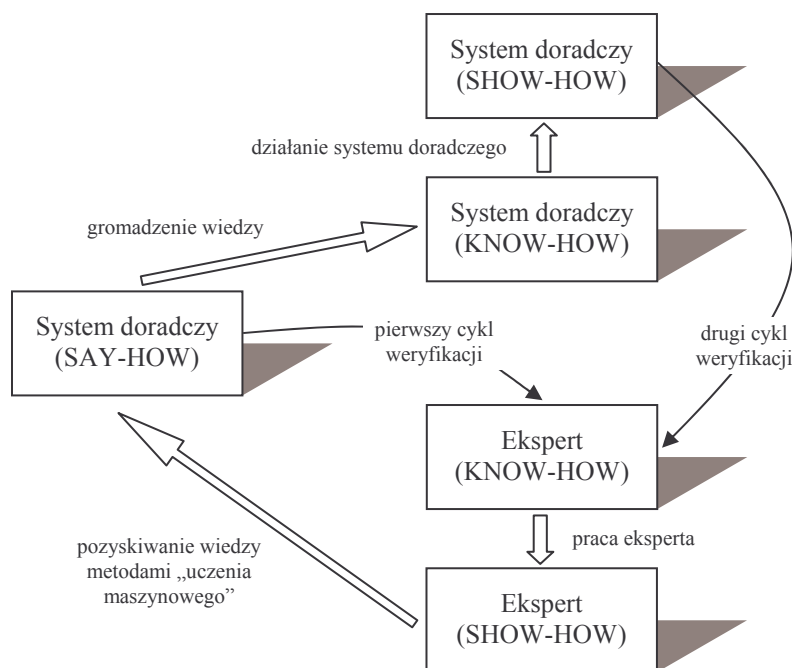


Rys. 2. Proces pozyskiwania wiedzy od eksperta wspomagany metodami „uczenia maszynowego” (na podst. [Michie, 1987])

Kolejny rysunek (Rys. 3) ukazuje proces pozyskiwania wiedzy, w którym ekspert nie przekazuje już swojej wiedzy, lecz dokonuje weryfikacji wiedzy pozyskanej metodami „uczenia maszynowego” oraz weryfikuje poprawność działania systemu doradczego (otrzymywanych wniosków i objaśnień). W przypadku stwierdzenia błędów i innych nieprawidłowości, ekspert wskazuje na konieczność dokonania poprawek w bazie wiedzy lub nawet ponownego pozyskania wiedzy na podstawie poprawionych i uzupełnionych danych.

Przedstawiony proces jest wydajniejszy od przedstawionych wcześniej, ponieważ:

- pozyskanie wiedzy metodami „uczenia maszynowego” jest znacznie szybsze,
- bazy wiedzy pozyskane „automatycznie” zawierają dużo mniejszą liczbę błędów lub ich nie zawierają w ogóle,
- weryfikacja wiedzy (pozyskanej metodami „uczenia maszynowego”) przez eksperta jest dla niego zadaniem prostszym i szybszym niż artikulacja tej wiedzy.



Rys. 3. Proces pozyskiwania wiedzy wspomagany metodami „uczenia maszynowego”
(na podst. [Michie, 1987])

METODY POZYSKIWANIA WIEDZY

Podstawową ideą uczenia się jest zdobywanie wymaganej wiedzy z zastosowaniem kilku metod rozumowania – indukcji, dedukcji lub analogii. W szczególnym przypadku uczenie może wymagać tylko powielania informacji dostarczonych przez otoczenie lub transformację tej informacji i/lub wydzielenie z niej pewnej istotnej dla nas części. Proces uczenia (pozyskiwania wiedzy) sklasyfikowany jest zależnie od wielu kryteriów. Jednym z takich kryteriów może być ilość informacji przekazanej do systemu doradczego. Można wyróżnić tutaj następujące metody uczenia (pozyskiwania wiedzy):

- bezpośrednie zapisanie wiedzy* (tzw. uczenie na pamięć) – nie wymaga od systemu podlegającego uczeniu wnioskowania i/lub transformacji wiedzy; realizowane np. przez bezpośrednie zaprogramowanie; metoda stosowana dla prostych baz wiedzy,
- pozyskiwanie wiedzy na podstawie instrukcji* (tzw. uczenie przez przekazanie informacji) – wymaga konieczności współdziałania uczącego z uczącym się; realizowane poprzez zastosowanie odpowiednich źródeł wiedzy wskazanych przez uczącego i ich przekształcenia na język akceptowalny przez uczącego się,
- pozyskanie na podstawie analogii* – polega na takim przekształceniu istniejącej informacji aby mogła być użyta do opisów faktów podobnych do tych, które już zostały zawarte w bazie wiedzy; realizowane przez np.: modyfikację programu komputerowego,
- pozyskiwanie wiedzy na podstawie przykładów* – metoda bardzo często stosowana przy konstruowaniu baz wiedzy; polega na generowaniu ogólnego opisu klas na podstawie zbioru przykładów i kontrprzykładów reprezentujących te klasy; ogólny opis otrzymywany jest na podstawie zasady indukcji,

- e) *pozyskiwanie wiedzy na podstawie obserwacji* (tzw. uczenie bez nauczyciela) – metoda ta wymaga większego udziału uczącego się podczas procesu uczenia; uczący może dokonywać obserwacji biernych oraz czynnych.

Śród wielu metod indukcji wiedzy na podstawie przykładów (patrz punkt d) opracowanych przez jednostki naukowe i dydaktyczne na całym świecie, wyróżnić można następujące metody:

- indukcja reguł za pomocą generowania pokryw [Michalski, 1983],
- indukcja drzew decyzyjnych [Quinlan, 1993],
- indukcja reguł za zastosowaniem zbiorów przybliżonych (np. [Øhrn, Komorowski, Skowron, Synak, 1998]).

W celu zastosowania tych metod konieczne jest użycie tzw. atrybutowego modelu opisu przestrzeni zainteresowania, czyli określenie tzw. dziedziny. Aby możliwe było pozyskanie wiedzy na podstawie przykładów konieczne jest wyszczególnienie:

- obiektów (zjawisk, problemów, itp.) podlegających klasyfikacji,
- atrybutów opisujących dane obiekty (zjawiska, problemy, itp.),
- oraz wartości jakie przyjmują poszczególne atrybuty.

Wartości atrybutów dla wybranej grupy obiektów tworzą tzw. zbiór uczący, a przykłady opisujące obiekty nazywa się przykładami uczącymi.

Poniżej szczegółowo opisano bardzo efektywną metodę indukcji drzew decyzyjnych, na podstawie której zbudowano system DeTreeX.

INDUKCJA DRZEW DECYZYJNYCH

Formalna definicja drzewa decyzyjnego brzmi:

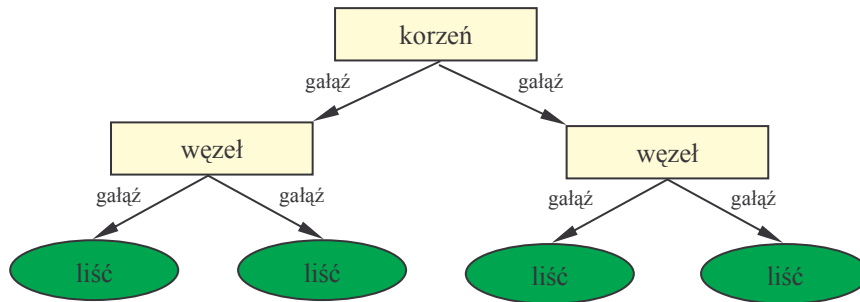
Drzewem decyzyjnym nazywamy strukturę drzewiastą, której każdy węzeł odpowiada przeprowadzeniu pewnego testu na wartości jednego atrybutu, zaś każdy liść zawiera decyzję o klasyfikacji przykładu. Z poszczególnych węzłów wychodzi tyle gałęzi, ile jest możliwych wyników testu odpowiadających tym węzłom. Każda z tych gałęzi prowadzi do poddrzewa (węzła) służącego do klasyfikacji tych obiektów, dla których ten test ma określony wynik.

Rys. 4 przedstawia przykład struktury drzewa decyzyjnego.

Głównym problemem w trakcie budowy drzewa decyzyjnego jest określenie kryterium, umożliwiającego wybór atrybutu stosowanego do rozbudowy tego drzewa. W tym przypadku stosuje się tzw. *entropię*. Informacja zawarta w zbiorze przykładów uczących (entropia) jest równa [Quinlan, 1993]:

$$I(E) = - \sum_{i=1}^{|E|} \frac{|E_i|}{|E|} \cdot \log_2 \left(\frac{|E_i|}{|E|} \right) \quad (1)$$

gdzie: E – zbiór przykładów uczących,
 $|E_i|$ – liczba przykładów, które opisują i -ty obiekt,
 $|E|$ – liczba przykładów w zbiorze uczącym E .



Rys. 4. Graficzna prezentacja struktury drzewa decyzyjnego z nazwami poszczególnych elementów [Wyleżoł, 1996]

Oczekiwana wartość informacji po podziale zbioru przykładów E na podzbiory $E^{(m)}$, $m = 1, \dots, |V_a|$, dla których atrybut a ma wartość v_m , określona jest jako [Quinlan, 1993]:

$$I(E, a) = \sum_{m=1, K, |V_a|, E^{(m)} \neq \emptyset} \frac{|E^{(m)}|}{|E|} \cdot I(E^{(m)}) \quad (2)$$

gdzie: $|E^{(m)}|$ – liczba przykładów po podziale zbioru E względem wartości m danego atrybutu,
 $|E|$ – liczba przykładów w zbiorze uczącym E .

W celu wyboru atrybutu, który będzie przypisany do tworzonego węzła drzewa decyzyjnego stosuje się kryterium względnego maksymalnego przyrostu informacji spowodowanego zastosowaniem danego atrybutu [Quinlan, 1993] (jako kolejny atrybut stosuje się ten, dla którego funkcja kryterialna ma wartość największą):

$$\Delta I(E, a) = I(E) - I(E, a) \quad (3)$$

Poniżej przedstawiono prosty przykład (na podst. [Quinlan, 1986]) ilustrujący zasadę budowy drzewa decyzyjnego stosując powyższe zależności.

Dany jest zbiór obiektów, do którego należą dwa obiekty oznaczone jako „-” oraz „+”. Rozpatrzmy zagadnienie klasyfikacji tych obiektów, przy czym uwzględnionymi atrybutami są: pogoda, temperatura, wilgotność oraz wiatr. Atrybuty te przyjmują następujące wartości:

pogoda:	słońce, zachmurzenie, deszcz
temperatura:	niska, średnia, wysoka
wilgotność:	wysoka, normalna
wiatr:	tak, nie

Tabela 1 przedstawia niewielki zbiór przykładów uczących. Przykłady te dotyczyć mogą np. problemu gry w golfa w zależności od warunków atmosferycznych – „Gramy w golfa” (obiekty „+”) lub „Zostajemy w domu” (obiekty „-”).

W celu wyboru atrybutu, od którego zaczyna się budowę drzewa, oblicza się wartość $I(E)$ dla całego zbioru uczącego. W zbiorze tym 9 przykładów przypisano do „+” a 5 do „-”. Zgodnie z zależnością (1) otrzymano:

$$I(E) = -\frac{9}{14} \log_2 \frac{9}{14} - \frac{5}{14} \log_2 \frac{5}{14} = 0,940 [\text{bit}].$$

Następnie oblicza się wartości $\Delta I(E, a)$ dla każdego atrybutu: atrybut „pogoda” o wartościach {słońce, zachmurzenie, deszcz}:

5 przykładów z całego zbioru posiada wartość „słońce”, przy czym 2 należą do „+” a 3 do „-”, więc

$$„+” = 2, „-” = 3, I(E_{s\acute{l}o\acute{n}ce}) = -2/5 \log_2 2/5 - 3/5 \log_2 3/5 = 0,971 \text{ [bit]}$$

i dla kolejnych wartości atrybutu „pogoda” podobnie:

$$„+” = 4, „-” = 0, I(E_{zachmurzenie}) = -4/5 \log_2 4/5 - 0/5 \log_2 0/5 = 0,0 \text{ [bit]},$$

$$„+” = 3, „-” = 2, I(E_{deszcz}) = -3/5 \log_2 3/5 - 2/5 \log_2 2/5 = 0,971 \text{ [bit]}.$$

$$I(E, pogoda) = \frac{5}{14} I(E_{s\acute{l}o\acute{n}ce}) + \frac{4}{14} I(E_{zachm.}) + \frac{5}{14} I(E_{desz.}) = 0,694 \text{ [bit]}$$

$$\Delta I(E, pogoda) = I(E) - I(E, pogoda) = 0,971 - 0,694 = \underline{0,246} \text{ [bit]}.$$

Tabela 1. Zbiór przykładów uczących (na podst. [Quinlan, 1986])

nr	pogoda	temperatura	wilgotność	wiatr	obiekty
1	słońce	wysoka	wysoka	nie	—
2	słońce	wysoka	wysoka	tak	—
3	zachmurzenie	wysoka	wysoka	nie	+
4	deszcz	średnia	wysoka	nie	+
5	deszcz	niska	normalna	nie	+
6	deszcz	niska	normalna	tak	—
7	zachmurzenie	niska	normalna	tak	+
8	słońce	średnia	wysoka	nie	—
9	słońce	niska	normalna	nie	+
10	deszcz	średnia	normalna	nie	+
11	słońce	średnia	normalna	tak	+
12	zachmurzenie	średnia	wysoka	tak	+
13	zachmurzenie	wysoka	normalna	nie	+
14	deszcz	średnia	wysoka	tak	—

Dla kolejnych atrybutów otrzymano ostatecznie:

$$\Delta I(E, temperatura) = \underline{0,029} \text{ [bit]},$$

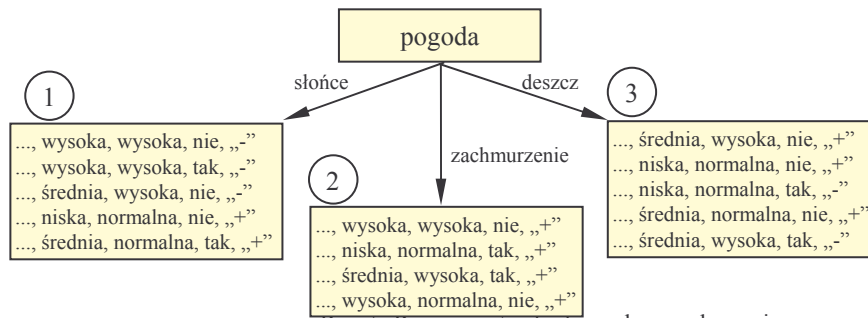
$$\Delta I(E, wilgotność) = \underline{0,151} \text{ [bit]},$$

$$\Delta I(E, wiatr) = \underline{0,048} \text{ [bit]}.$$

Tak więc budowę drzewa rozpoczęto od atrybutu „pogoda”, który będzie nazywany *korzeniem drzewa decyzyjnego*. Rys. 5 przedstawia pierwszy etap budowy drzewa decyzyjnego z wyszczególnionymi przykładami przynależnymi do kolejnych węzłów.

W celu wybrania kolejnego atrybutu do rozbudowy drzewa, oblicza się wartość $I(E)$ dla zbioru uczącego przynależnego do węzła oznaczonego symbolem np. ←. W zbiorze tym znajduje się 5 przykładów z czego 2 przypisano do „+” a 3 do „-”. Zgodnie z zależnością (1) otrzymano:

$$I(E) = 0,971 \text{ [bit]}.$$



Rys. 5. Pierwszy etap budowy drzewa decyzyjnego

Następnie obliczono wartości $\Delta I(E, a)$ dla każdego atrybutu: atrybut „pogoda” o wartościach {słońce, zachmurzenie, deszcz}:

$$„+” = 2, „-” = 3, \quad I(E_{\text{słońce}}) = 0,971 \text{ [bit]}$$

$$„+” = 0, „-” = 0, \quad I(E_{\text{zachmurzenie}}) = 0,0 \text{ [bit]},$$

$$„+” = 0, „-” = 0, \quad I(E_{\text{deszcz}}) = 0,0 \text{ [bit]}.$$

$$I(E, \text{pogoda}) = 0,971 \text{ [bit]}$$

$$\Delta I(E, \text{pogoda}) = 0,971 - 0,971 = 0,0 \text{ [bit]}.$$

Dla kolejnych atrybutów otrzymano:

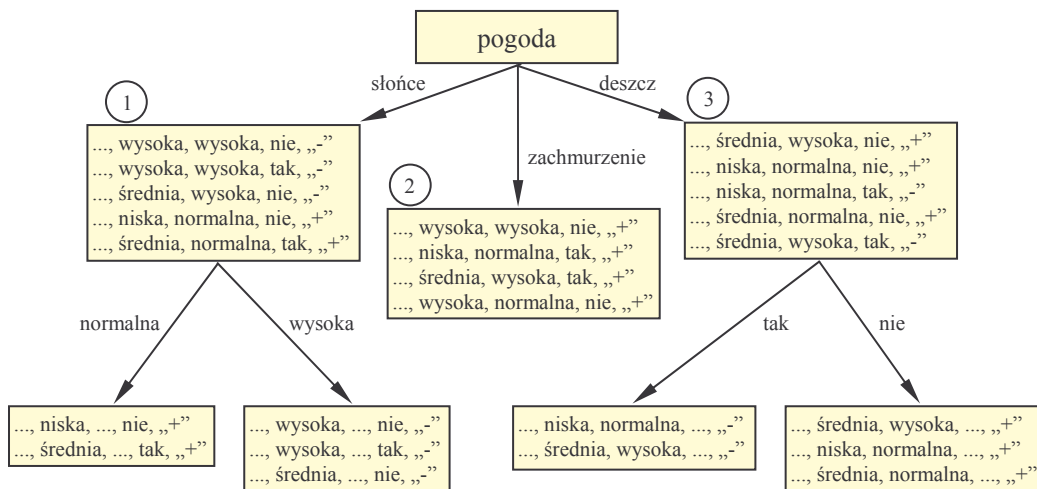
$$\Delta I(E, \text{temperatura}) = 0,571 \text{ [bit]},$$

$$\Delta I(E, \text{wilgotność}) = 0,971 \text{ [bit]},$$

$$\Delta I(E, \text{wiatr}) = 0,020 \text{ [bit]}.$$

Rozbudowę węzła oznaczonego symbolem \leftarrow rozpoczęto od atrybutu „wilgotność”. Rozbudowy węzła oznaczonego symbolem \rightarrow dokonuje się w ten sam sposób jak dla węzła oznaczonego symbolem \leftarrow . Rys. 6 przedstawia drugi i trzeci etap budowy drzewa decyzyjnego z wyszczególnionymi przykładami przynależnymi do kolejnych węzłów.

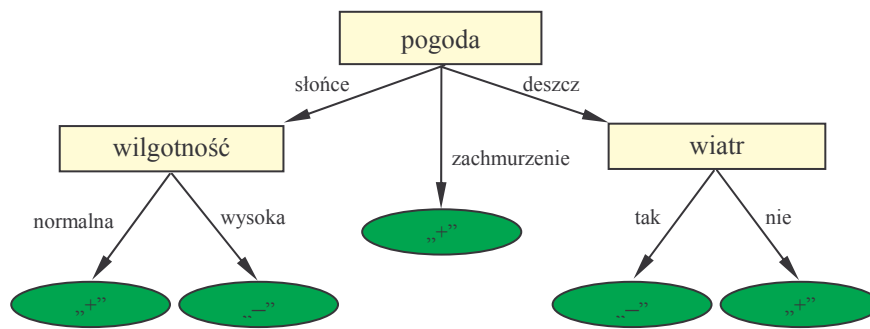
Należy zauważyć, że w węźle oznaczonym symbolem \uparrow wszystkie 4 przykłady są przypisane do „+”. Z tego też względu nie jest możliwe dalsze podzielenie tego podzbioru przykładów jak to miało miejsce w przypadku węzłów oznaczonych \leftarrow i \rightarrow . Jednocześnie



Rys. 6. Drugi i trzeci etap budowy drzewa decyzyjnego

po podziale węzłów \leftarrow i \rightarrow również otrzymano podzbiory, w których przykłady przypisane są tylko do „+” lub tylko do „-”. W sytuacji gdy nie jest możliwe dalsze dzielenie zbiorów przykładów, budowę drzewa

decyzyjnego kończy się. Węzły, w których nie można już podzielić zbiorów przykładów nazywa się *liśćmi* drzewa decyzyjnego. Rys. 7 przedstawia gotowe drzewo decyzyjne z odpowiadającym mu zbiorem reguł.



jeśli	(pogoda = słońce)	i (wilgotność = normalna)	to „+” – Gramy w golfa
jeśli	(pogoda = słońce)	i (wilgotność = wysoka)	to „-” – Zostajemy w domu
jeśli	(pogoda = zachmurzenie)		to „+” – Gramy w golfa
jeśli	(pogoda = deszcz)	i (wiatr = tak)	to „-” – Zostajemy w domu
jeśli	(pogoda = deszcz)	i (wiatr = nie)	to „+” – Gramy w golfa

Rys. 7. Gotowe drzewo decyzyjne i odpowiadający mu zbiór reguł

CZYM JEST DETREEX?

W rozdziale:

- Zastosowania DeTreexa
- Użytkownicy DeTreexa

DeTreex jest narzędziem służącym do wspomagania procesu pozyskiwania wiedzy. Dzięki zastosowanej indukcyjnej metodzie „uczenia maszynowego” możliwe jest budowanie drzew decyzyjnych i zapis tych drzew w postaci reguł (reguły są najczęściej stosowaną metodą reprezentacji wiedzy w bazach wiedzy systemów ekspertowych).

System DeTreex jest niezależny dziedzinowo i może zostać zastosowany zarówno do budowy drzew decyzyjnych w dziedzinie ekonomii, techniki (np. [Maniak, 1999]), medycyny jak również biologii. Ogólnie, DeTreex może być stosowany wszędzie tam, gdzie pojawia się problem:

- podejmowania decyzji (klasyfikacji obiektów),
- szybkiego pozyskania reguł decyzyjnych ze zbioru przykładów uczących,
- szybkiej weryfikacji pozyskanych reguł.

ZASTOSOWANIA DETREEXA

Rozwój urządzeń komputerowych w skali, w jakiej to można obserwować w ostatnich latach, powoduje, że coraz trudniej wskazać dziedzinę gospodarki, techniki, nauki, czy życia społecznego właściwie rozwijające się bez komputerów. Coraz częstsze zastosowania komputerów powodują gwałtowne zmiany komunikacji, medycynie, ekonomii, łączności, kształceniu i przemyśle zbrojeniowym.

Mimo, iż środki i techniki komputerowe są obecnie bardzo mocno rozwinięte, nie jest możliwe całkowite zastąpienie człowieka w jego czynnościach intelektualnych. Czynione są pomyślnie próby zastosowania wyspecjalizowanych *systemów ekspertowych*, które szczególnie wspomagają w podejmowaniu decyzji przez eksperta. W przypadku zastosowania systemów ekspertowych konieczne jest zbudowanie odpowiednich baz wiedzy, w których zapisana zostanie wiedza z wybranej dziedziny. Pozyskiwanie wiedzy dla takich baz jest bardzo trudnym i odpowiedzialnym zadaniem. Dzięki opracowanym indukcyjnym metodom pozyskiwania wiedzy i zastosowaniu komputerów, możliwe stało się bardzo efektywne budowanie takich baz wiedzy (Rys. 3).

System DeTreex został opracowany za zastosowaniem metody indukcji drzew decyzyjnych. Jak napisano w poprzednim rozdziale, indukcji tej dokonuje się na podstawie zgromadzonego zbioru przykładów uczących (baz danych ilościowych/numerycznych i jakościowych/dyskretnych). W każdym banku, biurze maklerskim, przedsiębiorstwie, jednostce naukowo-dydaktycznej prowadzącej szereg badań, klinice medycznej itp. gromadzone są obszerne bazy danych, które stosowane są w większości przypadków do analiz statystycznych i marketingu. Posiadane bazy danych to nie tylko szereg wartości liczbowych i znaków alfanumerycznych. Bazy danych zawierają wiedzę, która nie jest jeszcze znana *explicite* w danym momencie. Stosując DeTreexa można tą wiedzę wydobyć z baz danych i użyć znacznie efektywniej niż do tej pory.

UŻYTKOWNICY DETREEXA

System DeTreex może być stosowany przez:

- 1) departamenty kredytów konsumpcyjnych i gospodarczych banków,
- 2) firmy udzielające różnego rodzaju kredytów (pożyczek) osobom fizycznym,
- 3) firmy tworzące lub wykorzystujące hurtownie danych, w celu pogłębionej analizy danych i odkrywania wiedzy zawartej w bazach danych,
- 4) w trakcie szkoleń nowych specjalistów ds. kredytów konsumpcyjnych rozwijając umiejętności i doświadczenie w danej dziedzinie,
- 5) biura maklerskie, otwarte i zamknięte fundusze inwestycyjne i emerytalne oraz inne przedsiębiorstwa w zakresie podejmowania decyzji strategicznych i operacyjnych,
- 6) w dydaktyce i pracach badawczych wyższych uczelni.

PRZYKŁAD DRZEWA DECYZYJNEGO

KREDYTY KONSUMPCYJNE

Jak wspomniano w rozdziale 2, indukcję drzew decyzyjnych stosuje się głównie do klasyfikacji obiektów (problemów) a wygenerowana metodami indukcyjnymi wiedza, poprzez system ekspertowy, wspomaga w podejmowaniu decyzji.

Typowy problemem klasyfikacji można przedstawić na przykładzie podejmowania decyzji kredytowych przez departamenty kredytów konsumpcyjnych banków oraz firmy udzielające różnego rodzaju kredytów (pożyczek) osobom fizycznym. Specjalista ds. kredytów na podstawie dialogu z klientem uzyskuje szereg informacji, które w różnym stopniu warunkują o przyznaniu lub odmowie udzielenia kredytu. Bardzo często nie są znane dokładne reguły i zasady przyznawania kredytów i niejednokrotnie specjalista podejmuje decyzję „na wyczucie”.

Tabela 2 przedstawia fragment hipotetycznych historycznych danych jakie można zgromadzić w banku lub firmie udzielającej kredytów. W przypadku wątpliwości podczas podejmowania decyzji, przeglądanie tylko kilkunastu danych historycznych i porównywanie ich z danymi nowego klienta może okazać się zbyt czasochłonne a nawet niemożliwe. Stosując indukcyjną metodę generowania drzew decyzyjnych możliwe jest uogólnienie zawartej w danych wiedzy. W niniejszym przykładzie zdefiniowano następującą dziedzinę (atrybuty oraz ich wartości decydujące o przyznaniu lub odmowie przyznania kredytu):

Praca:	tak, nie
Kredyt na:	komputer, meble, mieszkanie, rower, samochód, sprzęt rtv-agd
Płeć:	kobieta, mężczyzna
Żonaty/Zamężna:	tak, nie
Wiek:	wartości liczb całkowitych dodatnich (>18)
Oszczędności:	wartości liczb rzeczywistych dodatnich
Kwota kredytu:	wartości liczb rzeczywistych dodatnich
Liczba rat:	wartości liczb całkowitych dodatnich
Staż pracy:	wartości liczb całkowitych dodatnich

Klasyfikowany obiekt (problem):

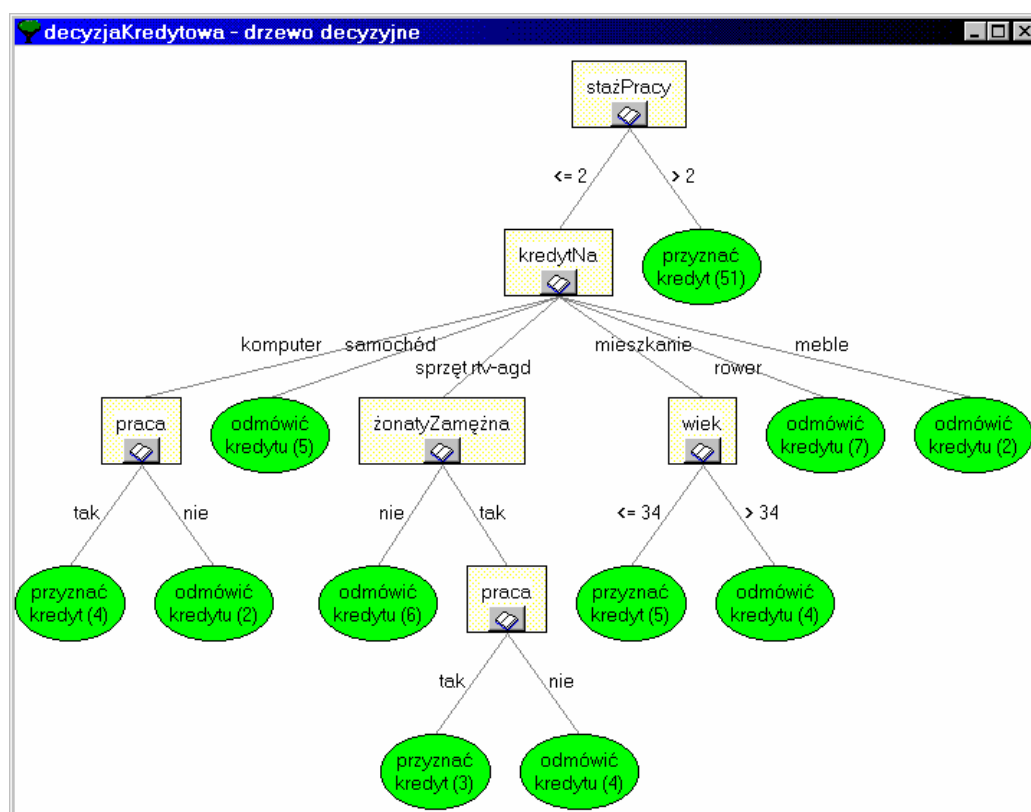
Decyzja kredytowa: przyznać kredyt, odmówić kredytu

Tabela 2. Fragment danych historycznych dotyczących kredytów

Praca	Kredyt na	Płeć	Żonaty/ Zamężna	Wiek	Oszczędności	Kwota kredytu	Liczba rat	Staż pracy	Decyzja kredytowa
tak	komputer	kobieta	nie	18	2,5	3,8	15	1	przyznać kredyt
tak	komputer	kobieta	nie	20	1,3	5,0	20	2	przyznać kredyt
nie	komputer	kobieta	tak	25	0,6	6,0	12	0	odmówić kredytu
tak	komputer	kobieta	tak	40	0,6	10,5	12	2	przyznać kredyt
...									
tak	samochód	kobieta	nie	18	6,3	20,0	20	1	odmówić kredytu
nie	samochód	kobieta	tak	20	6,3	25,0	20	0	odmówić kredytu
tak	samochód	kobieta	nie	25	6,3	12,5	20	5	odmówić kredytu
tak	samochód	kobieta	nie	38	18,8	25,0	20	15	przyznać kredyt
...									
tak	sprzęt rtv-agd	mężczyzna	tak	42	11,3	1,5	6	7	przyznać kredyt
tak	sprzęt rtv-agd	mężczyzna	nie	28	1,3	6,3	5	3	przyznać kredyt
nie	sprzęt rtv-agd	kobieta	tak	55	3,8	10,0	8	0	odmówić kredytu
tak	sprzęt rtv-agd	kobieta	tak	21	1,3	1,5	12	1	przyznać kredyt
...									

tak	mieszkanie	kobieta	nie	32	50	50,0	20	2	przyznać kredyt
tak	mieszkanie	kobieta	tak	37	12,5	100,0	10	1	przyznać kredyt
tak	mieszkanie	kobieta	tak	41	6,3	125,0	20	18	przyznać kredyt
tak	mieszkanie	kobieta	tak	50	37,5	75,0	20	27	przyznać kredyt
...									
tak	rower	mężczyzna	nie	18	1,9	1,3	10	1	przyznać kredyt
tak	rower	mężczyzna	nie	18	1,3	6,3	10	1	odmówić kredytu
tak	rower	mężczyzna	nie	22	2,5	6,3	10	5	przyznać kredyt
tak	rower	mężczyzna	nie	22	1,3	6,3	10	2	odmówić kredytu
...									
tak	meble	kobieta	tak	40	6,3	12,5	20	20	przyznać kredyt
tak	meble	kobieta	tak	50	12,5	15,0	12	20	przyznać kredyt
tak	meble	kobieta	nie	30	6,3	12,5	20	3	przyznać kredyt
tak	meble	kobieta	tak	35	2,5	18,8	15	13	przyznać kredyt
...									

Rys. 8 przedstawia drzewo decyzyjne zbudowane na podstawie 93 hipotetycznych decyzji kredytowych. Uzyskane drzewo o 11 liściach zostało zapisane w regułowym formacie reprezentacji wiedzy (Tabela 3).



Rys. 8. Przykładowe drzewo decyzyjne – decyzja kredytowa

Zbiór reguł jest wielokrotnie mniejszy od zbioru danych, na podstawie którego pozyskano te reguły. Reguły stanowią uogólnienie danych, dzięki temu możliwe jest klasyfikowanie nowych przypadków. Uzyskane zbiory reguł zastosować można do zapisu baz wiedzy systemu ekspertowego, który będzie w sposób bardzo efektywny wspomagał w podejmowaniu decyzji kredytowych.

Nr reguły	Część warunkowa reguły	Konkluzja „decyzjaKredytowa”
1	jeśli stażPracy > 2	przyznać kredyt
2	jeśli stażPracy <= 2 i kredytNa = rower	odmówić kredytu
3	jeśli stażPracy <= 2 i kredytNa = sprzęt rtv-agd i żonatyZamężna = nie	odmówić kredytu
4	jeśli stażPracy <= 2 i kredytNa = samochód	odmówić kredytu
5	jeśli stażPracy <= 2 i kredytNa = mieszkanie i wiek <= 34	przyznać kredyt
6	jeśli stażPracy <= 2 i kredytNa = komputer i praca = tak	przyznać kredyt
7	jeśli stażPracy <= 2 i kredytNa = sprzęt rtv-agd i żonatyZamężna = tak i praca = nie	odmówić kredytu
8	jeśli stażPracy <= 2 i kredytNa = mieszkanie i wiek > 34	odmówić kredytu
9	jeśli stażPracy <= 2 i kredytNa = sprzęt rtv-agd i żonatyZamężna = tak i praca = tak	przyznać kredyt
10	jeśli stażPracy <= 2 i kredytNa = meble	odmówić kredytu
11	jeśli stażPracy <= 2 i kredytNa = komputer i praca = nie	odmówić kredytu

OBSŁUGA SYSTEMU DeTREEX

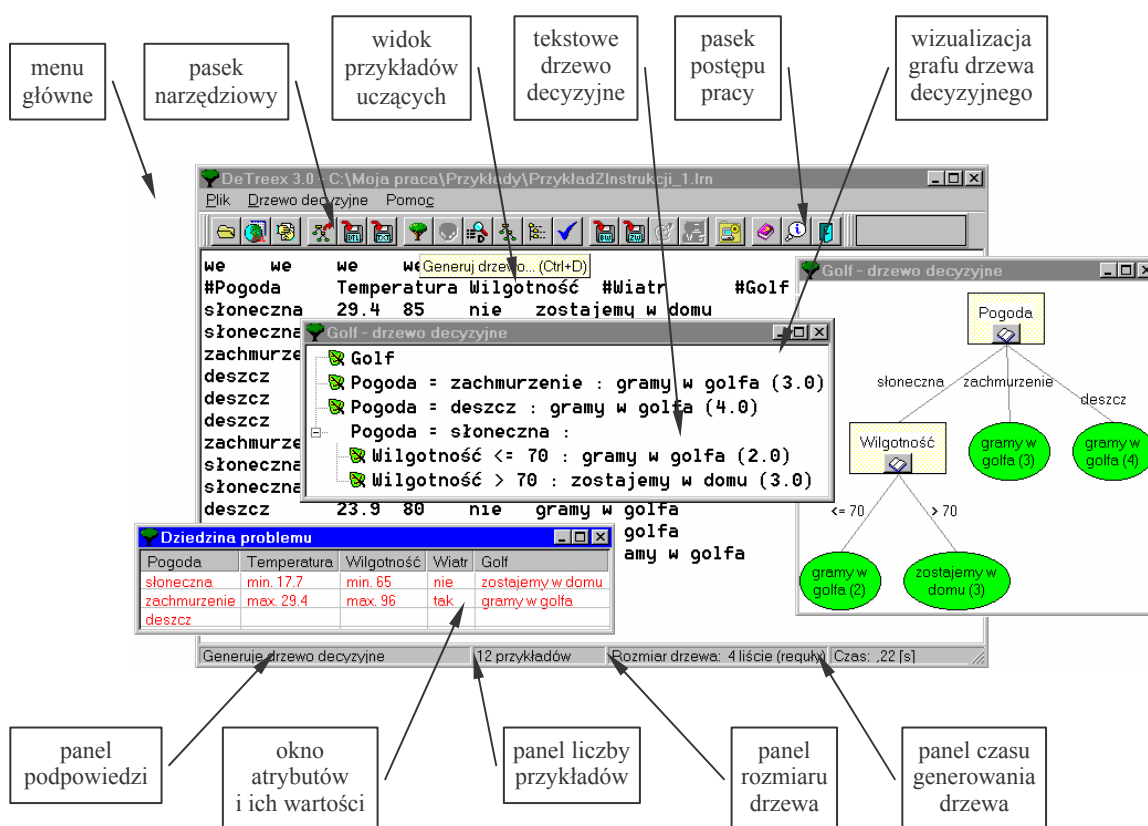
W rozdziale:

- Pierwsze uruchomienie
- Menu główne DeTreexa
- Format pliku uczącego i testowego
- Generowanie drzewa decyzyjnego i jego wizualizacja
- Zapis baz wiedzy i źródeł wiedzy
- Opcje generowania drzewa

PIERWSZE URUCHOMIENIE

System DeTreeX jest jednym z kilku systemów wchodzących w skład pakietu sztucznej inteligencji Sphinx firmy AITECH.

Po zainstalowaniu systemu jest obecny jako jeden z elementów utworzonej grupy i identyfikowany przez poniższą ikonę:



Rys. 9. Widok okna głównego systemu DeTreeX

Po uruchomieniu systemu ukazuje się jego główne okno (Rys. 9).

MENU GŁÓWNE DETREEXA

Główne okno systemu (Rys. 9) pozwala na dostęp do poleceń DeTreeXa. Umożliwiają one:

- wizualizację zbiorów przykładów uczących,
- wyników testowania drzewa,
- wizualizację zapisanych baz/źródeł wiedzy,
- przedstawiane w dodatkowych oknach: drzewo decyzyjne i opcje budowy drzewa,
- uruchomienie dodatkowych systemów: systemu ekspertowego PC-Shell oraz systemu wspomagania inżynierii wiedzy CAKE.

Funkcje systemu dostępne są poprzez wybór odpowiedniego polecenia z menu głównego. Dostęp do poleceń możliwy jest także za pośrednictwem przycisków (ikon) umieszczonych na pasku narzędziowym.



Włączenie lub wyłączenie prezentacji przykładów uczących w głównym oknie systemu DeTreeex. W przypadku otrzymania komunikatu o zbyt małej ilości pamięci operacyjnej w czasie wczytywania przykładów uczących, należy wyłączyć prezentację przykładów w oknie głównym systemu i ponownie otworzyć dany plik. Ikona uruchamiająca to polecenie nie ma swojego odpowiednika w menu głównym.

MENU „PLIK”



Otwórz plik uczący... (Ctrl+O)

Otwarcie pliku (z rozszerzeniem *.lrn) z zapisanymi w odpowiednim formacie przykładami uczącymi. Wybranie tego polecenia powoduje wyświetlenie standardowego okna służącego do wskazania i otwarcia pliku.



Import danych z bazy... (Ctrl+I)

Importowanie przykładów uczących z baz danych. Wybranie tego polecenia powoduje wyświetlenie kolejnych okien dialogowych umożliwiających wybranie: baz danych, tabel oraz atrybutów i zapis nowego pliku z przykładami uczącymi.



Otwórz graf drzewa... (Shift+Ctrl+O)

Otwarcie pliku (z rozszerzeniem *.dtl) z zapisanym w odpowiednim formacie grafem drzewa decyzyjnego. Wybranie tego polecenia powoduje wyświetlenie: standardowego okna służącego do wskazania i otwarcia pliku oraz (po wybraniu pliku) dodatkowego okna z graficzną prezentacją drzewa.



Zapisz graf drzewa... (Shift+Ctrl+S)

Zapisanie grafu drzewa decyzyjnego do pliku (z rozszerzeniem *.dtl). Wybranie tego polecenia powoduje wyświetlenie standardowego okna służącego do zapisu pliku.



Zapisz drzewo tekstowo (Ctrl+S)

Zapisanie drzewa decyzyjnego do pliku w formacie tekstowym (z rozszerzeniem *.txt).



Zakończ (Ctrl+Q)

Bezwarunkowe zamknięcie systemu.

MENU „DRZEWO DECYZYJNE”



Generuj drzewo (Ctrl+D)

Generowanie drzewa decyzyjnego dla wskazanych wcześniej przykładów uczących.



Przerwij pracę (Ctrl+Del)

Przerwanie budowy drzewa decyzyjnego. Polecenie dotyczy także przerwania testowania drzewa decyzyjnego oraz przerwania zapisu bazy/źródła wiedzy.



Pokaż dziedzinę... (Ctrl+A)

Wizualizacja dziedziny problemu. Uruchomienie tego polecenia powoduje wyświetlenie dodatkowego okna z tabelaryczną prezentacją atrybutów i ich wartości.



Pokaż drzewo graficzne... (Ctrl+G)

Wizualizacja drzewa decyzyjnego w postaci graficznej. Uruchomienie tego polecenia powoduje wyświetlenie dodatkowego okna z prezentacją drzewa.

**Pokaż drzewo tekstowe... (Shift+Ctrl+G)**

Wizualizacja drzewa decyzyjnego w postaci tekstowej. Uruchomienie tego polecenia powoduje wyświetlenie dodatkowego okna z prezentacją drzewa.

**Testuj drzewo (Ctrl+T)**

Testowanie drzewa decyzyjnego zbiorem przykładów testowych zapisanych w dodatkowym pliku (z rozszerzeniem *.tst).

**Zapisz reguły do bazy wiedzy (Ctrl+B)**

Zapisanie reguł z drzewa decyzyjnego do pliku bazy wiedzy (z rozszerzeniem *.bw).

**Zapisz reguły do źródła wiedzy (Ctrl+Z)**

Zapisanie reguł z drzewa decyzyjnego do pliku źródła wiedzy (z rozszerzeniem *.zw).

**Uruchom bazę wiedzy w PC-Shellu (Ctrl+P)**

Uruchamia zapisaną poleceniem „Zapisz reguły do bazy wiedzy” bazę wiedzy w systemie ekspertowym PC-Shell.

**Otwórz bazę/źródło wiedzy w CAKE'u (Ctrl+C)**

Uruchamia zapisaną poleceniem „Zapisz reguły do bazy wiedzy” lub „Zapisz reguły do źródła wiedzy” bazę lub źródło wiedzy w systemie CAKE.

**Opcje... (F10)**

Wybranie tego polecenia powoduje wyświetlenie dodatkowego okna dialogowego, w którym określa się parametry budowy drzewa decyzyjnego.

MENU „POMOC”**Tematy pomocy (F1)**

Uruchomienie pomocy ekranowej dotyczącej systemu DeTreex.

**O programie...**

Wyświetlenie okna z informacjami o aktualnej wersji systemu DeTreex i firmie AITECH.

FORMAT PLIKU UCZĄCEGO I TESTOWEGO

Jak napisano w rozdziale 1, indukcja drzew decyzyjnych jest metodą pozyskiwania wiedzy na podstawie przykładów. W celu wygenerowania drzewa decyzyjnego należy taki zbiór przykładów zapisać w pliku uczącym. Użytkownik może taki plik przygotować w dowolnym arkuszu kalkulacyjnym lub edytorze tekstu.

W celu testowania drzewa decyzyjnego (poprawności klasyfikacji przez to drzewo) konieczne jest przygotowanie tzw. pliku testowego. Format pliku uczącego i testowego jest taki sam, różnią się one rozszerzeniem nazw plików, np.:

plik uczący:	PrzykładZInstrukcji_1.lrn
plik testowy:	PrzykładZInstrukcji_1.tst

Rys. 10 i Rys. 11 przedstawiają przykładowe pliki: uczący i testowy. Utworzono 14 przykładów opisujących problem gry w golfa w zależności od warunków atmosferycznych. Warunki atmosferyczne zostały opisane 4 atrybutami. Ostatnia kolumna określa rozwiązanie problemu. 12 przykładów zapisano jako uczące, pozostałe 2 jako testowe.

Pierwszy wiersz w obu plikach oznacza liczbę atrybutów opisujących problem (symbole **we**) oraz rozwiązanie danego problemu (symbol **wy**).

Uwagi:

- 1) wszystkie kolumny w plikach *.ltn i *.tst muszą być oddzielone od siebie tylko jednym znakiem tabulatora,
- 2) pliki uczące i testowe muszą zawierać co najmniej jedną kolumnę oznaczoną symbolem we,
- 3) pliki uczące i testowe muszą zawierać tylko jedną kolumnę oznaczoną symbolem wy,
- 4) kolumna wyjściowa (symbol wy) oznaczająca rozwiązanie danego problemu musi być zapisana jako ostatnia,
- 5) liczba kolumn oznaczonych symbolami we i wy musi być zgodna w obu plikach uczącym i testowym.

we	We	we	we	wy
#Pogoda	Temperatura	Wilgotność	#Wiatr	#Golf
słoneczna	29.4	85	nie	zostajemy w domu
słoneczna	26.6	90	tak	zostajemy w domu
zachmurzenie	28.3	78	nie	gramy w golfa
deszcz	21.1	96	nie	gramy w golfa
deszcz	29.0	80	nie	gramy w golfa
deszcz	18.3	70	tak	zostajemy w domu
zachmurzenie	17.7	65	tak	gramy w golfa
słoneczna	22.2	95	nie	zostajemy w domu
słoneczna	20.5	70	nie	gramy w golfa
deszcz	23.9	80	nie	gramy w golfa
słoneczna	23.9	70	tak	gramy w golfa
zachmurzenie	22.2	90	tak	gramy w golfa

Rys. 10. Plik uczący „PrzykładZInstrukcji_1.ltn”

we	we	we	we	wy
#Pogoda	Temperatura	Wilgotność	#Wiatr	#Golf
zachmurzenie	27.2	75	nie	gramy w golfa
deszcz	21.6	80	tak	zostajemy w domu

Rys. 11. Plik testowy „PrzykładZInstrukcji_1.tst”

Drugi wiersz w obu plikach zawiera nazwy atrybutów opisujących problem oraz (w ostatniej kolumnie) nazwę problemu. Nazwy atrybutów mogą zawierać ciąg znaków **nie dłuższy niż 255** znaków. Kolumna pliku uczącego i/lub testowego zawierająca wartości symboliczne wyliczeniowe powinna być oznaczona znakiem specjalnym „#” bezpośrednio przed nazwą odpowiedniego atrybutu.

Uwagi:

- 1) nazwy atrybutów opisujących problem oraz nazwa problemu muszą być unikalne tzn. nie mogą powtórzyć się dwie takie same nazwy atrybutów (włącznie z nazwą problemu),
- 2) nazwy atrybutów opisujących problem oraz nazwa problemu w plikach uczącym i testowym muszą być identyczne i uszeregowane w tej samej kolejności,
- 3) w przypadku zapisu nazwy problemu (w ostatniej kolumnie oznaczonej przez wy) możliwe jest zastosowanie tylko wartości symbolicznych wyliczeniowych, co oznacza konieczność umieszczenia przed tą nazwą symbolu „#”,
- 4) w nazwach atrybutów opisujących problem oraz w nazwie problemu dozwolone jest stosowanie znaków z zakresu a-z i A-Z (w tym znaków polskich), znak „_” (podkreślenie) oraz cyfry (z wyjątkiem umieszczania cyfry na początku nazwy atrybutu),
- 5) stosowanie w nazwach atrybutów opisujących problem oraz w nazwie problemu znaków operacji matematycznych +, -, /, *, %, = oraz innych specjalnych np.: !, @, \$, ^, &, ~, (,), |, itp. jest niedozwolone,

- 6) nazwy atrybutów opisujących problem oraz nazwa problemu mogą zawierać spacje rozdzielające np.: „poranna temperatura powietrza” – w trakcie zapisu reguł do bazy/źródła wiedzy, znaki te zostaną zamienione znakami „_” (podkreślenia) np.: „poranna_temperatura_powietrza” (wymóg reprezentacji wiedzy w bazie wiedzy systemu ekspertowego PC-Shell),
- 7) nazwy atrybutów opisujących problem oraz nazwa problemu mogą rozpoczynać się dużą literą (lub być pisane w całości dużymi literami) np.: „Temperatura” – w trakcie zapisu reguł do bazy/źródła wiedzy, pierwszy znak zostanie zamieniony na znak małej litery np.: „temperatura” (wymóg reprezentacji wiedzy w bazie wiedzy systemu ekspertowego PC-Shell).

Kolejne wiersze zawierają przykłady opisujące dany problem lub obiekt. W pliku uczącym *.lrn zapisane są przykłady uczące, w pliku testowym *.tst zapisane są przykłady testowe służące do testowania poprawności klasyfikacji drzewa decyzyjnego. Do zapisu przykładów stosujemy wcześniej zdefiniowane wartości atrybutów. Opis dziedziny dla powyższego przykładu (Rys. 10 i Rys. 11) przedstawia się następująco:

Pogoda:	słońce, zachmurzenie, deszcz
Temperatura:	(wartości liczb rzeczywistych) °C
Wilgotność:	(wartości liczb całkowitych dodatnich) %
Wiatr:	tak, nie
Golf:	gramy w golfa, zostajemy w domu

W systemie DeTreex możliwe jest operowanie na następujących wartościach:

- liczbowych rzeczywistych (ilościowych),
- symbolicznych wyliczeniowych (jakościowych).

Wartości liczbowe mogą należeć do przedziału od $3.4 \cdot 10^{-38}$ do $3.4 \cdot 10^{+38}$. **Wartości symboliczne wyliczeniowe** mogą zawierać ciąg znaków **nie dłuższy niż 255 znaków**. Wartości te powinny reprezentować pewne kategorie, pomiędzy którymi nie zachodzi żadna relacja większości, np.: nazwy kolorów lub płeć.

Dodatkowo istnieje możliwość zapisu przykładów z brakującymi wartościami atrybutów oznaczonych jako wejściowe (symbol **we**), np.:

? zachmurzenie	26.6	? 78	tak	zostajemy w domu
	?		?	gramy w golfa

Wszystkie wartości atrybutu wyjściowego (symbol **wy**) muszą być określone, tzn. w przypadku zapisu przykładu, w którym atrybut wyjściowy przyjmie wartość „?”, wówczas wartość ta rozumiana będzie przez system DeTreex jako kolejna wartość tego atrybutu a nie jako wartość nieznana.

Uwagi:

- 1) kolejność zapisu wartości atrybutów w każdym wierszu musi być identyczna jak kolejność zapisanych nazw atrybutów w drugiej linii pliku uczącego (i/lub testowego), czyli muszą znajdować się w odpowiednich kolumnach danych,
- 2) podczas zapisu nieznanych wartości atrybutów wejściowych dozwolone jest stosowanie tylko jednego znaku „?”; zastosowanie więcej niż jednego znaku „?” lub innych, traktowane będzie przez system DeTreex jako kolejne znane wartości atrybutów wejściowych,
- 3) w zapisie wartości symbolicznych wyliczeniowych możliwe jest stosowanie dowolnych znaków,
- 4) podczas zapisu wartości liczbowych rzeczywistych jako separator liczb dziesiętnych można stosować zamiennie znak „.” (kropki) lub „,” (przecinka),
- 5) pliki uczące i testowe przygotowane dla systemu DeTreex można stosować do uczenia sztucznej sieci neuronowej przez system Neuronix bez jakichkolwiek zmian (z wyjątkiem plików, w których znajdują się przykłady z nieznanymi wartościami atrybutów),

6) *nie wszystkie pliki uczące i testowe przygotowane dla systemu Neuronix można zastosować do budowy drzewa decyzyjnego przez system DeTreex; w plikach Neuronixa mogą być stosowane:*

- a) *atrybuty o wartościach symbolicznych hierarchicznych, pomiędzy którymi zachodzi relacja większości (nazwy atrybutów oznaczone są symbolem „>”), oraz*
- b) *kolumny z dodatkowymi opisami wierszy (przykładów) oznaczone symbolem „–” (myślnik)*

co nie jest dozwolone w systemie DeTreex.

UŻYTKOWANIE DETREEXA




Po uruchomieniu DeTreexa użytkownikowi ukazuje się główne okno systemu (Rys. 9). Większość z ikon na pasku narzędziowym oraz poleceń w menu głównym jest nieaktywna co ma zabezpieczyć system przed wywołaniem nieodpowiedniego w danej chwili polecenia. Nie jest możliwe np.:

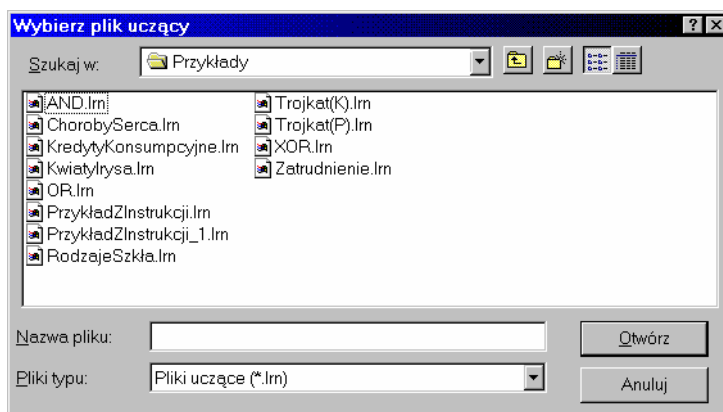
- uruchomienie systemu ekspertowego PC-Shell lub systemu wspomagania inżynierii wiedzy CAKE bez wcześniejszego zapisania bazy lub źródła wiedzy,
- nie jest możliwe zapisanie bazy/źródła wiedzy lub drzewa w pliku tekstowym bez wcześniejszego wygenerowania drzewa decyzyjnego,
- nie jest możliwe testowanie poprawności klasyfikacji drzewa decyzyjnego bez wcześniejszego wygenerowania tego drzewa oraz bez pliku testowego z rozszerzeniem *.tst,
- nie jest możliwa wizualizacja wygenerowanego drzewa w dodatkowych oknach oraz przedstawienie dziedziny problemu bez wcześniejszego wygenerowania tego drzewa,
- nie jest możliwe wywołanie okna opcji generowania drzewa decyzyjnego bez wcześniejszego wskazania pliku uczącego.

W trakcie wykonywania poszczególnych kroków stają się aktywne kolejne polecenia w menu głównym oraz ikony na pasku narzędziowym.

WSKAZANIE PLIKU UCZĄCEGO



W celu wygenerowania drzewa decyzyjnego konieczne jest wskazanie pliku uczącego ze zbiorem przykładów uczących.

Z menu głównego należy wybrać polecenie „Plik | Otwórz plik uczący...” lub wskazać kursorem ikonę . Wyświetlone zostanie okno wyboru pliku uczącego (Rys. 12). Jeśli istnieje potrzeba niewyświetlania przykładów uczących w oknie głównym systemu DeTreex (aby niepotrzebnie nie obciążać pamięci komputera), wówczas należy wskazać kursorem ikonę  przed otwarciem pliku z przykładami. W sytuacji gdy istnieje potrzeba wyświetlania przykładów uczących (ustawienie domyślne systemu DeTreex), wówczas przed otwarciem kolejnego pliku uczącego należy wskazać kursorem ikonę .







Rys. 12. Okno dialogowe wyboru pliku uczącego

Uwagi:

- 1) ikona  oraz  umieszczone są na tym samym klawiszu na pasku narzędzi,
- 2) w oknie głównym systemu DeTreeex, w którym wyświetlane są przykłady uczące, wyniki testowania drzewa oraz bazy/źródła wiedzy nie można dokonywać edycji.





GENEROWANIE DRZEWY DECYZYJNEGO

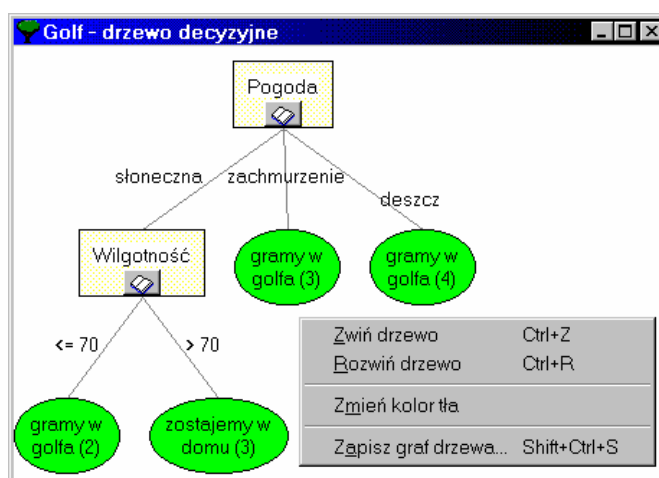
Po uprzednim wskazaniu pliku uczącego uaktywniają się nowe polecenia menu głównego „Drzewo decyzyjne | Generuj drzewo” „Drzewo decyzyjne | Opcje...” oraz odpowiadające tym poleceniom ikony:  i . Wygenerowanie drzewa możliwe jest poprzez wybór polecenia „Drzewo decyzyjne | Generuj drzewo” lub poprzez wskazanie kursorem myszki ikony . W czasie generowania drzewa nieaktywne są wszystkie polecenia oraz ikony z wyjątkiem polecenia „Drzewo decyzyjne | Przerwij pracę” (i ikony ). Polecenie to służy do przerywania obecnie wykonywanego etapu pracy w dowolnym jego momencie.

Uwagi:

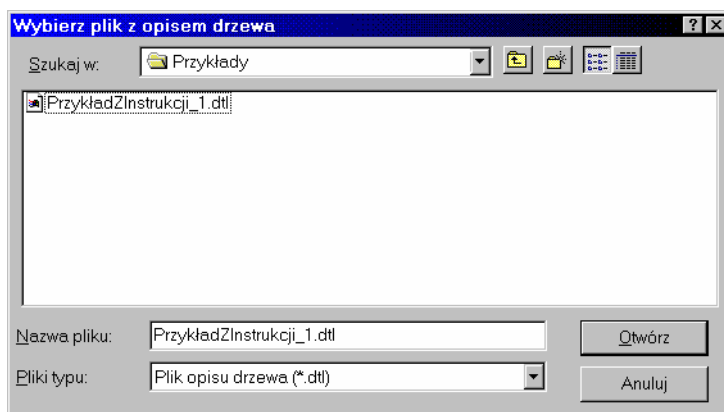
- 1) w trakcie generowania drzewa decyzyjnego na panelu podpowiedzi ukazują się komentarze dotyczące obecnie wykonywanego etapu pracy a na pasku postępu ukazuje się ich etap zaawansowania (Rys. 9),
- 2) po wczytaniu przykładów na panelu liczby przykładów (Rys. 9) ukazuje się liczba wczytanych przykładów z pliku uczącego,
- 3) po wygenerowaniu drzewa decyzyjnego na panelu rozmiaru drzewa (Rys. 9) ukazuje się liczba liści drzewa decyzyjnego,
- 4) po wygenerowaniu drzewa decyzyjnego na panelu czasu generowania drzewa (Rys. 9) ukazuje się czas budowy tego drzewa.

WIZUALIZACJA I ZAPIS DRZEWY DECYZYJNEGO

Po uprzednim zbudowaniu drzewa decyzyjnego możliwe jest jego przedstawienie w postaci graficznej i/lub tekstowej. Wydając polecenie „Drzewo decyzyjne | Pokaż drzewo graficzne...” lub wskazując kursorem ikonę  wywoływane jest dodatkowe okno z prezentacją drzewa decyzyjnego (Rys. 13). W celu wizualizacji drzew wcześniej wygenerowanych i zapisanych do pliku z rozszerzeniem *.dtl, należy wydać polecenie „Plik | Otwórz graf drzewa...” lub wskazać kursorem ikonę . Wywoływane jest wówczas standardowe okno wyboru pliku (Rys. 14) i (po wskazaniu tego pliku) dodatkowe okno z prezentacją drzewa decyzyjnego. Ikony umieszczone w węzłach drzewa służą do rozwijania  i zwiijania  poszczególnych gałęzi tego drzewa.



Rys. 13. Drzewo decyzyjne wygenerowane na podstawie pliku „PrzykładZInstrukcji_1.lrn”




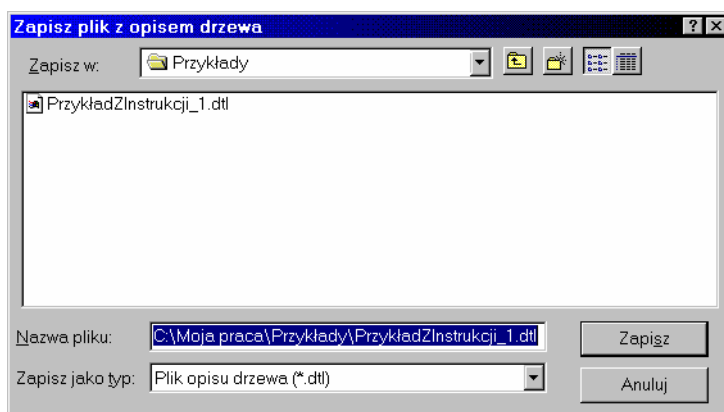
Rys. 14. Okno dialogowe wyboru pliku z grafem drzewa decyzyjnego

Naciskając prawy klawisz myszki w oknie wizualizacji uaktywniane jest dodatkowe menu. W menu tym możliwe jest szybkie zwinięcie oraz rozwinięcie całego drzewa decyzyjnego, zmiana koloru tła w widoku drzewa (Rys. 15) oraz zapisanie tego drzewa do pliku z rozszerzeniem *.dtl (Rys. 16).




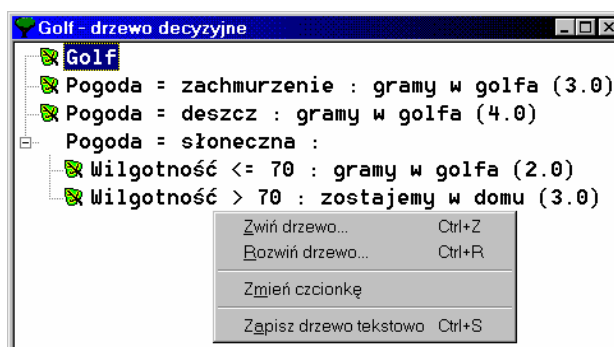
Rys. 15. Okno dialogowe zmiany koloru tła widoku drzewa decyzyjnego

Zapisu grafu drzewa decyzyjnego do pliku dokonać można także wybierając z menu głównego systemu polecenie „Plik | Zapisz graf drzewa...” lub wskazując kursorem ikonę .




Rys. 16. Okno dialogowe zapisu grafu drzewa decyzyjnego

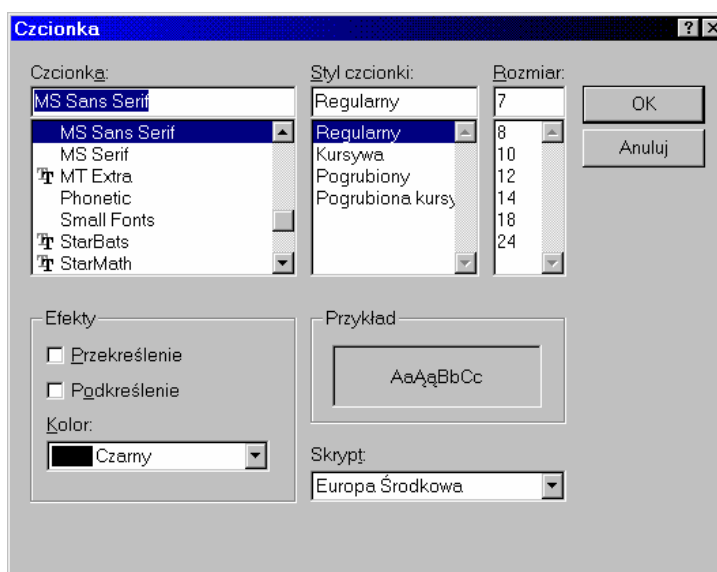
Wydając polecenie „Drzewo decyzyjne | Pokaż drzewo tekstowe...” lub wskazując kursorem ikonę  wywoływane jest dodatkowe okno z prezentacją drzewa decyzyjnego (Rys. 17).



Rys. 17. Drzewo decyzyjne wygenerowane na podstawie pliku „PrzykładZInstrukcji_1.lrn”


Naciskając prawy klawisz myszki uaktywniane jest dodatkowe menu. W menu tym możliwe jest szybkie zwiniecie oraz rozwinięcie całego drzewa decyzyjnego, zapis drzewa w pliku tekstowym oraz zmiana czcionki przedstawiającej nazwy atrybutów i ich wartości (Rys. 18).

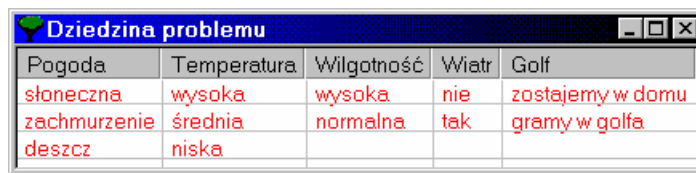
Możliwe jest także zapisanie drzewa decyzyjnego w pliku tekstowym wydając polecenie „Plik | Zapisz drzewo tekstowo” lub wskazując ikonę  (Rys. 9). Po wybraniu powyższych poleceń plik tekstowy z drzewem decyzyjnym zapisany jest w tym samym katalogu co plik uczący lecz z rozszerzeniem *.txt (np. gdy wczytany zostanie plik uczący „PrzykładZInstrukcji_1.lrn” wówczas zostanie zapisany plik „PrzykładZInstrukcji_1.txt”) i nie jest on przedstawiany w oknie głównym systemu DeTreeX.



Rys. 18. Okno dialogowe zmiany czcionki wizualizacji drzewa decyzyjnego

WIZUALIZACJA DZIEDZINY PROBLEMU


Po uprzednim zbudowaniu drzewa decyzyjnego możliwe jest przedstawienie zbioru atrybutów i ich wartości, na podstawie których zapisano przykłady uczące. Wydając polecenie „Drzewo decyzyjne | Pokaż dziedzinę...” lub wskazując kursorem ikonę  wywoływane jest dodatkowe okno z prezentacją dziedziny problemu (Rys. 19).



Pogoda	Temperatura	Wilgotność	Wiatr	Golf
słoneczna	wysoka	wysoka	nie	zostajemy w domu
zachmurzenie	średnia	normalna	tak	gramy w golfa
deszcz	niska			

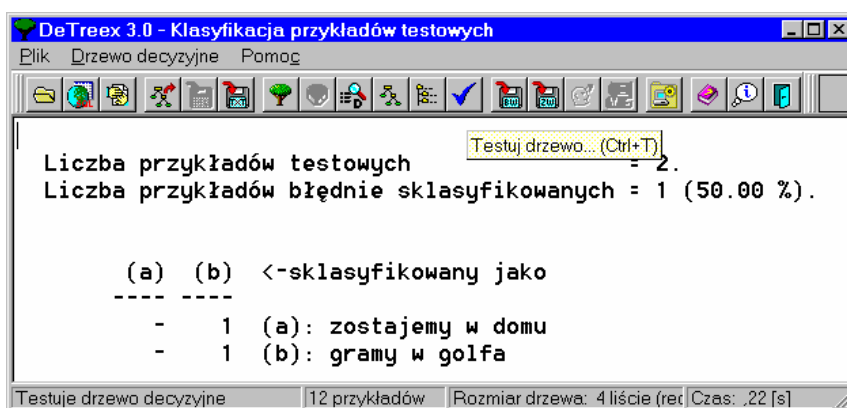
Rys. 19. Okno przedstawiające dziedzinę problemu

TESTOWANIE DRZEWY DECYZYJNEGO

Jeśli wygenerowano wcześniej drzewo decyzyjne oraz jeśli istnieje w tym samym katalogu co plik uczący, plik testowy o tej samej nazwie lecz z rozszerzeniem *.tst, wówczas możliwe jest testowanie poprawności klasyfikacji drzewa. Testowanie wykonuje się uruchamiając polecenie „Drzewo decyzyjne | Testuj drzewo” lub wskazując kursorem ikonę .




Po zakończeniu testowania, jego wyniki ukazują się w oknie głównym systemu DeTreex (Rys. 20). Wyniki te zawierają:

- liczbę przykładów testowych zapisanych w pliku testowym *.tst,
- liczbę przykładów testowych błędnie sklasyfikowanych przez drzewo decyzyjne,
- macierz, która przedstawia klasyfikację tych przykładów do poszczególnych rozwiązań wraz z liczbą przykładów (wartości umieszczone „po przekątnej” macierzy od górnego lewego rogu do dolnego prawego rogu oznaczają liczbę przykładów poprawnie sklasyfikowanych).



Rys. 20. Wyniki testowania drzewa decyzyjnego przykładami zawartymi w pliku testowym „PrzykładZInstrukcji_1.tst”



ZAPIS REGUŁ Z DRZEWY DECYZYJNEGO

Po uprzednim zbudowaniu drzewa decyzyjnego możliwe jest jego zapisanie w bazie wiedzy oraz źródle wiedzy systemu ekspertowego PC-Shell. Zapisu drzewa w postaci reguł w bazie wiedzy dokonać można wybierając polecenie z menu głównego „Drzewo decyzyjne | Zapisz reguły do bazy wiedzy” lub wskazując ikonę  (Rys. 9). Zapisu drzewa w postaci reguł w źródle wiedzy dokonać można wybierając polecenie z menu głównego „Drzewo decyzyjne | Zapisz reguły do źródła wiedzy” lub wskazując ikonę  (Rys. 9). Po wybraniu tych poleceń, baza (lub źródło) wiedzy przedstawiana jest w oknie głównym systemu DeTreex. W odróżnieniu od źródła wiedzy, baza wiedzy zawiera dodatkowy blok sterowania bazą „control” (zobacz: *CAKE – Komputerowy system wspomagania inżynierii wiedzy – Instrukcja użytkownika*). W czasie zapisu drzewa do bazy/źródła wiedzy nieaktywne są wszystkie polecenia oraz ikony z wyjątkiem polecenia „Drzewo decyzyjne | Przerwij pracę” (i ikony ). Polecenie to służy do przerywania obecnie wykonywanego etapu pracy w dowolnym jego momencie.

Uwagi:

- 1) w przypadku, gdy wczytany zostanie plik uczący np. „PrzykładZInstrukcji_1.lrn” wówczas baza wiedzy zostanie zapisana w tym samym katalogu pod nazwą „PrzykładZInstrukcji_1.bw”,
- 2) w przypadku, gdy wczytany zostanie plik uczący np. „PrzykładZInstrukcji_1.lrn” wówczas źródło wiedzy zostanie zapisane w tym samym katalogu pod nazwą „PrzykładZInstrukcji_1.zw”,
- 3) możliwe jest bezpośrednie uruchomienie zapisanej bazy lub źródła wiedzy z systemu DeTreeX w innych systemach (patrz punkt następny).


URUCHAMIANIE INNYCH SYSTEMÓW PAKIETU SPHINX

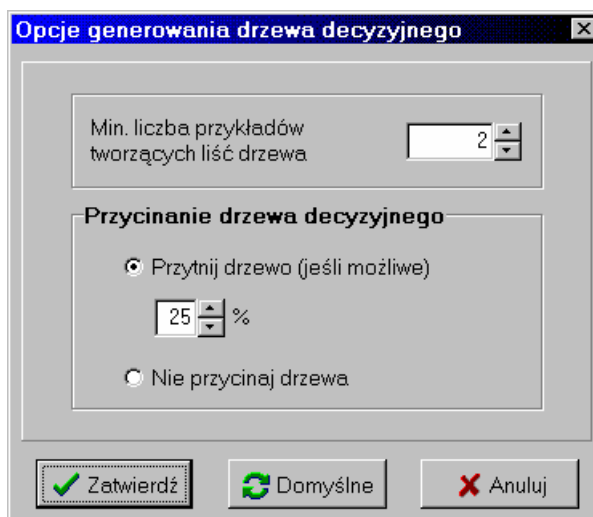
Jeśli zapisano wcześniej bazę wiedzy wówczas możliwe jest uruchomienie tej bazy w systemie ekspertowym PC-Shell wydając polecenie z menu głównego „Drzewo decyzyjne | Uruchom bazę wiedzy w PC-Shellu” lub wskazanie ikony  (Rys. 9). Jeśli zapisano wcześniej bazę i/lub źródło wiedzy wówczas możliwe jest otwarcie tej bazy i/lub źródła wiedzy w systemie CAKE wydając polecenie z menu głównego „Drzewo decyzyjne | Otwórz bazę/źródło wiedzy w CAKE’u” lub wskazanie ikony  (Rys. 9).

Uwagi:

- 1) w systemie ekspertowym PC-Shell nie jest możliwe otwarcie więcej niż jednej bazy wiedzy w tym samym czasie; w przypadku, gdy zachodzi konieczność uruchomienia kolejnej bazy wiedzy bezpośrednio z DeTreeXa podczas gdy jest już uruchomiony system PC-Shell, należy ten system zamknąć,
- 2) system CAKE nie otwiera bezpośrednio bazy/źródła wiedzy; w tej sytuacji należy samodzielnie z menu głównego CAKE’a wybrać odpowiednie polecenie otwarcia pliku bazy/źródła wiedzy lub wskazać kursorem odpowiednią ikonę na pasku narzędziowym.

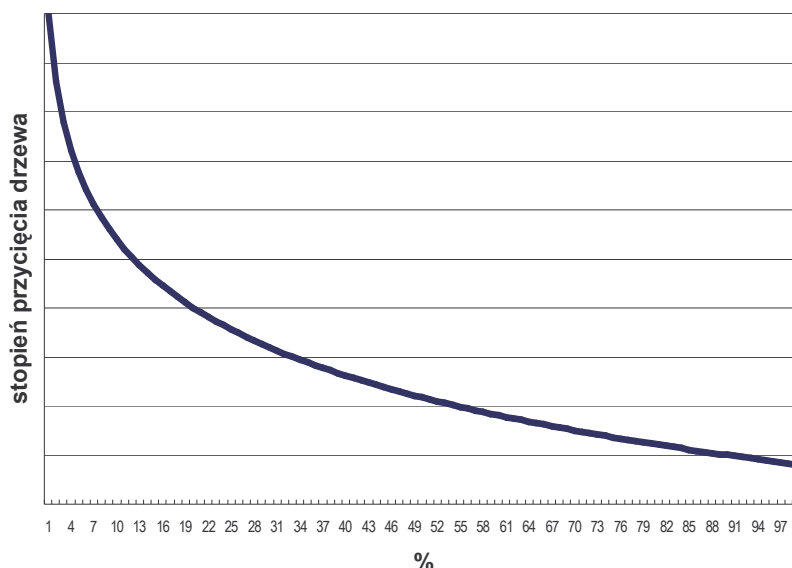
OPCJE GENEROWANIA DRZEWY DECYZYJNEGO

Po uprzednim wskazaniu pliku uczącego możliwa jest zmiana opcji generowania drzewa decyzyjnego wydając polecenie „Drzewo decyzyjne | Opcje...” lub wskazując kursorem ikonę . Polecenie to wywołuje dodatkowe okno z opcjami generowania drzewa decyzyjnego (Rys. 21).



Rys. 21. Okno dialogowe zmiany parametrów generowania drzewa decyzyjnego

Parametr „Minimalna liczba przykładów tworzących liść drzewa” umożliwia zmianę rozmiaru drzewa (a tym samym zmianę stopnia uogólnienia wiedzy reprezentowanej przez drzewo) **w trakcie budowy** tego drzewa. Im wartość tego parametru jest większa tym wygenerowane drzewo jest mniejszych rozmiarów (wiedza ogólniejsza). Wartość „1” tego parametru powoduje zbudowanie maksymalnie dopasowanego drzewa (drzewa, w którym wiedza została uogólniona w najmniejszym stopniu).



Rys. 22. Wykres zmiany stopnia przycięcia drzewa w zależności od wartości parametru przycięcia

Opcja „Przycinanie drzewa decyzyjnego” umożliwia zmniejszenie rozmiaru drzewa (a tym samym zwiększenie uogólnienia wiedzy reprezentowanej przez drzewo) **po zbudowaniu** tego drzewa. W sytuacji gdy wybrano opcję „Nie przycinaj drzewa” drzewo to nie zostanie zmniejszone i zostanie zachowany jego pierwotny kształt. W sytuacji gdy wybrano opcję „Przytnij drzewo (jeśli możliwe)”, drzewo zostanie zmniejszone zgodnie z zastosowanym parametrem. Parametr ten umożliwia wyznaczenie odpowiednich błędów klasyfikacji przed i po przycięciu drzewa. Jeśli błędy klasyfikacji po przycięciu drzewa (lub węzła drzewa) są większe od błędów przed przycięciem, wówczas drzewo (węzeł) nie jest przycinane. Wartość parametru zmienia się wykładniczo (Rys. 22) tzn. zmniejszając jego wartość uzyskujemy coraz większe przycięcie drzewa, natomiast zwiększając jego wartość uzyskujemy coraz mniejsze przycięcie.

Uwagi:


- 1) w przypadku wybrania wartości parametru „Minimalna liczba przykładów tworzących liść drzewa” np. „5” możliwe jest, iż po zbudowaniu drzewa liczba przykładów tworzących liść drzewa będzie mniejsza; nie jest to błąd systemu DeTreex; wynika to z faktu, iż poszczególne węzły drzewa mogą nie posiadać swoich poddrzew i nie jest możliwe ich zredukowanie do liścia, który zawierałby określoną przez użytkownika liczbę przykładów,
- 2) w przypadku wybrania opcji „Przytnij drzewo (jeśli możliwe)”, drzewo może nie zostać przycięte; nie jest to błąd systemu DeTreex; możliwe jest, iż dla wybranego przez użytkownika parametru przycinania, przycięcie drzewa nie odniesieżądanego skutku i ostatecznie drzewo zachowuje swój pierwotny kształt; w tej sytuacji należy zmienić (zmniejszyć) wartość tego parametru tak aby uzyskać żądany efekt przycięcia.

WSPÓŁPRACA SYSTEMU DETREEX Z BAZAMI DANYCH

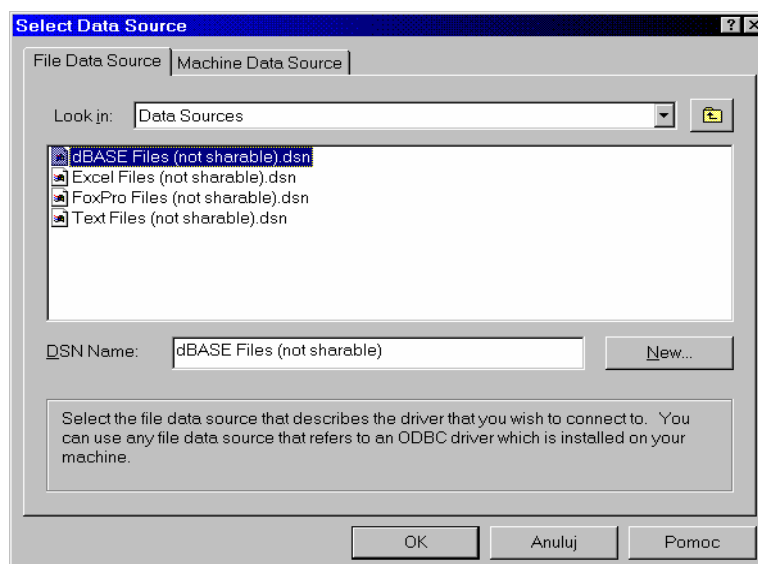
W większości przypadków dane zapisywane i przechowywane są w bazach danych. System DeTreeX umożliwia przygotowanie odpowiedniego pliku uczącego importując zapisane w takiej bazie dane. Wymagane jest aby w systemie Windows® zainstalowany był odpowiedni sterownik ODBC umożliwiający dostęp do danych w bazie.

IMPORT PRZYKŁADÓW UCZĄCYCH Z BAZY DANYCH

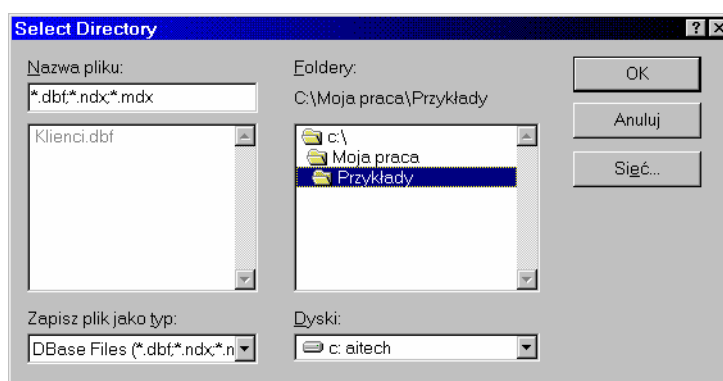
W celu zapisu pliku uczącego z danymi zawartymi w bazie danych należy postępować zgodnie z następującymi krokami:

- 1) Wybrać z menu głównego systemu polecenie „Plik | Import danych z bazy...” lub wskazać kursorem myszki ikonę . Wyświetlone zostanie okno dialogowe wyboru sterownika do bazy danych (Rys. 23).

W opisywanym przykładzie wybrany został sterownik do bazy danych w formacie dBASE (liczba i rodzaje sterowników do baz danych może różnić się od przedstawionych zależnie od konfiguracji komputera).



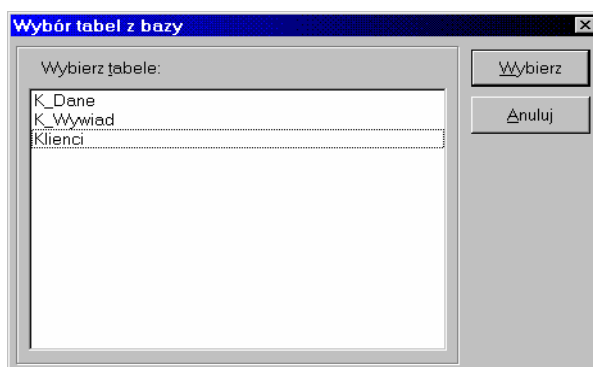
Rys. 23. Okno dialogowe wyboru sterownika do bazy danych



Rys. 24. Okno dialogowe wyboru katalogu bazy danych

- 2) Po zatwierdzeniu wyboru sterownika klawiszem „OK” ukazuje się kolejne okno dialogowe (Rys. 24). Służy ono do wskazania katalogu, w którym znajduje się odpowiednia baza danych zapisana w formacie dBASE (przedstawione okno dialogowe może różnić się wyglądem zależnie od rodzaju wybranego sterownika).

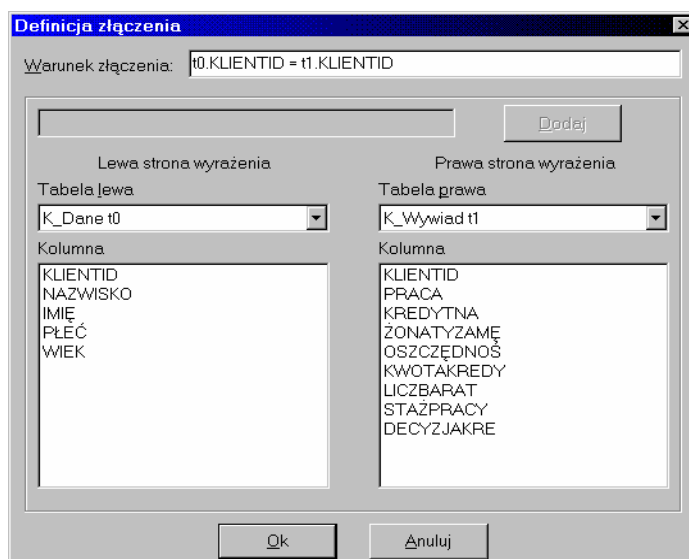
- 3) Wybrać table z bazy, z których będą importowane dane (Rys. 25).



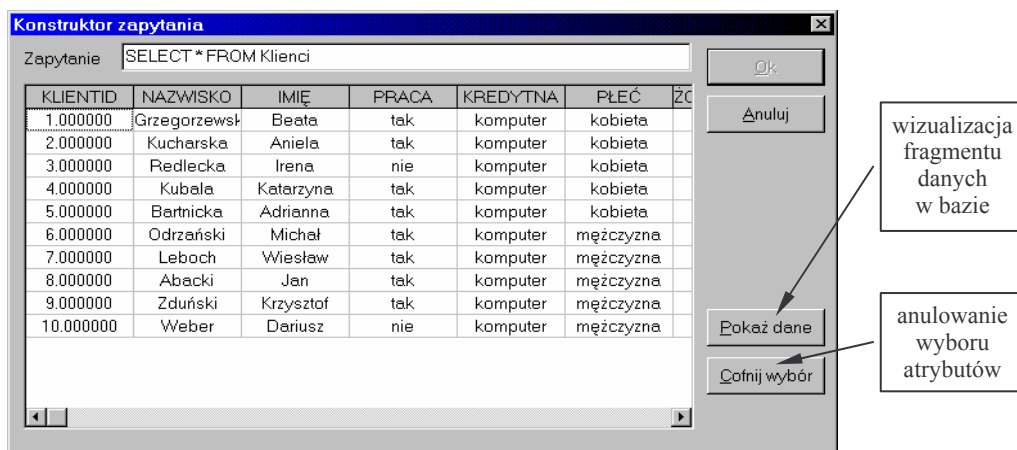
Rys. 25. Okno dialogowe wyboru tabel w bazie danych

Jeżeli zostanie wybrana tylko jedna tabela wówczas należy przejść do punktu 4. Po wskazaniu więcej niż jednej tabeli należy wykonać następujące kroki:

- Okno dialogowe „Definicja łączenia” (Rys. 26) służy do wskazania kolumn danych w tabelach, względem których nastąpi połączenie danych. W opisywanym przypadku wskazano w obu tabelach kolumny „KLIENTID”. Następnie należy nacisnąć klawisz „Dodaj” w celu zatwierdzenia warunku łączenia.
 - Po zatwierdzeniu warunku łączenia tabel klawiszem „Ok” wyświetlane jest okno dialogowe jak w punkcie 4 (Rys. 27) z tą różnicą, iż będzie ono zawierało kolumny danych z połączonych tabel.
- 4) Okno dialogowe „Kreator zapytania” (Rys. 27) umożliwia wybór atrybutów decydujących o rozwiązaniu problemu lub klasyfikacji obiektu. W tym celu należy wskazać kursorem myszki wybraną kolumnę danych i nacisnąć lewy klawisz.
- 5) Dla każdego wybranego atrybutu wyświetlone zostanie okno dialogowe (Rys. 28) pozwalające użytkownikowi na określenie, czy dany atrybut jest atrybutem wejściowym lub wyjściowym (symbole **w** i **wy** w formacie zapisu pliku uczącego) oraz czy jest to atrybut o wartościach numerycznych lub o wartościach symbolicznych (symbol „#” przed nazwą atrybutu w formacie zapisu pliku uczącego).



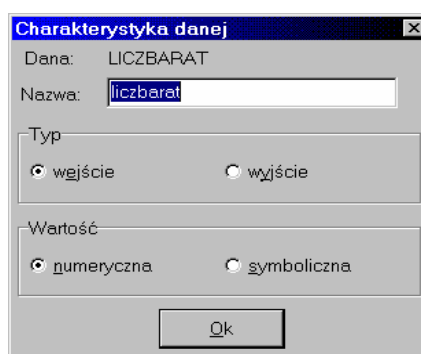
Rys. 26. Okno dialogowe łączenia danych zawartych w tabelach



Rys. 27. Okno dialogowe wyboru atrybutów

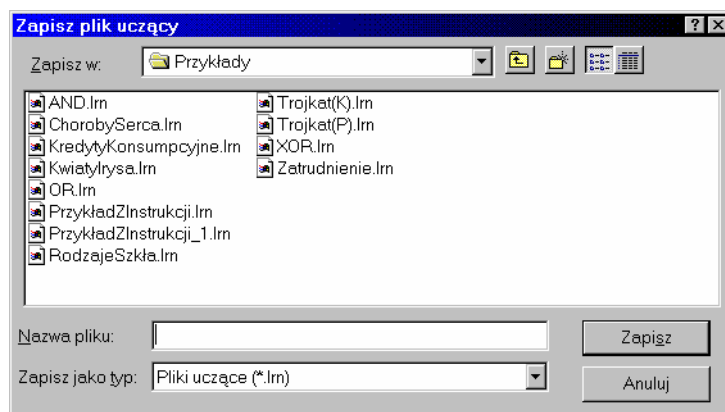
Uwagi:

- 1) w celu budowy drzewa decyzyjnego konieczne jest wybranie co najmniej jednego atrybutu wejściowego i tylko jednego atrybutu wyjściowego,
- 2) w przypadku oznaczenia atrybutu jako wyjściowy jest on automatycznie oznaczany jako atrybut o wartościach symbolicznych,
- 3) kolumny atrybutów oznaczonych jako wejściowe zmieniają swój kolor na granatowy, natomiast kolumna atrybutu wyjściowego zmienia swój kolor na żółty,
- 4) w sytuacji gdy atrybut wyjściowy nie znajduje się na końcu tabeli z danymi, wówczas w czasie zapisu pliku uczącego zostanie ten atrybut zapisany jako ostatni.



Rys. 28. Okno dialogowe charakterystyki danej

- 6) Po wskazaniu wymaganych atrybutów i zatwierdzeniu ich wyboru klawiszem „OK” (Rys. 27) ukazuje się standardowe okno dialogowe zapisu nowego pliku uczącego. Po podaniu nazwy nowego pliku i jej zatwierdzeniu, plik ten jest wyświetlany w oknie DeTreexa (lub też nie jest wyświetlany, zależnie od wybranej przez użytkownika opcji). W trakcie wpisywania nazwy pliku nie jest konieczne podawanie rozszerzenia tego pliku.



Rys. 29. Okno dialogowe zapisu nowego pliku uczącego

PRZYKŁADY DZIEDZINOWE

W rozdziale:

- Choroba serca
- Kwiaty irysa
- Zatrudnienie
- Rodzaje szkła

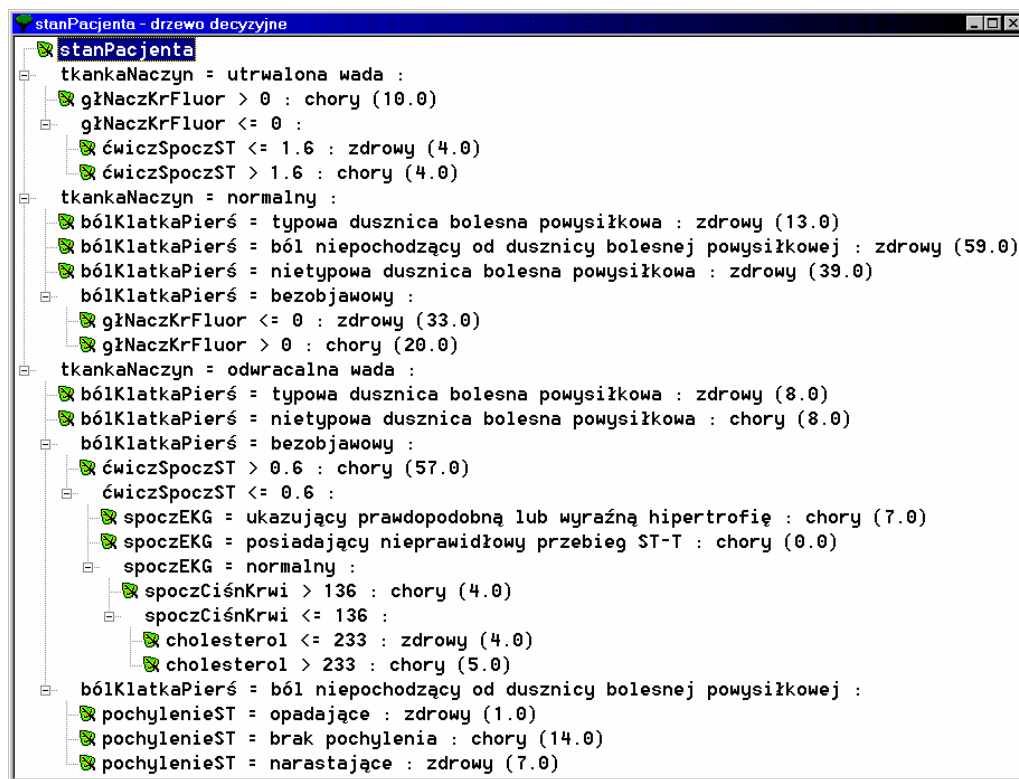
CHOROBA SERCA

- 1) Źródło przykładów: V.A. Medical Center, Long Beach and Cleveland Clinic Foundation: Robert Detrano, M.D., Ph.D.
- 2) Informacje o przykładach:
 - a) Przykłady przedstawiają wyniki diagnoz choroby serca u ludzi i mają na celu określenie czy dany pacjent ma chore serce.
 - b) Dane pacjentów zostały usunięte z przykładów.
- 3) Liczba przykładów: $303 - 5 = 297$ (usunięto 5 przykładów z nieznanymi wartościami atrybutów).
- 4) Opis atrybutów:

wiek:	wiek w latach;
płeć:	kobieta, mężczyzna;
bólKlatkaPierś:	typ bólu w klatce piersiowej: <ul style="list-style-type: none"> - typowa dusznica bolesna powysiłkowa, - nietypowa dusznica bolesna powysiłkowa, - ból nie pochodzący od dusznicy bolesnej powysiłkowej, - bezobjawowy;
spoczCiśnKrw:	ciśnienie krwi w czasie spoczynku (w mm Hg, po przyjęciu do szpitala);
cholesterol:	cholesterol w osoczu krwi (w mg/dcl);
stałyCukKrw:	stały cukier we krwi > 120 mg/dcl: <ul style="list-style-type: none"> - tak, - nie;
spoczEKG:	wyniki elektrokardiografu w czasie spoczynku: <ul style="list-style-type: none"> - normalny, - posiadający nieprawidłowy przebieg ST-T (przebieg T jest odwrócony i/lub wzniesienie lub obniżenie ST > 0.05 mV), - ukazujący prawdopodobną lub wyraźną hipertrofię (przerost) lewej komory serca wg kryteriów Estes,
maksTętno:	osiągnięte maksymalne tętno;
ćwiczDuszBol:	ćwiczenia wywołują dusznicę bolesną powysiłkową: <ul style="list-style-type: none"> - tak, - nie;
ćwiczSpoczST:	obniżenie ST wywołane ćwiczeniami w stosunku do ST w czasie spoczynku;
pochylenieST:	pochylenie wzniesienia przebiegu ST w trakcie ćwiczeń: <ul style="list-style-type: none"> - narastające, - brak pochylenia, - opadające;
głNaczKrFluor:	liczba głównych naczyń krwionośnych (0-3) zabarwionych przez flouroskopię;
tkankaNaczyn:	tkanka naczyniowa: <ul style="list-style-type: none"> - normalna, - utrwalona wada, - odwracalna wada;

- 5) Klasyfikowane obiekty: diagnostyka choroby serca (wynik angiografii) – pacjent:
- zdrowy (< 50% zwężenia średnicy),
 - chory: (> 50% zwężenia średnicy).

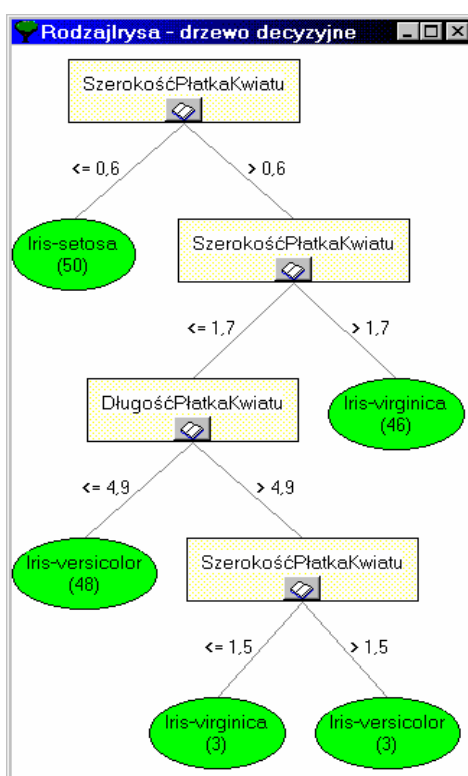
Rys. 30 przedstawia przykład drzewa decyzyjnego.



Rys. 30. Przykładowe drzewo decyzyjne – choroba serca

KWIATY IRYSA

- 1) Źródło przykładów: Michael Marshall (MARSHALL%PLU@io.arc.nasa.gov).
- 2) Informacje o przykładach: Dane przedstawiają przykłady klasyfikacji kwiatów irysa i mają na celu rozróżnienie tych kwiatów.
- 3) Liczba przykładów: 150.
- 4) Opis atrybutów:
 - DługośćDziałkiKielicha: wartości numeryczne [cm];
 - SzerokośćDziałkiKielicha: wartości numeryczne [cm];
 - DługośćPłatkaKwiatu: wartości numeryczne [cm];
 - SzerokośćPłatkaKwiatu: wartości numeryczne [cm];
- 5) Klasyfikowane obiekty: Iris Setosa, Iris Versicolor, Iris Virginica.



Rys. 31. Przykładowe drzewo decyzyjne – klasyfikacja kwiatów irysa

Rys. 31 przedstawia przykład drzewa decyzyjnego.

ZATRUDNIENIE

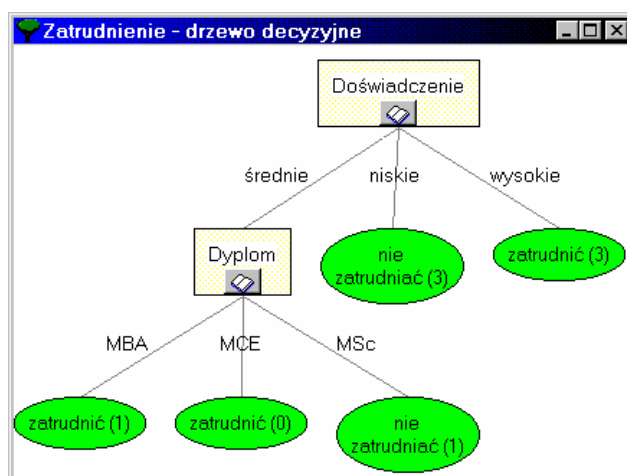
- 1) Źródło przykładów: nieznane.
- 2) Informacje o przykładach: Dane przedstawiają pozytywnie i negatywnie rozpatrzone wnioski osób ubiegających się o pracę i mają na celu określenie czy zatrudnić nowego pracownika.
- 3) Liczba przykładów: 8.
- 4) Opis atrybutów:

Dyplom: MCE (mgr budownictwa lądowego i wodnego),
MBA (mgr administracji i zarządzania),
MSc (mgr nauk ścisłych);

Doświadczenie: niskie, średnie, wysokie (Posiadane doświadczenie przez daną osobę);

Jangielski: tak, nie (Czy dana osoba zna j. angielski?);

Referencje: neutralne, dobre, wyśmienite (Posiadane referencje z poprzedniego miejsca zatrudnienia);
- 5) Klasyfikowane obiekty: zatrudnić, nie zatrudniać.



Rys. 32. Przykładowe drzewo decyzyjne – zatrudnienie

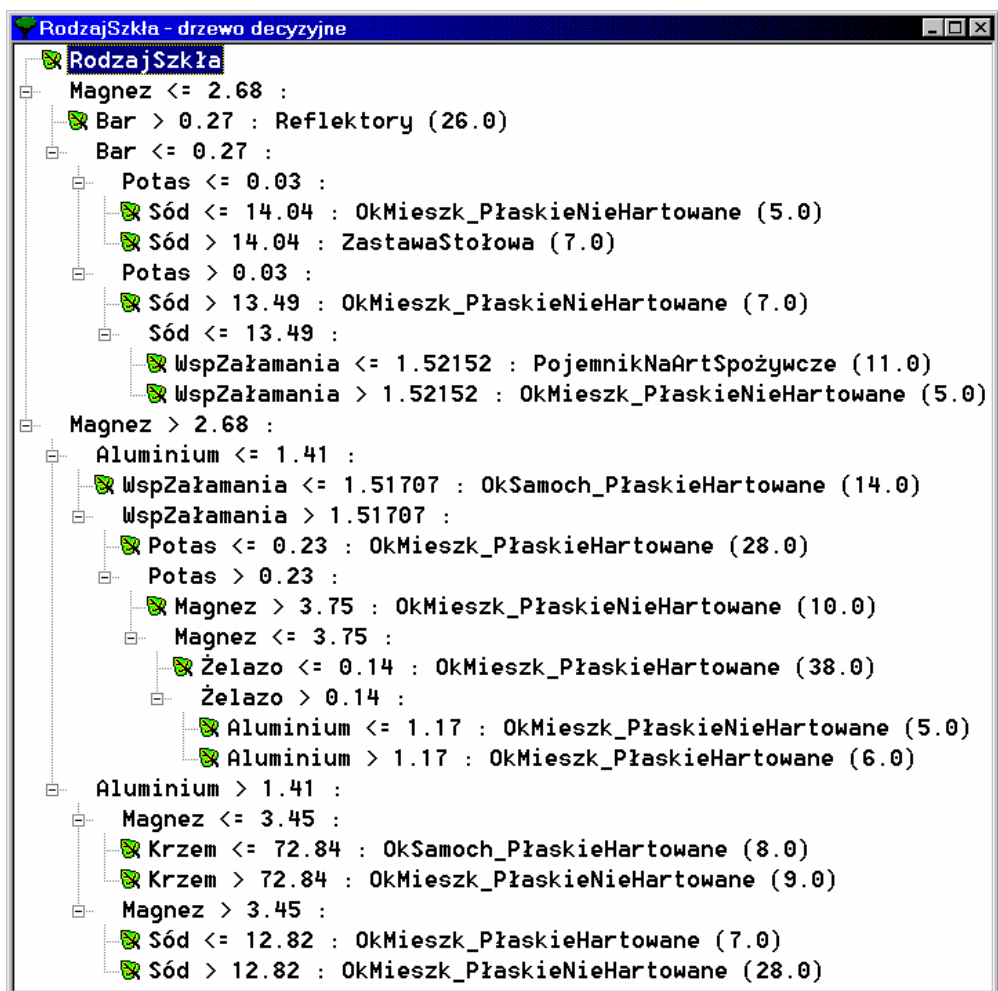
Rys. 32 przedstawia przykład drzewa decyzyjnego.

RODZAJE SZKŁA

- 1) Źródło przykładów: Vina Spiehler, Ph.D., DABFT
Diagnostic Products Corporation
(213) 776-0180 (ext 3014).
- 2) Informacje o przykładach: Dane przedstawiają przykłady klasyfikacji różnych rodzajów szkła i mają na celu określenie rodzaju szkła m.in. na podstawie składu chemicznego (klasyfikacja stosowana np. w badaniach kryminalistycznych).
- 3) Liczba przykładów: 214.
- 4) Opis atrybutów:

wspZałamania:	współczynnik załamania światła;
sód:	wartość numeryczna [%-owy udział ciężaru w tlenku];
magnez:	wartość numeryczna [%-owy udział ciężaru w tlenku];
aluminium:	wartość numeryczna [%-owy udział ciężaru w tlenku];
krzem:	wartość numeryczna [%-owy udział ciężaru w tlenku];
potas:	wartość numeryczna [%-owy udział ciężaru w tlenku];
wapń:	wartość numeryczna [%-owy udział ciężaru w tlenku];
bar:	wartość numeryczna [%-owy udział ciężaru w tlenku];
żelazo:	wartość numeryczna [%-owy udział ciężaru w tlenku].
- 5) Klasyfikowane obiekty: OkMieszk_PłaskieHartowane,
OkMieszk_PłaskieNieHartowane,
OkSamoch_PłaskieHartowane,
PojemnikNaArtSpożywcze,
ZastawaStołowa,
Reflektory.

Rys. 33 przedstawia przykład drzewa decyzyjnego.



Rys. 33. Przykładowe drzewo decyzyjne – rodzaje szkła

LITERATURA

- [Cholewa, Pedrycz, 1987] W. Cholewa, W. Pedrycz: *Systemy doradcze*, skrypt uczelniany Politechniki Śląskiej nr 1447, Gliwice, 1987.
- [Maniak, 1999] P. Maniak: *Badanie przydatności metod indukcyjnych do pozyskiwania wiedzy projektowej*, rozprawa doktorska, Katedra PKM, Politechnika Śląska, Gliwice, 1999.
- [Michalski, 1983] R. S. Michalski: *A Theory and Methodology of Inductive Learning*, Artificial Intelligence 20 (1983), pp. 111-161.
- [Michie, 1987] D. Michie: *Current Developments in Experts Systems*, [W:] J. R. Quinlan (Ed.) *Applications of Expert Systems*, Turing Institute Press in association with Addison-Wesley Publishing Company, 1987.
- [Øhrn, Komorowski, Skowron, Synak, 1998] A. Øhrn, J. Komorowski, A. Skowron, P. Synak: *The Design and Implementation of a Knowledge Discovery Toolkit Based on Rough Sets – The Rosetta System*, [W:] L. Polkowski and A. Skowron (Eds.): *Rough Sets in Knowledge Discovery*, Physica Verlag, 1998.
- [Quinlan, 1986] J. R. Quinlan: *Induction of Decision Trees*, Machine Learning, 1 (1986), pp. 81-106.
- [Quinlan, 1993] J. R. Quinlan: *C4.5 Program for Machine Learning*, Morgan Kaufmann, San Mateo, CA, 1993.
- [Wyleżoł, 1996] M. Wyleżoł: *Zastosowanie drzew decyzyjnych do identyfikacji wiedzy konstrukcyjnej*, praca dyplomowa magisterska, Katedra PKM, Politechnika Śląska, Gliwice, 1996.