METODY INŻYNIERII WIEDZY

Wprowadzenie do Metod Inżynierii Wiedzy

Adrian Horzyk

horzyk@agh.edu.pl

AGH

AGH University of Science and Technology
Krakow, Poland

Pytania

Czym jest wiedza?

Jak by Państwo zdefiniowali pojęcie wiedzy dla potrzeb informatyki?

Z jakimi problemami mamy do czynienia operując na danych?

Czego byście się Państwo chcieli dowiedzieć w trakcie tego kursu?

Zakres

- ✓ Eksploracja danych (*data mining*)
- ✓ Inżynieria wiedzy (*knowledge engineering*)
- ✓ Nauka o danych (*data science*)
- ✓ Algorytmy bezpośredniej eksploracji wiedzy z danych (poprzez ich przeszukiwanie)
- ✓ Tworzenie struktur skojarzeniowych pozwalających efektywnie modelować i szybko eksplorować wiedzę z danych na podstawie modelu wiedzy.
- ✓ Algorytmy asocjacyjnej eksploracji wiedzy z danych
- ✓ Algorytmy klasteryzacji i klasyfikacji danych
- ✓ Wnioskowanie oparte o wiedzę

Ewaluacja, Oceny i Egzamin

Przedmiot obejmuje:

- ✓ **Wykłady – w trakcie których będzie można zdobyć wiedzę teoretyczną pozwalającą na modelowanie i implementację systemów opartych o wiedzę, eksplorację danych, wnioskowanie, klasteryzację i klasyfikację danych.**
- ✓ **Ćwiczenia – w trakcie których będą implementowane wybrane rozwiązania w postaci kilku prostszych oraz jednego bardziej skomplikowanego zadania.**

Zaliczenie ćwiczeń laboratoryjnych obejmuje:

- ✓ **Implementację wszystkich zadań dotyczących eksploracji danych, klasteryzacji, klasyfikacji, kojarzenia i wnioskowania;**
- ✓ **Przygotowanie końcowej prezentacji zastosowanych metod i uzyskanych wyników, jak również ich interpretację.**
- ✓ **Wybrane ostatnie (skomplikowane) zadanie zaliczeniowe powinno prezentować rozwiązanie (strukturę) i wyniki w postaci graficznej.**

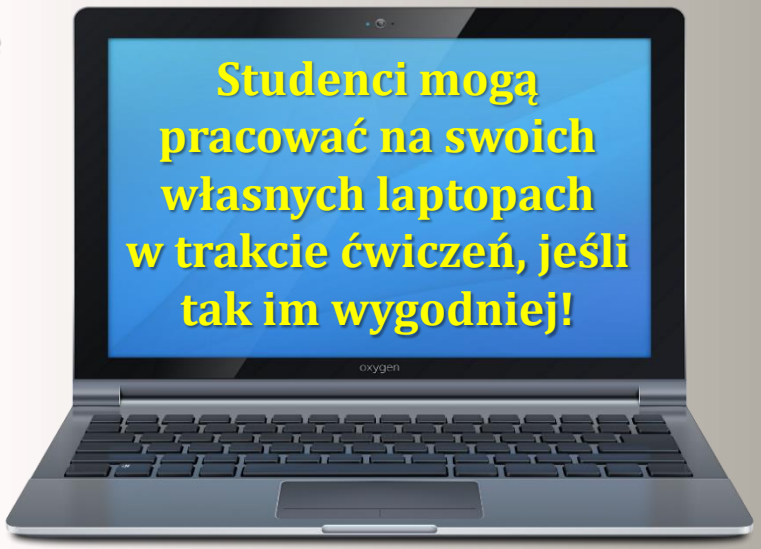
Wykład kończy się egzaminem obejmującym wiedzę teoretyczną z wykładów.

Realizacje zadań w trakcie ćwiczeń

Wszystkie zadania (aplikacje) powinny być zrealizowane w jednym z wiodących języków obiektowych, tj.: C#, C++, Java, Python, lub PHP oraz korzystać z baz danych. W laboratorium komputerowym dostępne są tylko środowiska: MS Visual Studio, Java i Python, więc w razie chęci skorzystania z innego środowiska lub języka programowania, proszę o wykorzystanie swojego własnego laptopa podczas ćwiczeń i prezentacji.

W trakcie ostatnich ćwiczeń w semestrze planowane są ok. 10 min. prezentacje studentów:

- ✓ **Przygotuj** swoją ostateczną **prezentację** zastosowanych metod, struktur danych, osiągniętych wyników itd.
- ✓ **Wyciągnij** dla nas cenne **wnioski** i **podsumowania**.
- ✓ **Inspiruj** nas, udziel porad i sugestii!
- ✓ **Opisz możliwości** swojego rozwiązania i pokaż nam, jak to działa dla przykładowych danych.
- ✓ **Zinterpretuj** i **porównaj** osiągnięte **wyniki**.



Ostateczne rozwiązania zadań z ćwiczeń (kody źródłowe wszystkich zadań, skompilowane aplikacje (*.exe), przykładowe dane, bazy danych i prezentacje końcowe itp.) należy przesłać prowadzącemu pod koniec semestru przed uzyskaniem oceny, gdyż wykładowca jest zobowiązany do przechowywania prac studentów co najmniej przez jeden rok jako dowód wystawionych ocen. W związku z tym tylko wysłane kompletne projekty mogą być ostatecznie ocenione, a oceny wpisane!

Od danych i informacji, do wiedzy i inteligencji!

Jak stworzyć i rozwijać komputerowe inteligentne systemy oparte o wiedzę?

Dane

Dane to zbiory liczb, znaków, symboli, sygnałów, bodźców, miar fizycznych lub empirycznych oraz surowych wartości opisujących różne przedmioty lub działania, np. *36.6ºC, T, $, φ, 25cm, !*

Niepowiązane dane są bezużyteczne, ponieważ dane przyjmują znaczenie, gdy są powiązane z innymi danymi. Dane mogą być surowe, niespójne, niezorganizowane…

Dane zazwyczaj opisują fakty i są nośnikiem informacji.

Tabele Danych

W informatyce stosujemy przede wszystkim tabele do przechowywania i zarządzania danymi,

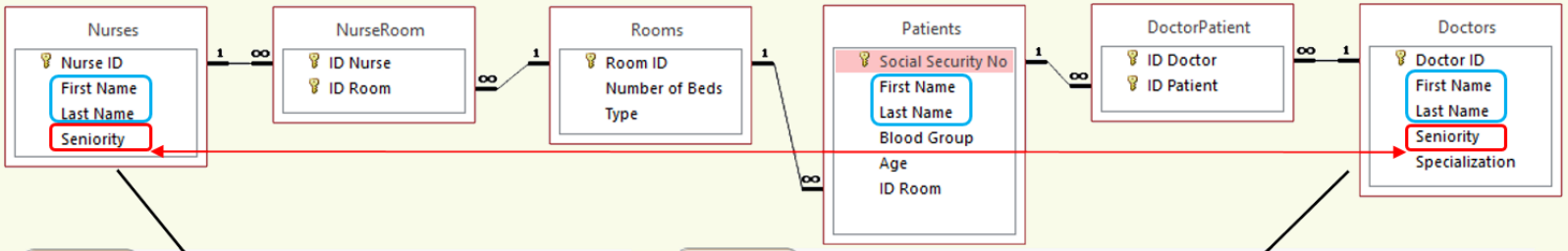
	ATTRIBUTES				
SAMPLE OBJECTS	SEPAL LENGTH	SEPAL WIDTH	PETAL LENGTH	PETAL WIDTH	CLASS LABEL
O1	5.4	3.0	4.5	1.5	Versicolor
O2	6.3	3.3	4.7	1.6	Versicolor
O3	6.0	2.7	5.1	1.6	Versicolor
O4	6.7	3.0	5.0	1.7	Versicolor
O5	6.0	2.2	5.0	1.5	Virginica
O6	5.9	3.2	4.8	1.8	Versicolor
O7	6.0	3.0	4.8	1.8	Virginica
O8	5.7	2.5	5.0	2.0	Virginica
O9	6.5	3.2	5.1	2.0	Virginica

lecz zwykle wykorzystywane relacje podczas wnioskowania, tj. podobieństwo, minima, maksima, ilość duplikatów, trzeba wyszukiwać, co wiąże się z wymiernym kosztem czasowym. Im więcej jest danych, tym więcej czasu i zasobów tracimy na samo wyszukiwanie relacji, które służą do dalszego przetwarzania danych i wnioskowania!

Such relations are not enough!

Relacyjne Bazy Danych

Relacyjne bazy danych wiążą dane tylko w poziomie, a nie w pionie (w tabelach), więc nadal musimy szukać duplikatów, sąsiadów lub podobnych wartości i obiektów.



Nurses

Nurse ID	First Name	Last Name	Seniority
N1	Amy	Moon	12
N2	Rose	Jolie	18
N3	Kate	Ford	24
N4	Lisa	Brown	9
N5	Sara	Pitt	4
N6	Kate	Lopez	12

Doctors

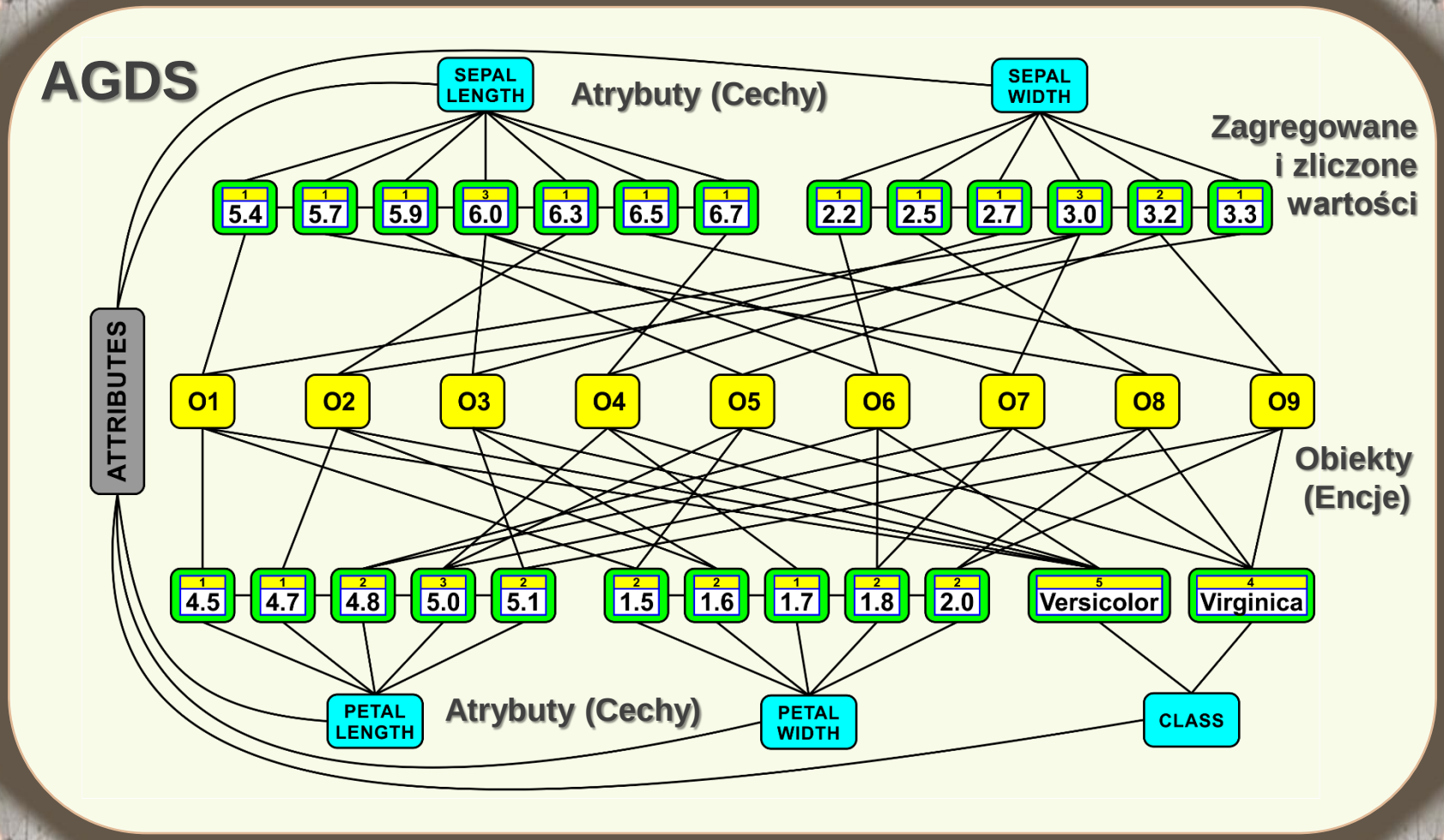
Doctor ID	First Name	Last Name	Seniority	Specialization
D1	Tom	Hanks	18	orthopedics
D2	Jack	Brown	15	surgery
D3	Lisa	Ford	23	pediatrician
D4	Tom	Trump	35	pediatrician
D5	Kate	Smith	7	surgery
D6	Amy	Hanks	12	surgery

Dane nie są perfekcyjnie powiązane nawet w poziomie, więc zwykle występuje wiele niepowiązanych duplikatów tych samych kategorii w różnych tabelach
W rezultacie tracimy dużo czasu obliczeniowego (i pamięci), aby wyszukać niezbędne relacje pomiędzy danymi, aby obliczyć wyniki lub wyciągnąć wnioski.

Czy to rozsądne tracić większość czasu obliczeń na poszukiwanie relacji pomiędzy danymi?!

AGDS

Asocjacyjne Grafowe Struktury Danych



Połączenia reprezentują różne relacje między elementami AGDS, tj. podobieństwo, bliskość, sąsiedztwo, definicja itp.

Eksploracja Danych

Eksploracja danych zajmuje się wyszukiwaniem powtarzających się wzorców danych, tzw. **wzorców częstych (*frequent patterns*)**, w celu określenia ich przydatności lub wyprowadzenia wniosków.

Wzorce danych mogą mieć różną postać:

- zbiorów (wzorce proste),
- sekwencji (wzorce sekwencyjne),
- struktur złożonych, np. desenie, tekstury, mapy, obrazy 2D, 3D (wzorce strukturalne).

Czasami ciekawe mogą być również **wzorce rzadkie (*infrequent patterns*)** lub **unikalne (*unique patterns*)**, do wykrycia nowych i ciekawych zjawisk, np. przyrodniczych czy ekonomicznych.

Eksplorację danych możemy wykonywać:

- bezpośrednio (przeszukując dane),
- z wykorzystaniem modelu wiedzy.

Informacja

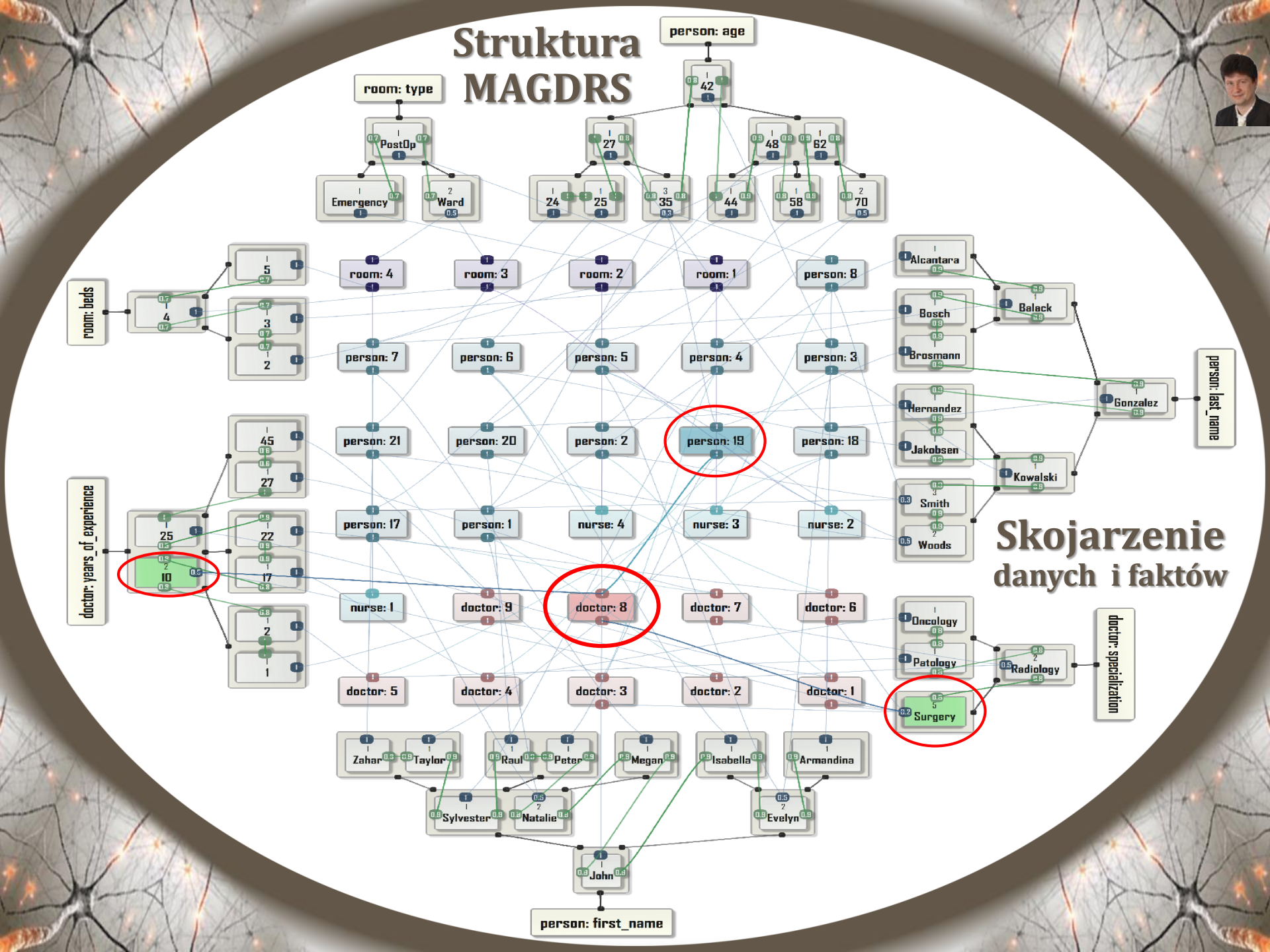
Fakt to zbiór uporządkowanych, spójnych i powiązanych danych.

Informacja to zbiór powiązanych danych (faktów) odbieranych przez odbiorcę, dla którego te dane mają określone znaczenie w kontekście posiadanej wiedzy, a odbiorca zmienia swój stan pod wpływem tych danych i/lub jego wiedza jest na ich podstawie aktualizowana lub poszerzana,

np. *normalna temperatura ciała człowieka wynosi 36,6 ºC*.

Informacja tworzy nowe lub modyfikuje istniejące powiązania pomiędzy znanymi obiektami oraz nowymi danymi.

Odbiorca informacji musi być w stanie powiązać te dane tak, aby zrozumieć przekazywaną informację.


Struktura
MAGDRS
person: age
room: type
PostOp
Emergency
Ward
42
27
48
62
24
25
35
44
58
70
room: beds
5
4
3
2
room: 4
room: 3
room: 2
room: 1
person: 8
person: 7
person: 6
person: 5
person: 4
person: 3
person: 21
person: 20
person: 2
person: 19
person: 18
person: 17
person: 1
nurse: 4
nurse: 3
nurse: 2
nurse: 1
doctor: 9
doctor: 8
doctor: 7
doctor: 6
doctor: 5
doctor: 4
doctor: 3
doctor: 2
doctor: 1
doctor: years_of_experience
45
27
25
22
10
17
7
1
Alcantara
Balack
Bosch
Brosmann
Gonzalez
Hernandez
Jakobsen
Kowalski
Smith
Woods
person: last_name
Oncology
Patology
Radiology
Surgery
doctor: specialization
Zahar
Taylor
Raul
Peter
Megan
Isabella
Armandina
Sylvester
Natalie
Evelyn
John
person: first_name
Skojarzenie
danych i faktów

Poznanie

Poznanie jest działaniem mentalnym prowadzącym do pozyskiwania wiedzy z danych i ich rozumienia poprzez procesy myślowe, doświadczenie i zmysły.

Poznanie obejmuje wiele aspektów funkcji i procesów intelektualnych, tj. uwaga, kształtowanie wiedzy, pamięć, ocena i ewaluacja, rozumowanie, rozwiązywanie problemów i podejmowanie decyzji, rozumienie, przetwarzanie i posługiwanie się językiem.

Procesy kognitywne wykorzystują istniejącą wiedzę i generują nową wiedzę dla przetwarzanych danych.

CZYM JEST WIEDZA?

Dane bombardują nas z każdej strony!

A nasze mózgi jakoś sobie z nimi radzą!

Wiedza

Wiedza jest abstrakcyjnym rezultatem kontekstowej, skojarzeniowej konsolidacji i reprezentacji wzorców, faktów i reguł oraz ich uogólniania, tworzącym nowe metody, reguły i algorytmy przetwarzania danych oraz wnioskowania.

W informatyce może być postrzegana jako zbiór informacji z powiązanym kontekstem, który występuje w formie relacji między różnymi, zebranymi w czasie informacjami.

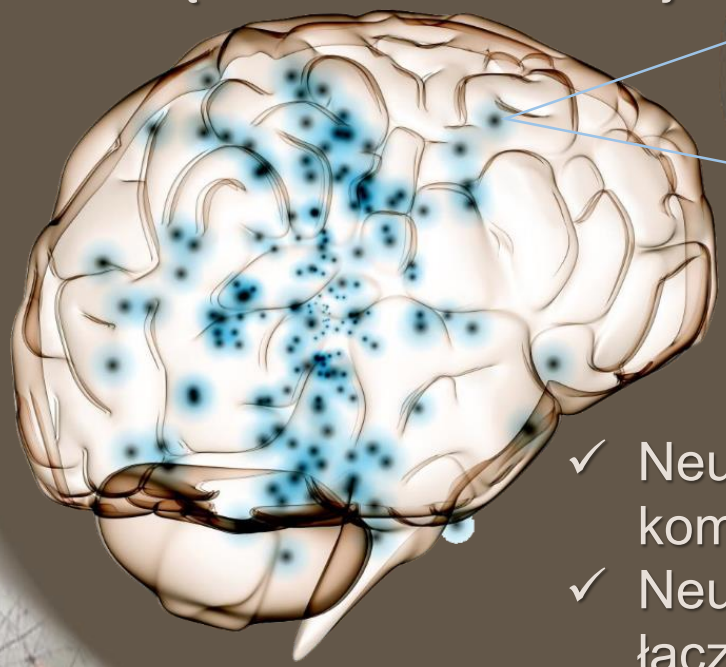
Wiedza jest ściśle związana z inteligencją, ponieważ umożliwia wnioskowanie i rozwój indywidualnej inteligencji, jak również wyodrębnienie własnego bytu i jestestwa.

Wiedza

Wiedza formuje się w naszym umyśle
i powstaje jako skutek aktywności skojarzeniowej zachodzącej w mózgu, który jest również siedliskiem naszej inteligencji.

Relacje pomiędzy danymi

Możemy poszukać rozwiązania
w strukturach mózgu,
w którym przechowywane
są dane oraz ich relacje.



- ✓ Neurony mogą reprezentować dowolny podzbiór kombinacji danych wejściowych, które je aktywują.
- ✓ Neuronalne procesy plastyczności automatycznie łączą neurony i wzmacniają połączenia, które reprezentują powiązane dane i obiekty.

Skorzystajmy z biologicznie zoptymalizowanego
rozwiązania do formowania wiedzy!

Wiedza

Wiedza tworzy się w wyniku:

- aktywnej nauki (teoretycznej),
- zdobywania doświadczenia (praktycznego),
- introspekcji i rozważań (wewnętrznych),
- intuicji (pasywnej i nieświadomej formy).

Wiedza przekazywana jest poprzez:

- wzorce oraz zbiory faktów i reguł, które nazywamy danymi uczącymi (*training data*).

Model wiedzy powstaje poprzez:

- skojarzeniowe i kontekstowe powiązanie ze sobą wzorców, faktów i reguł (danych uczących), pozwalających na uogólnianie i wyprowadzanie wniosków oraz tworzenie nowych reguł i algorytmów;
- wykorzystanie aktywnego mechanizmu wnioskowania wykorzystującego utrwalone relacje pomiędzy danymi i obiektami.

Czym nie jest wiedza!

Wiedza nie jest żadną:

~ kolekcją informacji, faktów, wzorców, danych ani reguł,

~ formą pamięci,

~ strukturą danych.

Wiedza nie zapewnia:

~ dokładnego zapamiętania wszystkich danych, wzorców, faktów ani reguł, lecz je konsoliduje, wiąże i uogólnia.

Wiedza nie skupia się:

~ na danych lecz na łączących je relacjach, na podstawie których umożliwia wnioskowanie.

MONKEY

poniżej przykład zestawu faktów opisujących tę małpkę:

”I have a monkey. My monkey is very small.
It is very lovely. It likes to sit on my head.
It can jump very quickly. It is also very clever.
It learns quickly. My monkey is lovely.
I have also a small dog.”

Jaką wiedzę na podstawie opisu zgromadziliśmy na temat tej małpki?
Spróbujmy teraz odpowiedzieć na pytanie: What is this monkey like?

GRAF WIEDZY

Konstrukcja asocjacyjnego grafu neuronowego dla następującego zestawu wzorców sekwencyjnych:

1x S1 I HAVE A MONKEY

1x S2 MY MONKEY IS VERY SMALL

1x S3 IT IS VERY LOVELY

1x S4 IT LIKES TO SIT ON MY HEAD

1x S5 IT CAN JUMP VERY QUICKLY

1x S6 IT IS ALSO VERY CLEVER

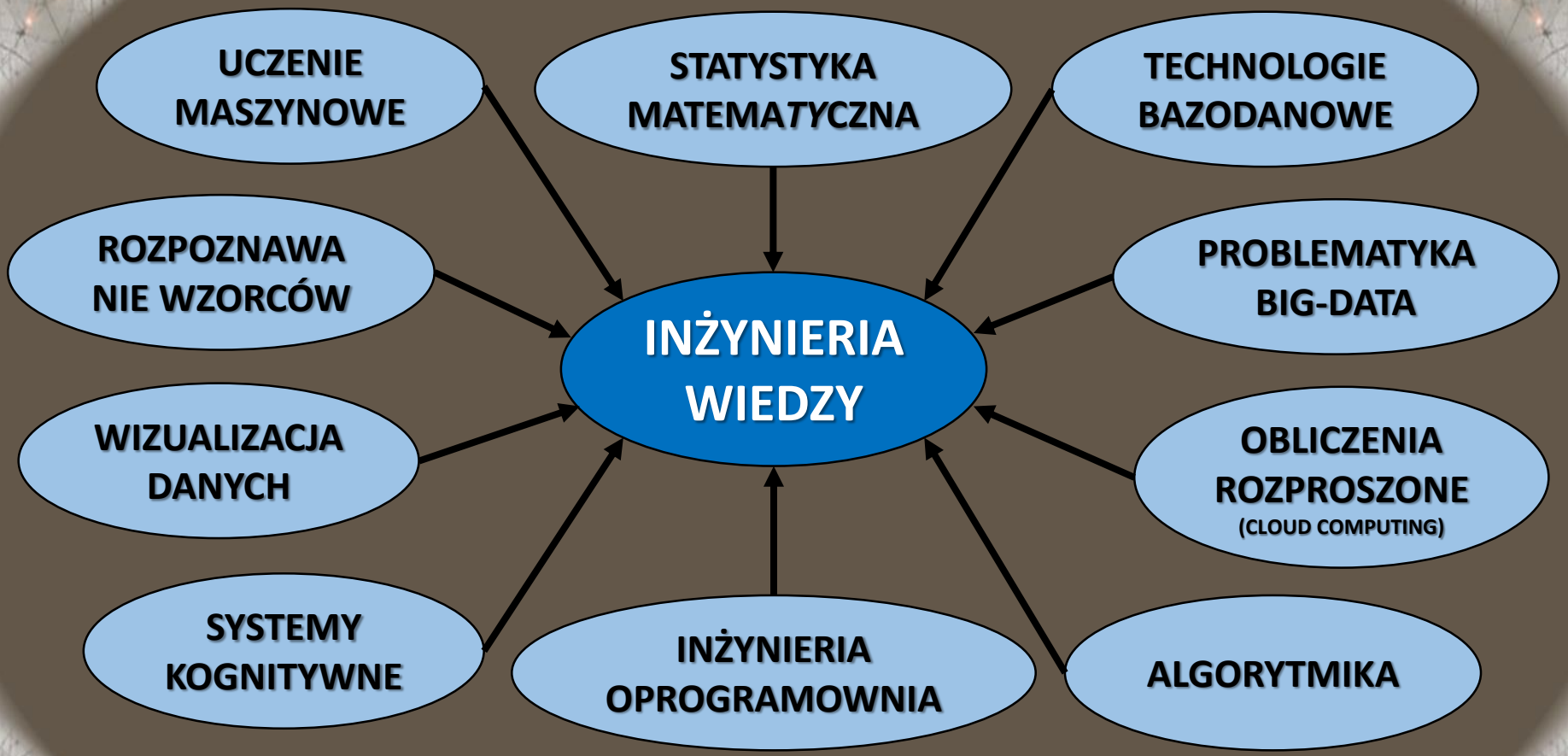
1x S7 IT LEARNS QUICKLY

1x S8 MY MONKEY IS LOVELY

1x S9 I HAVE ALSO A SMALL DOG

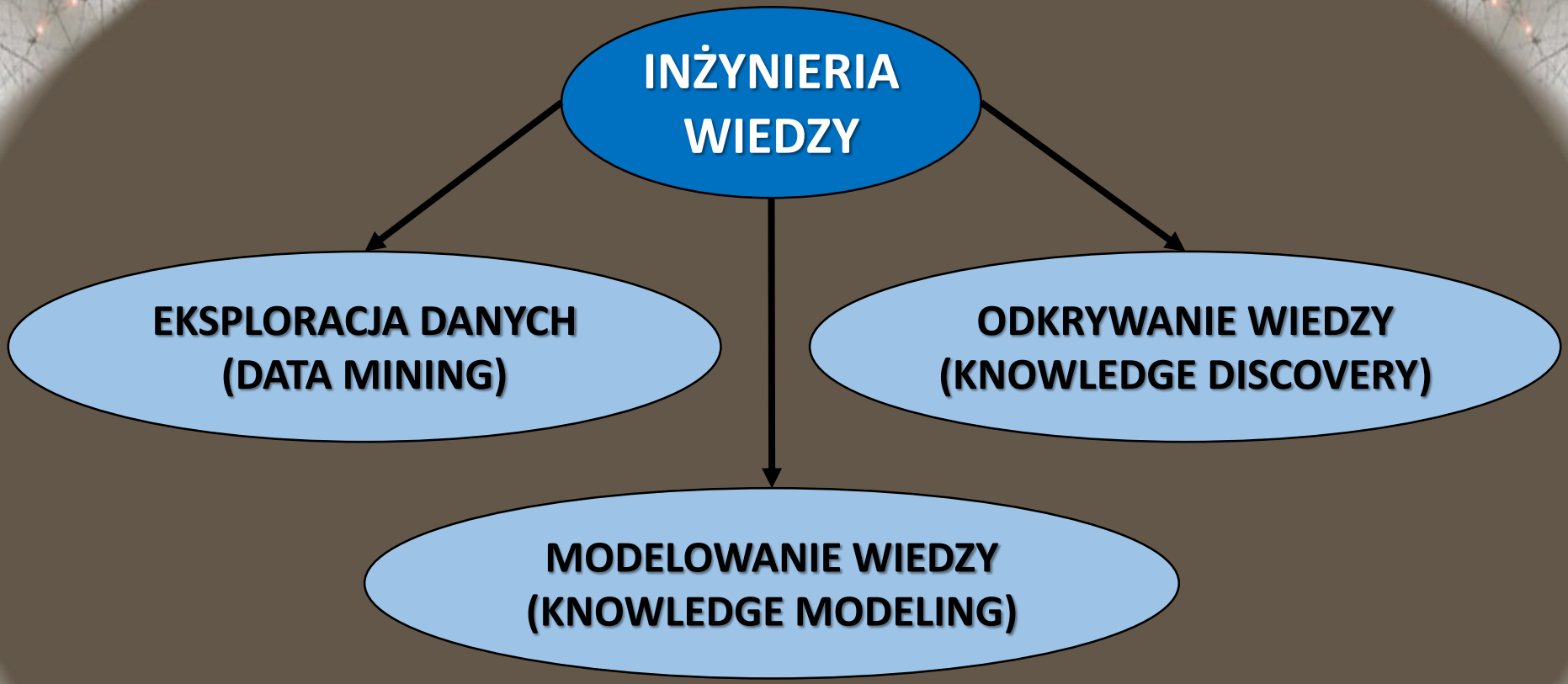

Open Data
Parameters
Learn
Learn Monkey
Learn Slower
2
Learn Faster
Balance Graph
View -
100
View +
Ask
Ask Monkey
Ask Faster
5
Ask Slower
Weights
Background
Exit
Ask times:
1
INPUT:
(Enter a new sentence)
SIMTIME:
0
OUTPUT:
(Results will be brought back by the ANAKG after asking)
Google: Adrian Horzyk (http://home.agh.edu.pl/~horzyk/index-eng.php) to cooperate in developing associative knowledge-based artificial intelligence solutions: horzyk@agh.edu.pl

Inżynieria Wiedzy



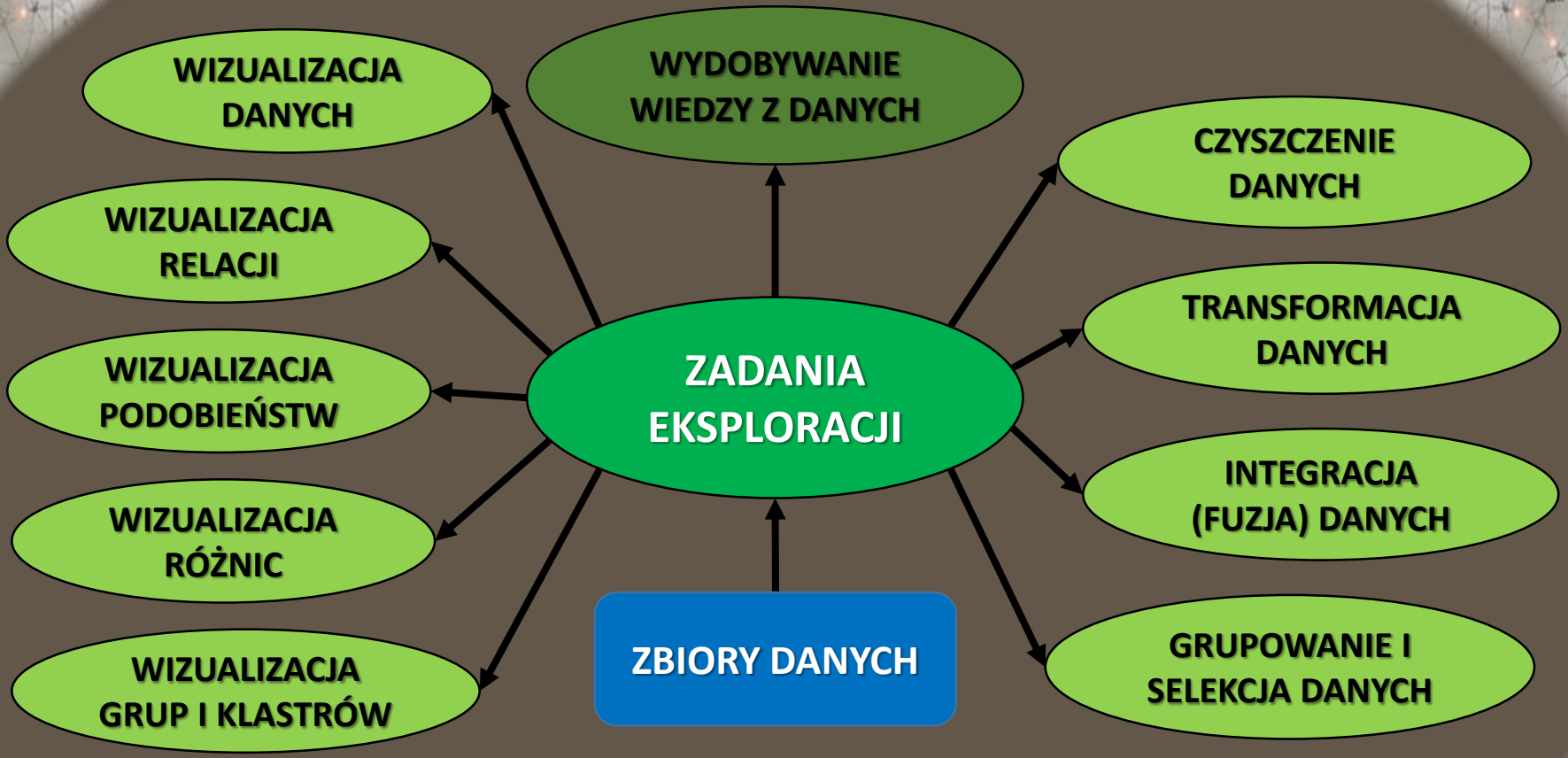
Inżynieria wiedzy to obszar informatyki zajmujący się metodami eksploracji, reprezentacji i modelowania wiedzy z danych (ich zbiorów, reguł, baz danych) oraz metodami wnioskowania na ich podstawie.

Odkrywanie Wiedzy



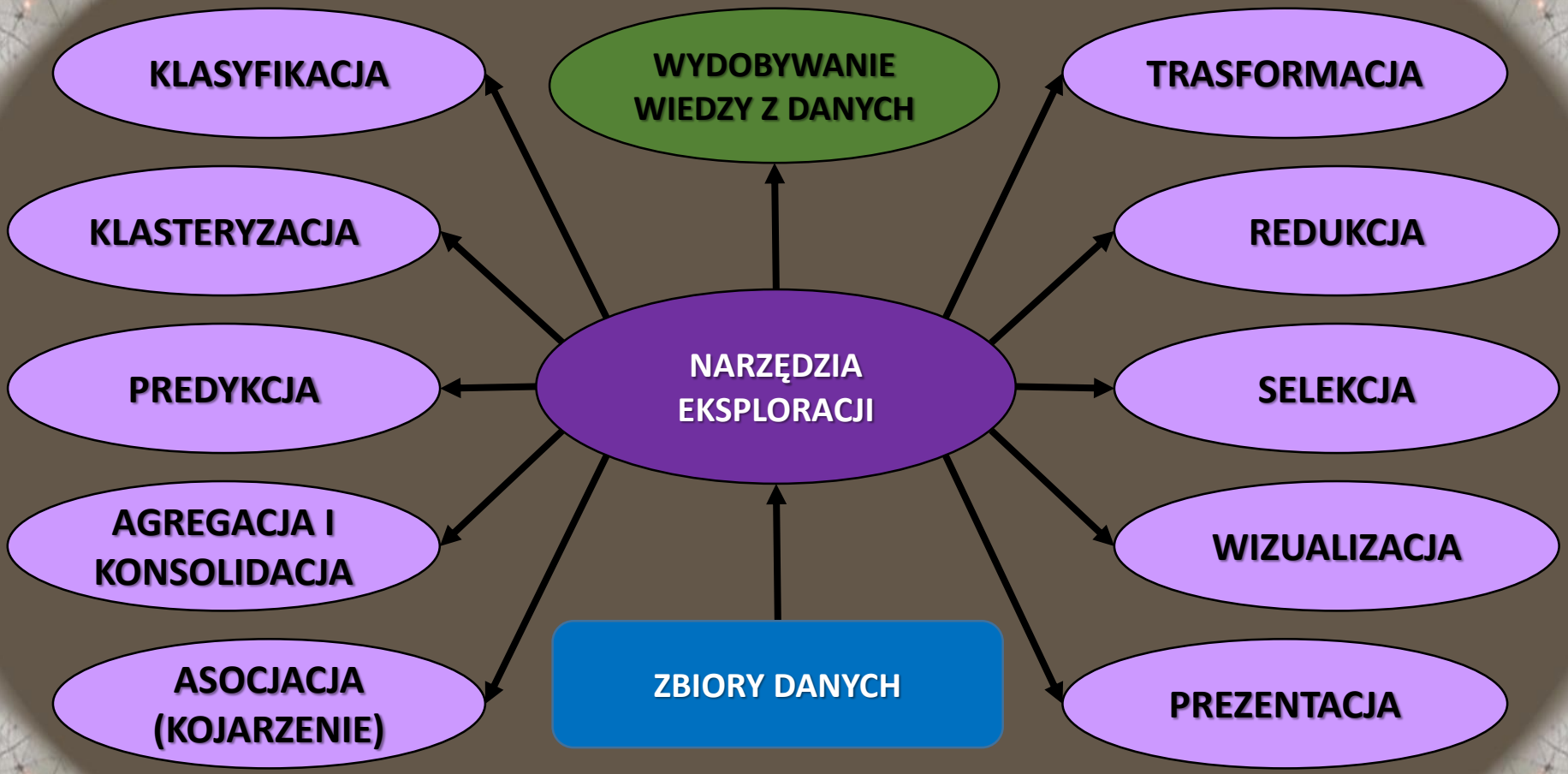
Odkrywanie wiedzy z (baz) danych to proces odkrywania wiedzy ukrytej w danych lub ich zbiorach (czyli bazach danych) polegający na wyszukiwaniu prawidłowości, powtarzalności, podobieństw i zależności (relacji) pomiędzy danymi.

Zadania Eksploracji



Eksploracja danych to zwykle proces wieloetapowy związany z wstępną obróbką danych (czyszczenie, normalizacja, standaryzacja lub inny rodzaj transformacji), porównywaniem, integracją, grupowaniem i selekcją danych oraz wizualizacją danych, ich cech, grup, podobieństw, różnic i zależności (relacji).

Narzędzia Eksploracji



Narzędzia eksploracji danych to metody i algorytmy informatyczne służące do realizacji celów eksploracji wiedzy z danych.

Klasyfikacja

Klasyfikacja to zadanie przyporządkowania **obiektu** do **pewnej klasy** na podstawie podobieństwa, czyli rozpoznawania **obiektu** jako elementu **pewnej klasy**. W wyniku klasyfikacji wzorcowi zostaje przyporządkowana pewna klasa, reprezentowana zwykle przez pewną **etykietę klasy**.

Klasa – to pewna **grupa wzorców** charakteryzujących się podobnymi cechami/właściwościami dla określających je atrybutów/parametrów.

Jeśli **wzorzec** należy równocześnie do kilku klas, wtedy mówimy o zagadnieniu **multiklasyfikacji** (*multiclass classification*), np.:

ser Mozzarella należy do klas: serów, nabiału, produktów spożywczych.

Sklasyfikowanie wzorca jako przynależnego do określonej klasy może być rozważane jako proces:

- rozmyty / predyktywny / ciągły: o określonym stopniu przynależności do klasy
- binarny / zero-jedynkowy / dyskretny: należy lub nie należy do klasy

Klasyfikacja to bardzo ważny proces,
bez którego trudne byłoby formowanie wiedzy,
jak również inteligentne działanie!

Klasteryzacja

Klasteryzacja to proces grupowania obiektów (wzorców) na podstawie ich podobieństwa w taki sposób:

- iż wzorce należące do różnych klastrów są rozłączne (klasteryzacja silna),
- iż wzorce mogą równocześnie należeć do kilku klastrów (klasteryzacja słaba).

Klaster to grupa obiektów podobnych, czyli takich, które są bliskie siebie w pewnej przestrzeni w porównaniu do obiektów innych klastrów, do których są względnie dalekie (niepodobne).

Do najpopularniejszych **metod klasteryzacji** należą:

- Algorytm k-średnich (*k-means clustering*)
- Klasteryzacja hierarchiczna (*hierarchical clustering*)
- Klasteryzacja spektralna (spectra clustering)

Model

Model – to zwykle pewien algorytm lub wzór matematyczny połączony z pewną strukturą lub sposobem reprezentacji przetworzonych danych źródłowych, określany w trakcie procesu uczenia, adaptacji lub konstrukcji.

Obserwacja – to zestaw pomiarów tworzących jeden rekord danych (krotkę).

Predykcja – to wynik procesu regresji lub kojarzenia, w którym otrzymujemy odpowiedź w postaci liczbowej lub innego obiektu.

Redukcja – to proces kompresji stratnej polegający na zmniejszeniu wymiaru wektorów lub macierzy obserwacji poprzez eliminację mało reprezentatywnych lub niekompletnych atrybutów albo w wyniku określania pochodnych reprezentatywnych cech (np. PCA, ICA).

Uczenie

Adaptacja – to polegający na przedstawieniu danych uczących oraz dobraniu, dopasowaniu lub obliczeniu wartości modelu tak, aby dostosował swoje działanie do określonego zbioru, typu i ew. pożądanych wartości wyjściowych danych uczących.

Uczenie – to proces iteracyjny polegający na wielokrotnym przedstawianiu danych uczących oraz poprawianiu wartości modelu tak, aby dostosował swoje działanie do określonego zbioru, typu i ew. pożądanych wartości wyjściowych danych uczących.

Uczenie może być:

- nienadzorowane (bez nauczyciela, *unsupervised*),
- nadzorowane (z nauczycielem, *supervised*),
- konkurencyjne (*competitive*),
- przez wzmacnianie (*reinforcement*),
- motywowane (*motivated*),
- Hebbowskie,
- Bayesowskie (*Bayes*),
- skojarzeniowe (*associative*).

Testowanie

Testowanie – to proces sprawdzania jakości modelu przeprowadzanym w trakcie procesu uczenia lub adaptacji modelu:

- na zbiorze danych chwilowo wydzielonych i wykluczonych z procesu uczenia (tzw. **walidacja** np. krzyżowa – *n-fold cross validation*) lub
- na zbiorze danych testowych całkowicie wykluczonych z procesu uczenia/adaptacji modelu (testowanie właściwe).

Wzorzec – to zestaw lub sekwencja albo inna struktura danych reprezentowanych w postaci zbioru, wektora, macierzy, sekwencji albo grafu danych stosowana do budowy, adaptacji, uczenia, walidacji i testowania modelu.

Wzorce stosowane w trakcie:

- uczenia nazywamy **wzorcami uczącymi**;
- walidacji nazywamy **wzorcami walidacyjnymi**;
- testowania nazywamy **wzorcami testującymi**.

Etapy Eksploracji

1. Zrozumienie zadania i zdefiniowanie celu praktycznego eksploracji, czyli przyporządkowanie zadania do grupy: klasyfikacji, grupowania, predykcji lub asocjacji.
2. Przygotowanie bazy danych do analizy poprzez wyselekcjonowanie rekordów z baz danych najlepiej charakteryzujących rozważany problem.
3. Czyszczenie i wstępna transformacja danych poprzez ich normalizację, standaryzację, usuwanie danych odstających, usuwanie lub uzupełnianie niekompletnych wzorców.
4. Transformacja danych z postaci symbolicznej na postać numeryczną poprzez przypisanie im wartości lub rozmywanie (*fuzzification*) w zależności od stosowanej metody ich dalszego przetwarzania.
5. Redukcja wymiaru danych i selekcja najbardziej znaczących i dyskryminujących cech pozwalających uzyskać najlepsze zdolności uogólniające projektowanego systemu.
6. Wybór techniki i metody eksploracji danych na podstawie możliwości danej metody oraz rodzaju i liczności danych: numeryczne, symboliczne, sekwencyjne...
7. Wybór algorytmu lub aplikacji implementującej wybraną technikę eksploracji danych oraz określenie optymalnych parametrów adaptacji/uczenia wybranej metody (przydatne mogą tutaj być metody ewolucyjne, genetyczne, walidacja krzyżowa).
8. Przeprowadzenie procesu konstrukcji, adaptacji lub uczenia wybraną metodą.
9. Eksploatacja systemu: wnioskowanie, określanie grup, podobieństw, różnic, zależności, następstwa lub implikacji.
10. Douczanie systemu na nowych danych lub utrwalanie zebranych wniosków z eksploracji.

Atrybuty i Cechy

Atrybut – to jedna z cech (parametrów) opisujących obiekt za pośrednictwem wartości reprezentujących ten atrybut. Wartości te są określonego typu i mogą posiadać wartości z pewnego zakresu lub zbioru.

Cecha diagnostyczna – deskryptor numeryczny charakteryzujący i opisujący analizowany proces, zwany również atrybutem procesu.

Ekstrakcja cech diagnostycznych – to proces tworzenia atrybutów wejściowych dla modelu eksploracji na podstawie wyników pomiarowych. Proces ten nazywany jest również **generacją cech**.
Proces ten może być powiązany z **normalizacją**, **standaryzacją** lub inną transformacją danych, mających na celu uwydatnienie głównych cech modelowanego procesu, które mają istotny wpływ na budowę modelu oraz uzyskiwane wyniki i uogólnienie.

Normalizacja

Normalizacja – to przeskalowanie danych względem wielkości skrajnych (min i max) danego wektora danych najczęściej do zakresu [0, 1] (czasami do [-1, 1]) zgodnie z następującą zależnością:

$$y_i = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

$x = [x_1, x_2, \ldots, x_N]$ – to N-elementowy wektor danych źródłowych,

$y = [y_1, y_2, \ldots, y_N]$ – to N-elementowy wektor danych po normalizacji.

Normalizacja jest wrażliwa na wartości odstające i o dużym rozrzucie, gdyż wtedy właściwe dane zostaną ściśnięte w wąskim przedziale, co może znacząco utrudnić ich dyskryminację!

Przeprowadzenie normalizacji jest czasami niezbędne do zastosowania metody, która wymaga, aby dane wejściowe lub wyjściowe mieściły się w pewnym zakresie, np. stosując funkcje sigmoidalną lub tangens hiperboliczny.

Standaryzacja

Standaryzacja – to powszechnie stosowana w statystyce operacja polegająca na przeskalowaniu danych każdego elementu zbioru względem wartości średniej oraz odchylenia standardowego zgodnie z wzorem:

$$y_i = \frac{x_i - m}{\sigma}$$

$x = [x_1, x_2, \ldots, x_N]$ – to N-elementowy wektor danych źródłowych,

$y = [y_1, y_2, \ldots, y_N]$ – to N-elementowy wektor danych po standaryzacji.

m – to wartość średnia wyznaczona z tych danych,

σ – to odchylenie standardowe.

W wyniku standaryzacji otrzymujemy wektor cech, którego wartość średnia jest zerowa, natomiast odchylenie standardowe jest równe jedności.

Nie należy stosować dla danych o odchyleniu standardowym bliskim zeru!

Asocjacje

Asocjacja / Skojarzenie (*associations*) – to proces stowarzyszenia ze sobą dwu lub więcej obserwacji (danych, obiektów, wzorców, encji).

W najprostszej postaci opisywana jest często przez **reguły asocjacyjne**.

Asocjacje są również postawą działania ludzkiego mózgu, pamięci i inteligencji, więc mogą być reprezentowane przez skomplikowane sieci neuronowe.

Uogólnienie / Generalizacja (*generalization*) – to zdolność lub właściwość modelu eksploracji danych polegająca na możliwości poprawnego działania (np. przewidywania, klasyfikacji, regresji) modelu na innych danych niż dane uczące.

Transformacje

Metody redukcji i transformacji danych – mają za zadanie doprowadzić do optymalnej reprezentacji dużych ilości danych, tj. takiej ich reprezentacji, żeby dane w dalszym ciągu były reprezentatywne dla rozważanego problemu, np. klasyfikacji, czyli umożliwiały poprawną dyskryminację wzorców, tj. rozróżnienie ich według pozostałych po redukcji danych.

- **Optymalna reprezentacja** danych może być osiągnięta na skutek:
- **Redukcji wymiaru danych** – czyli **usuwania** mniej istotnych atrybutów danych, oraz **selekcji** atrybutów najistotniejszych pod kątem rozwiązywanego zadania.
- **Transformacji danych** – czyli przekształcenia danych do innej, bardziej oszczędnej lub mniej wymiarowej postaci, która dalej pozwala na ich poprawne rozróżnianie i przetwarzanie, np.:
  - metoda analizy głównych składowych (PCA – Principal Component Analysis),
  - metoda analizy składowych niezależnych (ICA – Independent Component Analysis).
- **Agregacji i Asocjacji danych (*Aggregate & Associate*)** – czyli takiej reprezentacji danych, która polega na zagregowaniu reprezentacji takich samych i/lub podobnych danych i ich grup oraz ich odpowiednim do rozwiązywanego zadania powiązaniu w celu przyspieszenia ich przeszukiwania i przetwarzania.

Wizualizacja

Wizualizacja i prezentacja to zadania związane z graficzną reprezentacją danych w takiej postaci, żeby zaprezentować dane w taki sposób, aby możliwe było:

- porównanie liczności danych określonego typu/grupy/zbioru/klasy,
- wskazanie zależności (relacji) pomiędzy danymi i ich grupami,
- wskazanie minimów, maksimów, średnich, odchyleń i wariancji danych,
- wskazanie rozkładów, agregacji, środków ciężkości,
- wskazanie podobieństw i różnic pomiędzy danymi i ich grupami,
- wskazanie reprezentantów, typowych i nietypowych danych,
- wskazanie wzorców lub wartości odstających od przeciętnych (*outlier*), błędnych, brakujących lub szczególnych,
- podział, odfiltrowanie lub selekcja pewnej grupy wzorców,
- oceny pokrycia przestrzeni danych i ich reprezentatywności dla zadania,
- oceny jakości, zaszumienia, poprawności, dokładności i pełności danych.

Inteligencja

Inteligencja to mentalna zdolność postrzegania informacji i wykorzystywania jej do formowania wiedzy, w celu jej zastosowania do adaptacji do środowiska, rozwiązywanego problemu lub efektywnego osiągania celów.

Inteligencja to mentalna zdolność rozumowania, planowania, rozwiązywania problemów, abstrakcyjnego myślenia, rozumienia złożonych idei, szybkiego uczenia się i efektywnego wykorzystywania zasobów.

Inteligencja obejmuje procesy uczenia się, rozpoznawania, klasyfikacji, rozumienia, logiki, planowania, kreatywności, rozwiązywania problemów i samoświadomości.

Mądrość

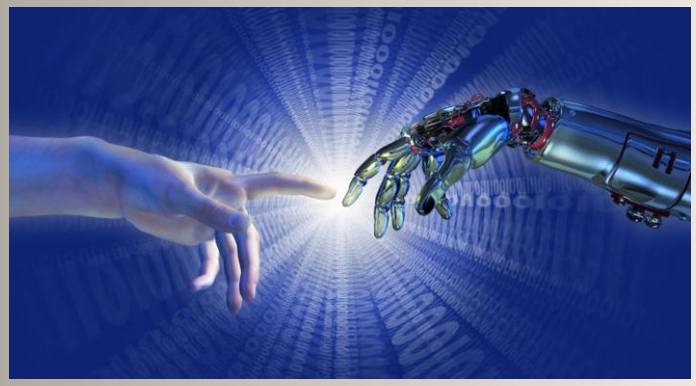
Mądrość to umiejętność wyboru najlepszego, rozsądnego, wydajnego i najbardziej dochodowego sposobu osiągnięcia pożądanego rezultatu w oparciu o wiedzę, potrzeby, inteligencję i etyczne priorytety.

Mądrość pozwala na dobrą ocenę oraz wysoką jakość bycia.

Mądrość jest zwykle wynikiem wcześniejszych prób osiągnięcia pomyślnego wyniku na postawie posiadanego doświadczenia, wiedzy i inteligencji.

Mądrość jest więc traktowana jako przejaw wysokiej inteligencji oraz posiadanej szerokiej wiedzy.

Zbudujmy systemy oparte na wiedzy!



- ✓ Pytania?
- ✓ Uwagi?
- ✓ Sugestie?
- ✓ Życzenia?

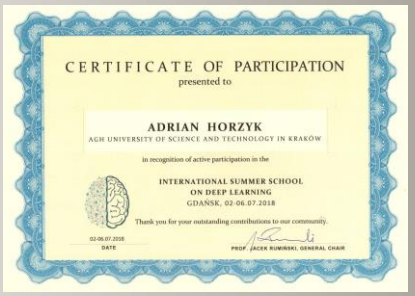
Bibliografia i Literatura

1. Stanisław Osowski, *Metody i narzędzia eksploracji danych*, BTC, Legionowo, 2013.
2. Andrzej Łachwa, *Rozmyty świat zbiorów*, liczb, relacji, faktów, reguł i decyzji, Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2001.
3. Leszek Rutkowski, *Metody i techniki sztucznej inteligencji*, PWN, Warszawa, 2012.
4. D. T. Larose, *Odkrywanie wiedzy z danych*. Wprowadzenie do eksploracji danych, PWN, Warszawa 2006.
5. Stanisław Osowski, *Sieci neuronowe do przetwarzania informacji*, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa, 2013.
6. Nikola K. Kasabov, *Time-Space, Spiking Neural Networks and Brain-Inspired Artificial Intelligence*, In Springer Series on Bio- and Neurosystems, Vol 7., Springer, 2019.
7. Ian Goodfellow, Yoshua Bengio, Aaron Courville, *Deep Learning*, MIT Press, 2016, ISBN 978-1-59327-741-3 or PWN 2018.
8. Holk Cruse, *Neural Networks as Cybernetic Systems*, 2nd and revised edition
9. R. Rojas, *Neural Networks*, Springer-Verlag, Berlin, 1996.
10. *Convolutional Neural Network* (Stanford)
11. *Visualizing and Understanding Convolutional Networks*, Zeiler, Fergus, ECCV 2014
12. IBM: https://www.ibm.com/developerworks/library/ba-data-becomes-knowledge-1/index.html
13. NVIDIA: https://developer.nvidia.com/discover/convolutional-neural-network
14. Horzyk, A., *How Does Generalization and Creativity Come into Being in Neural Associative Systems and How Does It Form Human-Like Knowledge?*, Elsevier, Neurocomputing, Vol. 144, 2014, pp. 238 - 257, DOI: 10.1016/j.neucom.2014.04.046.
15. A. Horzyk, J. A. Starzyk, J. Graham, *Integration of Semantic and Episodic Memories*, IEEE Transactions on Neural Networks and Learning Systems, Vol. 28, Issue 12, Dec. 2017, pp. 3084 - 3095, 2017, DOI: 10.1109/TNNLS.2017.2728203.
16. A. Horzyk, J.A. Starzyk, *Multi-Class and Multi-Label Classification Using Associative Pulsing Neural Networks*, IEEE Xplore, In: 2018 IEEE World Congress on Computational Intelligence (WCCI IJCNN 2018), 2018, (in print).
17. A. Horzyk, J.A. Starzyk, *Fast Neural Network Adaptation with Associative Pulsing Neurons*, IEEE Xplore, In: 2017 IEEE Symposium Series on Computational Intelligence, pp. 339 -346, 2017, DOI: 10.1109/SSCI.2017.8285369.
18. A. Horzyk, K. Gołdon, *Associative Graph Data Structures Used for Acceleration of K Nearest Neighbor Classifiers*, LNCS, In: 27th International Conference on Artificial Neural Networks (ICANN 2018), 2018, (in print).
19. A. Horzyk, *Deep Associative Semantic Neural Graphs for Knowledge Representation and Fast Data Exploration*, Proc. of KEOD 2017, SCITEPRESS Digital Library, pp. 67 - 79, 2017, DOI: 10.13140/RG.2.2.30881.92005.
20. A. Horzyk, *Neurons Can Sort Data Efficiently*, Proc. of ICAISC 2017, Springer-Verlag, LNAI, 2017, pp. 64 - 74, ICAISC BEST PAPER AWARD 2017 sponsored by Springer.

Adrian Horzyk

horzyk@agh.edu.pl

Google: Horzyk



Akademia Górniczo-Hutnicza im. St. Staszica w Krakowie