

Eksploracja danych – Przykładowe pytania na kolokwium i/lub egzamin

Nazwisko i imię Nr albumu

ZADANIA	ODP.
Charakterystyczny efekt końcowy procesu opracowania aplikacji analitycznej w technologii OLAP stanowi:	
a. Specyfikacja wymagań b. Kostka danych c. Diagram UML d. Diagram DFD	
Częstość pojawiania się konkretnego zbioru pozycji w zbiorze transakcji jest mierzona za pomocą:	
a. Zaufania b. Wsparcia c. Hierarchii wymiaru d. Tabeli faktów	
Do której z wymienionych technik analizy danych odnosi się metoda <i>Apriori</i>?	
a. Przetwarzanie OLAP b. Generowanie drzew decyzyjnych c. Analiza asocjacji d. Segmentacja (clustering)	
Jeśli analizie poddawany jest pewien, ograniczony tematycznie, fragment funkcjonowania firmy, wówczas zamiast globalnej hurtowni zwykle wykorzystuje się dane pochodzące:	
a. wyłącznie z własnych zasobów b. wyłącznie ze źródeł zewnętrznych c. wyłącznie z plików tekstowych d. ze składnicy danych (<i>data mart</i>)	
Która z podanych charakterystyk NIE JEST prawdziwa dla grupowania?	
a. Dąży się do tego, aby obiekty wewnątrz grupy (klastra) były maksymalnie do siebie podobne b. Dąży się do tego, aby obiekty z różnych grup (klastrow) były jak najmniej do siebie podobne c. Używa się danych trenujących dla których wcześniej określono etykiety klas d. Żaden podział na klasy nie jest z góry zadany	
Które z podanych pojęć bezpośrednio odnoszą się do wymiaru w modelu wielowymiarowym:	
a. Spójność i wrażliwość b. Hierarchia i uczestnik c. Inżynieria prosta i inżynieria odwrotna d. Wsparcie i zaufanie	
Które z poniższych zdań NIE JEST prawdziwe dla hurtowni danych:	
a. Znalazły zastosowanie w systemach wspomagania decyzji b. Ich zawartość może być dowolnie modyfikowana przez każdego użytkownika c. Zawierają zwykle duże ilości danych, które mogą pochodzić z różnych źródeł d. Mogą być zbudowane w oparciu o model gwiazdy	
Które ze zdań NIE JEST prawdziwe w stosunku do operacyjnych (tradycyjnych) baz danych ?	
a. Mogą zawierać elementy obiektowości b. Udostępniają dane tylko do odczytu c. We współczesnych wersjach współpracują efektywnie z danymi XML d. Wspierają codzienną (operacyjną) działalność firmy lub instytucji	
Które ze zdań POPRAWNIE określa ważną cechę eksploracji danych?	
a. Proces eksploracji jest całkowicie sterowany przez analityka b. Można analizować tylko dane pochodzące z baz relacyjnych c. Eksploracja koncentruje się na samych danych, a nie na udzielaniu odpowiedzi na ściśle określone pytania d. Można analizować tylko dane pochodzące z hurtowni danych	
Można wyodrębnić następujące kategorie technik analizy danych:	
a. Współczesna analiza strukturalna, inżynieria informacji oraz analiza obiektowa b. Analiza trójwymiarowa, analiza funkcjonalna oraz prototypowanie c. Ekstrakcja, transformacja oraz ładowanie d. Zapytania/raporty, analiza wielowymiarowa OLAP oraz eksploracja danych	
Operacja przechodzenia na wyższy poziom w ramach pewnego wymiaru przy analizie wielowymiarowej to:	
a. <i>data mining</i> b. <i>drill-down</i> c. <i>OLAP</i> d. <i>roll-up</i>	
Parametry pozwalające ograniczyć zbiór wygenerowanych reguł asocjacyjnych do tzw. reguł mocnych (interesujących) to:	
a. Średnie wsparcie i średnie zaufanie b. Minimalne wsparcie i minimalne zaufanie c. Wsparcie i zaufanie d. Maksymalne wsparcie i maksymalne zaufanie	
Podstawowe etapy odkrywania wiedzy z danych stanowią:	
a. Wstępna obróbka danych – analiza – ocena i interpretacja wyników b. Ekstrakcja – transformacja – ładowanie c. Analiza – projektowanie – implementacja d. Czyszczenie – integracja – transformacja – redukcja	
W drzewie decyzyjnym każdy węzeł nie będący liściem reprezentuje:	
a. operację skalowania danych na przedział [0;1] b. określoną operację wstępnej obróbki danych b. docelową klasę analizowanego obiektu d. test wartości pewnego atrybutu	