

Tytuł Integracja i eksploracja danych	Kod 1010332421010330687
Kierunek Informatyka	Rok / Semestr 1 / 2
Specjalność -	Przedmiot obowiązkowy
Godziny Wykłady: 2 Ćwiczenia: - Laboratoria: 1 Projekty / seminaria: 1	Liczba punktów 6
	Język prowadzenia przedmiotu polski

Prowadzący:

dr hab. Tadeusz Pankowski
Instytut Automatyki i Inżynierii Informatycznej
ul. Piotrowo 3A
60-965 Poznań
tel. +48 (61) 665 3592
e-mail: Tadeusz.Pankowski@put.poznan.pl

Wydział:

Wydział Elektryczny
ul. Piotrowo 3A
60-965 Poznań
tel. (061) 665-2539, fax. (061) 665-2548
e-mail: office_deef@put.poznan.pl

Miejsce przedmiotu w programie studiów:

Przedmiot obowiązkowy na kierunku Informatyka, studia magisterskie, na Wydziale Elektrycznym.

Założenia i cele przedmiotu:

Poznanie technologii hurtowni danych oraz eksploracji danych. Opanowanie podstaw inegracji danych w środowisku hurtowni danych i metod korzystania z tych danych w przetwarzaniu analitycznym. Poznanie podstawowych metod i algorytmów analizy i eksploracji danych. Poznanie wybranych narzędzi do budowy hurtowni i eksploracji danych. Opanowanie wybranych metod analizy statystycznej.

Treści programowe przedmiotu (opis przedmiotu):

Systemy transakcyjnego (OLTP) a systemy analitycznego (OLAP) przetwarzania danych. Zadania hurtowni danych. Modele danych dla hurtowni danych: schemat gwiazdy, płątka śniegu i znormalizowany. Wielowymiarowy model danych: kostki, wymiary, pomiary, funkcje agregujące. Proces ETL integracji danych w hurtowni danych. Język MDX: wyrażenia i zastosowania. Miejsce i znaczenie eksploracji danych (ang. data mining). Charakterystyka zadań, metod i algorytmów. Drzewa decyzyjne: tworzenie i zastosowanie w problemach klasyfikacji i predykcji. Klasyfikacja metodą Bayesa. Eksploracja danych tekstowych. Odkrywanie reguł asocjacyjnych. Elementy analizy statystycznej. Wykorzystanie systemów SQL Server Analysis Services i SQL Server Business Intelligence. Język DMX i jego wykorzystanie.

Laboratoria

1. Wprowadzenie do hurtowni danych - możliwości SSIS.
2. Eksploracja danych za pomocą SSAS - wstęp do języka MDX.
3. OLAP tworzenie baz analitycznych, tworzenie źródła danych, wymiarów, hierarchii, agregacji oraz wartości wyliczeniowych.
4. Eksploracja danych na przykładach algorytmu drzew decyzyjnych oraz reguł asocjacyjnych.
5. Eksploracja danych na przykładach algorytmów cd. - naiwnego klasyfikatora Bayesa oraz wstęp do eksploracji tekstu (text mining).
6. Techniki eksploracji danych. Tworzenie projektu eksploracji danych korzystając z języka DMX (struktury, modele, ocena).
7. Reprezentacja wiedzy na pomocą SSRS, Excel 2007 jako klient baz analitycznych.

Projekt

Wybrane metody statystycznej analizy danych z wykorzystaniem systemu Statistica i Excel. Statystyka opisowa: miary położenia, rozproszenia, koncentracji i asymetrii. Techniki wnioskowania statystycznego: testy t-Studenta i inne testy istotności. Analiza wariacji. Metody analizy korelacji i regresji. Opracowywanie samodzielnych przykładów z zastosowaniem omawianych technik.

Przedmioty wprowadzające i wymagane wiadomości wstępne:

Wiedza z zakresu baz danych, podstaw teorii informacji, rachunku prawdopodobieństwa i statystyki. Język SQL i programowanie w środowisku .NET.

Forma zajęć i metody dydaktyczne:

Wykład, ćwiczenia laboratoryjne i projektowe (środowisko MS SQL Server). Przygotowanie projektów z zastosowaniem analizy statystycznej.

Forma i warunki zaliczenia przedmiotu – wymagania i system oceniania:

Zaliczenie zajęć laboratoryjnych i projektowych na podstawie umiejętności formułowania zadań i ich rozwiązywania oraz opanowania wybranych środowisk przeznaczonych do budowy hurtowni danych, a także do integracji, eksploracji danych i statystycznej analizy danych.

Bibliografia podstawowa:

-

Bibliografia uzupełniająca:

-