

KARTA OPISU MODUŁU KSZTAŁCENIA		
Nazwa modułu/przedmiotu Bazy i hurtownie danych		Kod 1010515311010510637
Kierunek studiów Informatyka	Profil kształcenia (ogólnoakademicki, praktyczny) ogólnoakademicki	Rok / Semestr 1 / 1
Ścieżka obieralności/specjalność Technologie wytwarzania oprogramowania	Przedmiot oferowany w języku: polski	Kurs (obligatoryjny/obieralny) obligatoryjny
Stopień studiów: II stopień	Forma studiów (stacjonarna/niestacjonarna) niestacjonarna	
Godziny Wykłady: 16 Ćwiczenia: - Laboratoria: 16 Projekty/seminaria: -		Liczba punktów 4
Status przedmiotu w programie studiów (podstawowy, kierunkowy, inny) kierunkowy		(ogólnouczelniany, z innego kierunku) z danego kierunku
Obszar(y) kształcenia i dziedzina(y) nauki i sztuki nauki techniczne		Podział ECTS (liczba i %) 4 100%
Odpowiedzialny za przedmiot / wykładowca:		
<p>dr inż. Krzysztof Dembczyński email: krzysztof.dembczynski@put.poznan.pl tel. 61 6652936 Instytut Informatyki ul. Piotrowo 2, 60-965 Poznań</p>		
Wymagania wstępne w zakresie wiedzy, umiejętności, kompetencji społecznych:		
1	Wiedza:	Efekty kształcenia ze studiów I stopnia zdefiniowane w Uchwale Senatu PP, a szczególnie efekty K_W1-2, K_W4, K_W6-15, weryfikowane w procesie rekrutacji na studia 2 stopnia ? efekty te prezentowane są w serwisie internetowym wydziału www.fc.put.poznan.pl
2	Umiejętności:	Efekty kształcenia ze studiów I stopnia zdefiniowane w Uchwale Senatu PP, a szczególnie efekty K_U1-2, K_U4, K_U7-8, K_U14-20, K_U22-23, K_U26, weryfikowane w procesie rekrutacji na studia 2 stopnia ? efekty te prezentowane są w serwisie internetowym wydziału www.fc.put.poznan.pl
3	Kompetencje społeczne	Efekty kształcenia ze studiów I stopnia zdefiniowane w Uchwale Senatu PP, a szczególnie efekty K_K1-9, weryfikowane w procesie rekrutacji na studia 2 stopnia ? efekty te prezentowane są w serwisie internetowym wydziału www.fc.put.poznan.pl Ponadto w zakresie kompetencji społecznych student musi prezentować takie postawy jak uczciwość, odpowiedzialność, wytrwałość, ciekawość poznawcza, kreatywność, kultura osobista, szacunek dla innych ludzi.
Cel przedmiotu:		
<p>Przekazanie studentom podstawowej wiedzy w zakresie przetwarzania masywnych danych (bardzo dużych zbiorów danych), a dokładniej podstawowych metod organizacji, dostępu i przetwarzania danych, oraz efektywnych algorytmów związanych z prostą analizą masywnych danych.</p> <p>Rozwijanie u studentów umiejętności rozwiązywania problemów dotyczących zarządzania, dostępu, przetwarzania oraz podstawowej analizy masywnych danych.</p>		
Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia		
Wiedza:		
<p>1. Ma uporządkowaną, podbudowaną teoretycznie wiedzę ogólną w zakresie przetwarzania masywnych danych, a dokładniej podstawowych metod organizacji, dostępu i przetwarzania danych, oraz efektywnych algorytmów związanych z prostą analizą masywnych danych. - [K_W4]</p> <p>2. Ma podbudowaną teoretycznie szczegółową wiedzę związaną z wybranymi zagadnieniami, takimi jak: efektywna organizacja i modelowanie masywnych danych w modelu relacyjnym (schemat gwiazdowy) oraz w modelu wielowymiarowym, języki przetwarzania masywnych zbiorów danych (rozszerzenia języka SQL, język MDX), - [K_W5]</p> <p>3. Ma podbudowaną teoretycznie szczegółową wiedzę związaną z wybranymi zagadnieniami, takimi jak: podstawowe struktury danych w przetwarzaniu masywnych danych (funkcje i tabele mieszające, spójne haszowanie, filtry Blooma, indeksy), zaawansowane algorytmy łączenia danych, algorytmy przybliżania wyników zapytań. - [K_W5]</p> <p>4. Ma wiedzę o trendach rozwojowych i najistotniejszych nowych osiągnięciach w informatyce związanych z eksplozją danych i technologiami przetwarzania masywnych zbiorów danych. - [K_W6]</p> <p>5. Ma podstawową wiedzę o cyklu życia systemów przetwarzania masywnych danych. - [K_W7]</p> <p>6. Zna podstawowe metody, techniki i narzędzia stosowane w przetwarzaniu masywnych danych. - [K_W8]</p>		
Umiejętności:		

1. Potrafi pozyskiwać informacje z literatury (w języku ojczystym i angielskim), integrować je, dokonywać ich interpretacji i krytycznej oceny, wyciągać wnioski oraz formułować i wyczerpująco uzasadniać opinie. - [K_U1]
2. Potrafi określić kierunki dalszego uczenia się i zrealizować proces samokształcenia. - [K_U5]
3. Potrafi planować i przeprowadzać eksperymenty, w tym pomiary i symulacje komputerowe, interpretować uzyskane wyniki i wyciągać wnioski. - [K_U8]
4. Potrafi wykorzystać do formułowania i rozwiązywania zadań inżynierskich i prostych problemów badawczych metody analityczne oraz eksperymentalne. - [K_U9]
5. Potrafi - przy formułowaniu i rozwiązywaniu zadań inżynierskich - integrować wiedzę z różnych obszarów informatyki oraz zastosować podejście systemowe, uwzględniające także aspekty pozatechniczne. - [K_U10]
6. Potrafi formułować i testować hipotezy związane z problemami inżynierskimi i prostymi problemami badawczymi w zakresie przetwarzania masywnych danych. - [K_U12]
7. Potrafi ocenić przydatność i możliwość wykorzystania nowych osiągnięć (metod i narzędzi) oraz nowych produktów informatycznych w zakresie przetwarzania masywnych danych. - [K_U13]
8. Potrafi - stosując m.in. koncepcyjnie nowe metody - rozwiązywać złożone zadania informatyczne, w tym zadania nietypowe oraz zadania zawierające komponent badawczy dotyczące przetwarzania masywnych danych - [K_U25]
9. Potrafi odpowiednio zorganizować masywne zbiory danych i przetwarzać je za pomocą takich technologii jak relacyjne systemy baz danych (oraz język SQL) oraz wielowymiarowe systemy baz danych (oraz język MDX). - [-]

Kompetencje społeczne:

1. Rozumie, że w informatyce, a zwłaszcza w przetwarzaniu masywnych danych, wiedza, technologie i umiejętności bardzo szybko stają się przestarzałe. - [K_K1]
2. Zna możliwości dalszego dokształcania się w zakresie przetwarzania masywnych danych. - [K_K3]
3. Zna przykłady i rozumie przyczyny wadliwie działających systemów informatycznych, które doprowadziły do poważnych strat finansowych, społecznych lub też do poważnej utraty zdrowia, a nawet życia. - [K_K4]
4. Potrafi odpowiednio określić priorytety służące realizacji określonego przez siebie lub innych zadania. - [K_K6]

Sposoby sprawdzenia efektów kształcenia

Efekty kształcenia przedstawione wyżej weryfikowane są w następujący sposób:

Ocena formująca:

- a) w zakresie wykładów:
 - na podstawie odpowiedzi na pytania dotyczące materiału omówionego na wykładach.
- b) w zakresie laboratoriów / ćwiczeń:
 - na podstawie oceny bieżącego postępu realizacji zadań.

Ocena podsumowująca:

- a) w zakresie wykładów weryfikowanie założonych efektów kształcenia realizowane jest przez:
 - ocenę wiedzy i umiejętności wykazanych na egzaminie pisemnym o różnej charakterystyce problemów do rozwiązania: 40% pytań dotyczy podstawowej wiedzy i jest przedstawiona w postaci prostych zadań, 40% pytań stanowią zadania obliczeniowe (lub algorytmiczne), natomiast pozostałe 20% pytań to zadania problemowe o większej złożoności; liczba pytań na egzaminie to ok. 10; wszystkie pytania są podobnie punktowane, łącznie można otrzymać 100 punktów; zaliczenie egzaminu jest od 50 punktów; na ostateczną ocenę składa się w 60% ocena z egzaminu pisemnego i w 40% ocena z laboratorium.
 - omówienie wyników egzaminu,
- b) w zakresie laboratoriów weryfikowanie założonych efektów kształcenia realizowane jest przez:
 - ocenę realizacji zadań związanych z danymi zajęciami laboratoryjnymi: podczas każdego zajęcia laboratoryjnego student otrzymuje listę zadań do wykonania: zadania dzielą się na niepunktowane, obowiązkowe punktowane do realizacji na zajęciach, obowiązkowe punktowane zadania domowe oraz punktowane nieobowiązkowe zadania domowe; możliwe jest uzyskanie dodatkowych punktów za aktywność podczas zajęć.

Treści programowe

Program wykładu obejmuje następujące zagadnienia:

- Przedstawienie problemu eksplozji danych we współczesnym świecie oraz rozróżnienie systemów informatycznych pod względem wykorzystywania danych na systemy operacyjne, w których dane służą do wspomaganie codziennych, urzędniczych czynności, oraz na systemy analityczne, w których stara się wydobyć jak największą wiedzę ze zgromadzonych danych. Omówienie zastosowania metod eksploracji danych oraz wskazanie pułapek związanych z przetwarzaniem dużych zbiorów danych.
 - Przedstawienie historii i ewolucji systemów baz danych oraz dokładne omówienie modeli danych w rozróżnieniu na rodzaje systemów przetwarzania danych. Przypomnienie podstaw modelu relacyjnego, wprowadzenie i dokładne scharakteryzowanie modelu wielowymiarowego będącego podstawą systemów hurtowni danych, oraz modelu nierelacyjnego (NoSQL) związanego z przetwarzaniem masywnych danych w zastosowaniach internetowych.
 - Omówienie systemów hurtowni danych, modelowania wielowymiarowego i schematu gwiazdy, przedstawienie podstaw procesu ekstrakcji, transformacji i ładowania danych (proces ETL) będącego podstawą przenoszenia danych z systemów operacyjnych do systemów analitycznych, oraz wprowadzenie rozróżnienia systemów hurtowni danych na systemy relacyjne i wielowymiarowe.
 - Wprowadzenie do analitycznych zapytań wielowymiarowych i ich specyfiki, tabel i raportów przestawnych, analitycznych rozszerzeń języka SQL oraz języka MDX, który został stworzony specjalnie do definiowania tabel i raportów przestawnych.
 - Kolejna grupa wykładów dotyczy struktur danych i algorytmów przetwarzania masywnych danych. Przypomnienie wiedzy teoretycznej z zakresu funkcji i tabel mieszających, omówienie filtrów Blooma, indeksów stosowanych w przetwarzaniu masywnych danych (indeksy bitmapowe, segmentowe, projekcji i połączeniowe), podstawowe zagadnienia dotyczące partycjonowania danych, przetwarzania zapytań łączenia i grupowania.
- Zajęcia laboratoryjne prowadzone są w formie ośmiu dwugodzinnych ćwiczeń, odbywających się w laboratorium. Ćwiczenia realizowane są indywidualnie, z wyjątkiem niektórych zadań, które mogą być realizowane w zespołach 2-osobowych. Program laboratorium obejmuje następujące zagadnienia:
- Organizacja danych w systemie informatycznym dla przykładowego dużego zbioru danych, np. z dziedziny systemów rekomendacyjnych; wprowadzenie i utrzymywanie danych w systemie relacyjnym baz danych oraz w innych reprezentacjach; ocena możliwości przetwarzania przykładowego zbioru danych. Podstawowe analityczne zapytania w języku SQL.
 - Studium przypadku z modelowania wielowymiarowego dla przykładowego dużego zbioru danych, jak również dla typowych scenariuszy biznesowych. Ocena możliwości realizacji zapytań SQL dla danego modelu.
 - Rozszerzenia języka SQL dla zapytań analitycznych; studium przypadku dla przykładowego dużego zbioru danych oraz dla typowego scenariusza biznesowego.
 - Model wielowymiarowy oraz język MDX; studium przypadku dla przykładowego dużego zbioru danych oraz dla typowego scenariusza biznesowego.

Metody dydaktyczne:

1. wykład: prezentacja multimedialna ilustrowana przykładami podawanymi na tablicy.
2. ćwiczenia laboratoryjne: rozwiązywanie zadań, ćwiczenia praktyczne dotyczące przetwarzania konkretnego dużego zbioru danych połączone z dyskusją, zapisywanie zapytań w języku SQL i MDX.

Literatura podstawowa:

1. Systemy baz danych. Kompletny podręcznik. Wydanie II, Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom
2. Hurtownie danych: logiczne i fizyczne struktury danych, Z. Królikowski, Wydawnictwo Politechniki Poznańskiej 2007

Literatura uzupełniająca:

1. Mining of Massive Datasets, A. Rajaraman, J. D. Ullman, Cambridge University Press, 2012 (podręcznik jest legalnie dostępny w wersji elektronicznej: <http://infolab.stanford.edu/~ullman/mmds.html>)
2. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, R. Kimball, M. Ross, John Wiley & Sons 2002

Bilans nakładu pracy przeciętnego studenta

Czynność	Czas (godz.)
1. Udział w zajęciach wykładowych:	16
2. Udział w zajęciach laboratoryjnych:	16
3. Dokończenie (w ramach pracy własnej) zadań z ćwiczeń laboratoryjnych:	20
4. Udział w konsultacjach (mogą być realizowane drogą elektroniczną) związanych z realizacją procesu kształcenia, w szczególności ćwiczeń laboratoryjnych / projektu	4 20
5. Zapoznanie się ze wskazaną literaturą / materiałami dydaktycznymi (~10 stron tekstu naukowego = 1 godz.), 200 stron.	20
6. Przygotowanie do egzaminu	2
7. Obecność na egzaminie	1
8. Omówienie wyników egzaminu	

Obciążenie pracą studenta		
forma aktywności	godzin	ECTS
Łączny nakład pracy	99	4
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	39	1
Zajęcia o charakterze praktycznym	36	1