

| 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 | | Podział ECTS (liczba i %) |
| 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: | Student rozpoczynający ten przedmiot powinien posiadać podstawową wiedzę z zakresu systemów baz danych, algorytmiki, metod probabilistycznych oraz statystycznej analizy danych. |
| 2 | Umiejętności: | Powinien posiadać umiejętności programistyczne (zwłaszcza w zakresie systemów baz danych), rozwiązywania zadań z algorytmiki, metod probabilistycznych i statystycznej analizy danych. |
| 3 | Kompetencje społeczne | W zakresie kompetencji społecznych student musi prezentować takie postawy jak uczciwość, odpowiedzialność, wytrwałość, ciekawość poznawcza, kreatywność, kultura osobista, szacunek dla innych ludzi. |
| Cel przedmiotu: | | |
| Przekazanie studentom podstawowej wiedzy w zakresie organizacji, zarządzania i przetwarzania masywnych danych (bardzo dużych zbiorów danych). | | |
| Rozwijanie u studentów umiejętności rozwiązywania problemów dotyczących organizacji, zarządzania i przetwarzania masywnych danych. | | |
| Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia | | |
| Wiedza: | | |
| 1. Ma uporządkowaną i podbudowaną teoretycznie wiedzę ogólną związaną z kluczowymi zagadnieniami dotyczącymi przetwarzania masywnych danych. - [K2st_W2] 2. Ma zaawansowaną wiedzę szczegółową dotyczącą wybranych zagadnień z zakresu przetwarzania masywnych danych. - [K2st_W3] 3. Ma wiedzę o trendach rozwojowych i najistotniejszych nowych osiągnięciach dokonanych w przetwarzaniu masywnych danych. - [K2st_W4] 4. Ma zaawansowaną i szczegółową wiedzę o procesach zachodzących w cyklu życia systemów przetwarzania masywnych danych. - [K2st_W5] 5. Zna zaawansowane metody, techniki i narzędzia stosowane przy rozwiązywaniu złożonych zadań w obszarze przetwarzania masywnych danych. - [K2st_W6] | | |
| Umiejętności: | | |

| |
|---|
| <p>1. Potrafi pozyskiwać informacje z literatury, baz danych oraz innych źródeł (w języku polskim i angielskim), integrować je, dokonywać ich interpretacji i krytycznej oceny, wyciągać wnioski oraz formułować i wyczerpująco uzasadniać opinie. - [K2st_U1]</p> <p>2. Potrafi planować i przeprowadzać eksperymenty, w tym pomiary i symulacje komputerowe, interpretować uzyskane wyniki i wyciągać wnioski oraz formułować i weryfikować hipotezy związane ze złożonymi problemami inżynierskimi i prostymi problemami badawczymi. - [K2st_U3]</p> <p>3. Potrafi wykorzystać do formułowania i rozwiązywania zadań inżynierskich i prostych problemów badawczych metody analityczne, symulacyjne oraz eksperymentalne. - [K2st_U4]</p> <p>4. Potrafi - przy formułowaniu i rozwiązywaniu zadań inżynierskich - integrować wiedzę z różnych obszarów informatyki (a w razie potrzeby także wiedzę z innych dyscyplin naukowych) oraz zastosować podejście systemowe, uwzględniające także aspekty pozatechniczne. - [K2st_U5]</p> <p>5. Potrafi ocenić przydatność i możliwość wykorzystania nowych osiągnięć (metod i narzędzi) oraz nowych produktów w obszarze przetwarzania masywnych danych. - [K2st_U6]</p> <p>6. Potrafi - stosując m.in. koncepcyjnie nowe metody - rozwiązywać złożone zadania w obszarze przetwarzania masywnych danych, w tym zadania nietypowe oraz zadania zawierające komponent badawczy. - [K2st_U10]</p> <p>7. Potrafi współdziałać w zespole, przyjmując w nim różne role. - [K2st_U15]</p> <p>8. Potrafi określić kierunki dalszego uczenia się i zrealizować proces samokształcenia, w tym innych osób. - [K2st_U16]</p> |
| Kompetencje społeczne: |
| <p>1. Rozumie, że wiedza i umiejętności dotyczące przetwarzania masywnych danych bardzo szybko stają się przestarzałe. - [K2st_K1]</p> <p>2. Rozumie znaczenie wykorzystywania najnowszej wiedzy z zakresu informatyki w rozwiązywaniu problemów badawczych i praktycznych. - [K2st_K2]</p> |

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 i złożoności problemów do rozwiązania (proste zadania dotyczące wiedzy podstawowej, zadania trudniejsze wymagające obliczeń lub symulacji algorytmów, zadania problemowe o dużej złożoności); łączna liczba pytań na egzaminie to ok. 10; wszystkie pytania są podobnie punktowane, łącznie można otrzymać 100 punktów; zaliczenie egzaminu jest od 50 punktów; ostateczna ocena jest średnią ważoną z egzaminu pisemnego i 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 (składającą się z zadań niepunktowanych, zadań punktowanych oraz zadań domowych); zaliczenie laboratorium jest od 50% zdobytych punktów podczas całego semestru; możliwe jest uzyskanie dodatkowych punktów za aktywność podczas zajęć.

Treści programowe

Program wykładu obejmuje następujące zagadnienia:

- Problem eksplozji danych we współczesnym świecie; rozróżnienie systemów informatycznych pod względem wykorzystywania danych na systemy operacyjne oraz na systemy analityczne; zastosowania metod eksploracji danych oraz pułapki związane z przetwarzaniem masywnych danych.
- Historia i ewolucja systemów baz danych; modele danych w rozróżnieniu na rodzaje systemów przetwarzania danych: model relacyjny, wielowymiarowy i nierelacyjny (NoSQL).
- Systemy hurtowni danych, modelowanie wielowymiarowe i schemat gwiazdy, przedstawienie podstaw procesu ekstrakcji, transformacji i ładowania danych (proces ETL), relacyjne i wielowymiarowe systemy hurtowni danych
- Wprowadzenie do paradygmatu MapReduce na przykładzie oprogramowania Hadoop i Spark, podstawowe algorytmy takie jak zliczanie, operacje algebry relacji (projekcja, selekcja, grupowanie, łączenie), oraz mnożenie macierzy.
- Struktury i algorytmy przetwarzania masywnych danych: funkcje i tabele mieszające, indeksy stosowane w przetwarzaniu masywnych danych, filtry Blooma, podstawowe zagadnienia dotyczące partycjonowania danych, przetwarzanie zapytań.

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 dwuosobowych. Program laboratorium obejmuje następujące zagadnienia:

- Organizacja danych w systemie informatycznym dla przykładowego dużego zbioru danych, np. z dziedziny systemów rekomendacyjnych.
- Modelowania wielowymiarowe.
- Wprowadzenie do MapReduce na przykładzie oprogramowania Hadoop i Spark: przedstawienie podstawowych zagadnień technicznych oraz implementacja prostych algorytmów, takich jak zliczanie, operacje algebry relacji, mnożenie macierzy
- Studium przypadku na podstawie przykładowego dużego zbioru danych.

Metody dydaktyczne:

1. wykład: prezentacja multimedialna ilustrowana przykładami podawanymi na tablicy, dyskusja i analiza problemów.
2. ćwiczenia laboratoryjne: rozwiązywanie zadań, dyskusja, praca w zespole.

Literatura podstawowa:

1. Systemy baz danych. Kompletny podręcznik. Wydanie II, Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom

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
3. Hurtownie danych: logiczne i fizyczne struktury danych, Z. Królikowski, Wydawnictwo Politechniki Poznańskiej 2007

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 |
| 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 |