

KARTA PRZEDMIOTU

1. Informacje ogólne

Nazwa przedmiotu i kod (wg planu studiów):	Hurtownie i eksploracja danych D1_5
Nazwa przedmiotu (j. ang.):	Data Warehouses and Data Mining
Kierunek studiów:	Informatyka
Specjalność/specjalizacja:	Technologie internetowe i bazy danych
Poziom kształcenia:	studia I stopnia
Profil kształcenia:	praktyczny (P)
Forma studiów:	studia stacjonarne / studia niestacjonarne
Obszar kształcenia:	nauki techniczne (wg wykazu)
Dziedzina:	nauki techniczne (wg wykazu)
Dyscyplina nauki:	(wg wykazu)
Koordinator przedmiotu:	dr hab. Jan Bazan

2. Ogólna charakterystyka przedmiotu

Przynależność do modułu:	kształcenie specjalnościowe
Status przedmiotu:	Obowiązkowy
Język wykładowy:	Polski
Rok studiów, semestr:	III, 5,6
Forma i wymiar zajęć według planu studiów:	stacjonarne - wykład 15 h, ćw. laboratoryjne 45 h niestacjonarne – wykład 15 h, ćw. laboratoryjne 30 h
Interesariusze i instytucje partnerskie (nieobowiązkowe)	
Wymagania wstępne / Przedmioty wprowadzające:	Znajomość relacyjnych baz danych oraz języka SQL. Umiejętność projektowania relacyjnych baz danych oraz programowania w języku SQL. Programowanie w języku orientowanym obiektowo Java (poziom podstawowy).

3. Bilans punktów ECTS

Całkowita liczba punktów ECTS: (A + B)		5	Stacjonarne	Niestacjonarne
A. Liczba godzin wymagających bezpośredniego udziału nauczyciela z podziałem na typy zajęć oraz całkowita liczba punktów ECTS osiągniętych na tych zajęciach:	obecność na wykładach		15	15
	obecność na ćwiczeniach laboratoryjnych		45	30
	udział w konsultacjach		5	5
	w sumie: ECTS		65	50
B. Poszczególne typy zadań do samokształcenia studenta (nie-wymagających bezpośredniego udziału nauczyciela) wraz z planowaną średnią liczbą godzin na każde i sumaryczną liczbą ECTS:	przygotowanie do zajęć laboratoryjnych (zad. domowe)		15	15
	przygotowanie do kolokw. zalicz. na papierze		10	10
	przygotowanie do kolokw. zalicz. przy komputerze		15	20
	wykonanie projektu		15	20
	przygotowanie do egzaminu		10	15
w sumie: ECTS		65	75	
C. Liczba godzin praktycznych / laboratoryjnych w ramach przedmiotu oraz związana z tym liczba punktów ECTS:	udział w ćwiczeniach		45	30
	praca praktyczna samodzielna (zadania domowe, przygotowanie do kolokwium przy komputerze i wykonanie projektu)		45	55
	w sumie: ECTS		90	85
			3.0	3.0

4. Opis przedmiotu

<p>Cel przedmiotu: Celem zajęć jest przygotowanie studentów do realizacji projektów inżynierskich związanych z klasycznymi hurtowniami danych i systemami eksploracji danych.</p>
<p>Metody dydaktyczne: wykład, praktyczne ćwiczenia laboratoryjne</p>
<p>Treści kształcenia: Wykłady:</p> <ol style="list-style-type: none"> 1. Hurtownia danych, jej cechy i cele. 2. Dwa podejścia do gromadzenia i przetwarzania danych: OLTP i OLAP. 3. Hurtownie danych w przedsiębiorstwach (motywacja, aktualna sytuacja, korzyści, trendy). 4. Typowe modele architektury hurtowni danych. 5. Trzy poziomy projektowania hurtowni danych. 6. Metody projektowania hurtowni danych. 7. Wielowymiarowy model danych. 8. Trzy typowe schematy do reprezentowania danych w modelach wielowymiarowych. 9. Retrospekcja jako sposób wykonywania zmian w hurtowniach danych. 10. Procesy ETL w hurtowni danych (waga procesu ETL, programy typu wrapper, transformacja i czyszczenie danych, metody ładowania danych, techniki wykrywania zmian w danych, problem aktualizacji danych). 11. Aktualizacja perspektyw i strategię ich odświeżania. 12. Obsługiwalność perspektyw.

13. Agregacja wartości w hurtowniach danych.
14. Nawigacja po agregacjach oraz typowe operacje związane z nawigacją po agregacjach (zwijanie, rozwijanie, selekcja, filtrowanie, zawężanie, obracanie).
15. Modele pamięci w hurtowniach danych.
16. Ogólne etapy realizacji narzędzia BI opartego na hurtowni danych.
17. Realizacja projektu na budowę hurtowni danych (analiza wymagań, projektowanie hurtowni, implementacja hurtowni, testowanie i strojenie hurtowni).
18. Przegląd systemów do tworzenia hurtowni danych.
19. Wprowadzenie do komputerowych metod eksploracji danych (tablice danych, formaty danych, import danych, problem pustych miejsc w tablicach, dyskretyzacja atrybutów, klasyfikacja i klasyfikatory, generowanie decyzji na podstawie klasyfikatora, metody oceny jakości klasyfikatora, podstawowe scenariusze eksperymentów związane z eksploracją danych).
20. Przegląd najbardziej znanych systemów i bibliotek programistycznych do eksploracji danych.
21. Omówienie obsługi systemów RSES i WEKA.
22. Omówienie WEKA API na przykładzie tworzenia konkretnych programów do eksploracji danych.
23. Omówienie metod integracji RDBMS i metod eksploracji danych w celu uzyskania hurtowni danych.
24. Perspektywy rozwoju hurtowni danych.

Ćwiczenia laboratoryjne:

1. Zadania na projektowanie hurtowni danych.
2. Zadania na implementację hurtowni danych z wykorzystaniem języka SQL, w tym zapytań grupujących (PostgreSQL).
3. Zadania implementację hurtowni danych z wykorzystaniem typowych rozszerzeń języka SQL (w typ polecen ROLLUP, CUBE i GROUPING SETS).
4. Zadania realizację rozmaitych nawigacji po agregacjach.
5. Zadanie na tworzenie tablic przestawnych i innych złożonych zapytań analitycznych.
6. Kolokwium na papierze z projektowania i implementowania hurtowni danych.
7. Prezentacja systemu WEKA (Rough Set Exploration System) i wykonywanie zadań eksploracyjnych z użyciem tego systemu.
8. Przykłady użycia WEKA API do czytania danych i wyliczania statystyk dla zbiorów danych. Przykłady użycia WEKA API do budowy klasyfikatorów.
9. Przykłady użycia WEKA API do grupowania obiektów w danych.
10. Przykłady użycia WEKA API do wyliczania reguł asocjacyjnych.
11. Przykłady integracji bazy danych zaimplementowanej w systemie RDBMS z metodami eksploracji danych w hurtownię danych.
12. Kolokwium przy komputerze z wykorzystania WEKA API do eksploracji danych.
13. Przewiedzenie propozycji tematów projektów indywidualnych oraz metodologii realizacji projektu, którego celem jest wykonanie hurtowni danych.
14. Raporty z realizacji etapu projektowania i implementowania wybranej indywidualnie przez studenta hurtowni danych.
15. Zaliczenie projektu.

5. Efekty kształcenia i sposoby weryfikacji

Efekty kształcenia		
Efekt przedmiotu <i>(kod przedmiotu + kod</i>	Student, który zaliczył przedmiot (spełnił minimum wymagań)	Efekt kierunkowy

efektu kształcenia)				
Wiedza:				
D1_5_W01	1. Ma podstawową wiedzę dotyczącą metod eksploracji danych.	K_W06, K_W07, K_W08, K_W16		
D1_5_W02	2. Ma podstawową wiedzę w zakresie projektowania hurtowni danych.	K_W07, K_W08.		
D1_5_W03	3. Ma podstawową wiedzę potrzebną do implementacji hurtowni danych.	K_W06, K_W08, K_W16		
Umiejętności				
D1_5_U01	1. Potrafi stosować wybrane narzędzia i biblioteki informatyczne do eksploracji danych, w tym do realizacji w nich podstawowych zadań eksploracji.	K_U03, K_U07, K_U08, K_U11, K_U19, K_U28, K_U32		
D1_5_U02	2. Potrafi projektować hurtownie danych w oparciu o relacyjną bazę danych.	K_U03, K_U08, K_U12, K_U20		
D1_5_U03	3. Potrafi implementować hurtownie danych za pomocą relacyjnej bazy danych i języka programowania Java.	K_U03, K_U07, K_U08, K_U11, K_U19, K_U20		
Sposoby weryfikacji efektów kształcenia:				
Lp.	Efekt przedmiotu	Sposób weryfikacji	Ocena formująca	Ocena końcowa
1	D1_5_W01	egzamin pisemny	ocena z egzaminu w zakresie tego efektu	średnia z ocen formujących, sprawdzających nabytą wiedzę i umiejętności
2	D1_5_W02	egzamin pisemny	ocena z egzaminu w zakresie tego efektu	
3	D1_5_W03	egzamin pisemny	ocena z egzaminu w zakresie tego efektu	
7	D1_5_U01	kolokwium przy komputerze, projekt	średnia ocen z kolokwium i projektu w zakresie tego efektu	
8	D1_5_U02	kolokwium pisemne, projekt	średnia ocen z kolokwium i projektu w zakresie tego efektu	
9	D1_5_U03	kolokwium pisemne, projekt	średnia ocen z kolokwium i projektu w zakresie tego efektu	
Kryteria oceny				
w zakresie wiedzy			Efekt kształcenia	
Na ocenę 3,0	Student ma podstawową wiedzę dotyczącą metod eksploracji danych, zna wybrane narzędzia informatyczne do eksploracji danych oraz zna techniki realizacji w nich podstawowych zadań eksploracji. W szczególności, zna		D1_5_W01	

	techniki wczytywania danych, ich przygotowania do eksploracji, wyliczania dla nich podstawowych statystyk, tworzenia podstawowych modeli eksploracji oraz podstawowe techniki mierzenia jakości wyznaczonych modeli i formułowania opinii na temat jakości tych metod. Jednak nie zawsze zna podstawy teoretyczne stosowanych modeli eksploracji danych.	
Na ocenę 5,0	Student ma podstawową wiedzę dotyczącą metod eksploracji danych, zna wybrane narzędzia informatyczne do eksploracji danych oraz zna techniki realizacji w nich podstawowych zadań eksploracji. W szczególności, zna techniki wczytywania danych, ich przygotowania do eksploracji, wyliczania dla nich podstawowych statystyk, tworzenia podstawowych modeli eksploracji oraz podstawowe techniki mierzenia jakości wyznaczonych modeli i formułowania opinii na temat jakości tych metod. Ponadto, zna podstawy teoretyczne stosowanych modeli eksploracji danych.	
Na ocenę 3,0	Student ma podstawową wiedzę w zakresie metod projektowania hurtowni danych w oparciu relacyjną bazę danych przedstawianą na wykładzie (patrz treści powyżej). Jednak nie zna lub ma problemy ze zrozumieniem zaawansowanych elementów projektowania hurtowni danych, takich jak retrospekcja danych, obsługiwalność perspektyw, techniki wykrywania zmian danych w hurtowni, nawigacje po agregacjach.	D1_5_W02
Na ocenę 5,0	Student ma wiedzę w zakresie metod projektowania hurtowni danych w oparciu relacyjną bazę danych przedstawianą na wykładzie (patrz treści powyżej). Ponadto, zna i rozumie zaawansowane elementy projektowania hurtowni danych (patrz wyżej).	
Na ocenę 3,0	Student ma podstawową wiedzę w zakresie implementacji hurtowni danych, która jest potrzebna do implementacji hurtowni danych za pomocą relacyjnej bazy danych. Jednak czasem brakuje mu wiedzy do poprawnego tworzenia zaawansowanych zapytań analitycznych.	D1_5_W03
Na ocenę 5,0	Student ma podstawową wiedzę w zakresie implementacji hurtowni danych, która jest potrzebna do implementacji hurtowni danych za pomocą relacyjnej bazy danych, w tym wiedzę na temat tworzenia zaawansowanych zapytań analitycznych.	
w zakresie umiejętności		Efekt kształcenia
Na ocenę 3,0	Student potrafi stosować wybrane narzędzia informatyczne do eksploracji danych, w tym techniki do realizacji w nich podstawowych zadań eksploracji. W szczególności, potrafi stosować techniki wczytywania danych, ich przygotowania do eksploracji, wyliczania dla nich podstawowych statystyk, tworzenia podstawowych modeli eksploracji i wizualizacji podstawowych informacji o ich strukturze oraz potrafi stosować podstawowe techniki mierzenia jakości wyznaczonych modeli. Jednak nie zawsze potrafi poprawnie użyć wyznaczone modele eksploracji danych zarówno w typowej jak i nie typowej sytuacji. Ma także problemy z wyliczaniem bardziej złożonych statystyk dotyczące danych lub wyznaczonych modeli.	D1_5_U01
Na ocenę 5,0	Student potrafi stosować wybrane narzędzia informatyczne do eksploracji danych, w tym techniki do realizacji w nich podstawowych zadań eksploracji. W szczególności, potrafi stosować techniki wczytywania danych, ich przygotowania do eksploracji, wyliczania dla nich podstawowych statystyk, tworzenia podstawowych modeli eksploracji i wizualizacji podstawowych informacji o ich strukturze oraz potrafi stosować podstawowe techniki mierzenia jakości wyznaczonych modeli. Ponadto, potrafi poprawnie użyć wyznaczone modele eksploracji danych w typowej sytuacji oraz wylicza bardziej złożone statystyki dotyczące danych lub wyznaczonych modeli. Wreszcie potrafi poprawnie użyć wyznaczone modele eksploracji danych	

	w nietypowej sytuacji lub w nietypowy sposób, tj. np. wtedy, gdy dane testowe nie są bezpośrednio dostępne w narzędziu i trzeba je zaimportować w nie trywialny sposób.	
Na ocenę 3,0	Student potrafi projektować hurtownie danych w oparciu relacyjną bazę danych. Jednak czasami ma problem z poprawnym zaprojektowaniem bardziej złożonych hurtowni, np. takich, które są projektowane według schematu konstelacyjnego.	D1_5_U02
Na ocenę 5,0	Student potrafi projektować hurtownie danych w oparciu relacyjną bazę danych, w tym poprawnie projektuje bardziej złożone hurtownie danych.	
Na ocenę 3,0	Student potrafi implementować zaprojektowaną hurtownię danych w oparciu o relacyjną bazę danych. Jednak nie zawsze potrafi implementować zaawansowane operacje analityczne (np. złożone zapytania analityczne i trudniejsze nawigacje po agregacjach kostek).	D1_5_U03
Na ocenę 5,0	Student potrafi implementować zaprojektowaną hurtownię danych w oparciu o relacyjną bazę danych, w tym potrafi poprawnie implementować zaawansowane operacje analityczne.	

Student, który osiągnął zakładany poziom wiedzy, posiadał wymagane umiejętności, cechuje się określonymi kompetencjami społecznymi, które są zdefiniowane w efektach kształcenia dla przedmiotu, zalicza przedmiot.

Student, który nie osiągnął zakładanych efektów kształcenia, nie zalicza przedmiotu.

<p>Kryteria oceny końcowej:</p> <p><u>Zaliczenie przedmiotu</u> następuje na podstawie zaliczenia wszystkich efektów weryfikowanych przez planowane w danym okresie metody weryfikacji. Przy czym zakłada się, że każda metoda weryfikacji dostarcza osobne oceny dla każdego z weryfikowanych przez nią efektów kształcenia. Jeśli dany efekt jest weryfikowany przez więcej niż jedną metodę, to ocena weryfikująca osiągnięcie tego efektu jest obliczana jako średnia arytmetyczna ocen uzyskanych w poszczególnych metodach weryfikowania tego efektu.</p> <p>Student otrzymuje ocenę niedostateczny, gdy metody weryfikacji wykażą, iż przynajmniej jeden efekt nie został osiągnięty (średnia ocena dla tego efektu jest niższa niż 3.0);</p> <p>Student otrzymuje ocenę dostateczny, gdy przeciętnie każdy z efektów zostanie osiągnięty na poziomie co najmniej 3.0, ale chociaż jeden z efektów został osiągnięty na poziomie mniejszym od 3.75;</p> <p>Student otrzymuje ocenę dobry, gdy przeciętnie każdy z efektów zostanie osiągnięty na poziomie co najmniej 3.75, ale chociaż jeden z efektów został osiągnięty na poziomie mniejszym od 4.75;</p> <p>Student otrzymuje ocenę bardzo dobry, gdy przeciętnie każdy z efektów zostanie osiągnięty na poziomie co najmniej 4.75;</p>
<p>Zalecana literatura :</p> <ol style="list-style-type: none"> 1. Wykłady: http://fenix.univ.rzeszow.pl/bazan/ 2. Wykłady z przedmiotu Hurtownie danych, na PJWSTK w Warszawie, Autorzy: Jakub Wróblewski, Agnieszka Chądryńska, Maciej Wawrzynek, Michał Wilbrandt oraz Dominik Ślęzak (http://edu.pjwstk.edu.pl/wyklady/hur/scb/) 3. Morzy T., Eksploracja danych; Metody i algorytmy, PWN, 2013 (biblioteka PWSZ). 4. Todman C., Projektowanie hurtowni danych. Helion 2011 (biblioteka PWSZ). 5. Larose D.T., Odkrywanie wiedzy z danych; wprowadzenie do eksploracji danych, PWN, 2006 (dostępna w Internecie w pliku pdf).

- | |
|---|
| 6. Dokumentacja systemu WEKA (http://www.cs.waikato.ac.nz/ml/weka/documentation.html). |
| 7. Dokumentacja systemu PostgreSQL (https://www.postgresql.org/docs/). |
| 8. Witryna internetowa API WEKA (https://weka.wikispaces.com/Use+WEKA+in+your+Java+code) |

Informacje dodatkowe:

Dodatkowe obowiązki prowadzącego wraz z szacowaną całkowitą liczbą godzin:
Konsultacje – 5 godzin
Poprawa prac zaliczeniowych i projektowych – 2 godziny
Przygotowanie wykładów - 10 godzin
Przygotowanie ćwiczeń laboratoryjnych, w tym programistycznych - 5 godzin
W sumie: 22 godziny