

## Sylabus modułu kształcenia na studiach wyższych

Nazwa Wydziału	Wydział Matematyki i Informatyki		
Nazwa jednostki prowadzącej moduł	Instytut Informatyki i Matematyki Komputerowej		
Nazwa modułu kształcenia	<b>Hurtownie danych i metody eksploracji danych</b>		
Kod modułu			
Język kształcenia	język polski		
Efekty kształcenia dla modułu kształcenia	Symbol	Efekty kształcenia	Odniesienie do efektów kierunkowych
	E1	Ma wiedzę w zakresie matematyki wyższej obejmującą zagadnienia analizy matematycznej, algebry, matematyki dyskretnej (elementy logiki i teorii mnogości, kombinatoryki i teorii grafów), metod probabilistycznych i statystyki (ze szczególnym uwzględnieniem metod dyskretnych) oraz metod numerycznych	K_W01
	E2	Ma uporządkowaną, podbudowaną teoretycznie wiedzę ogólną w zakresie architektury systemów komputerowych, systemów operacyjnych, technologii sieciowych, baz danych, inżynierii oprogramowania itp.	K_W04
	E3	Zna podstawowe narzędzia wspomagające pracę informatyka	K_W05
	E4	Posiada umiejętności efektywnego posługiwania się oprogramowaniem istniejącym – systemami operacyjnymi, bazami danych, sieciami komputerowymi	K_U06
	E5	Posiada umiejętność przygotowania, realizacji i	K_U05

		weryfikacji projektów informatycznych, zarówno indywidualnie , jak i pracy zespołowej	
	E6	Potrafi pracować zespołowo; rozumie konieczność systematycznej pracy nad projektami o charakterze długofalowym	K_K03
Typ modułu kształcenia (obowiązkowy/fakultatywny)	fakultatywny		
Rok studiów	III rok, studia I stopnia		
Semestr	semestr 3		
Imię i nazwisko osoby/osób prowadzących moduł	dr Jerzy Martyna		
Imię i nazwisko osoby/osób egzaminującej/egzaminujących bądź udzielającej zaliczenia, w przypadku gdy nie jest to osoba prowadząca dany moduł	dr Jerzy Martyna		
Sposób realizacji	<b>Wykład ilustrowany prezentacją komputerową oraz ćwiczeniami w laboratorium komputerowym z użyciem SAS/Enterprise Miner Software firmy SAS, który jest własnością Wydziału Matematyki i Informatyki. Przewiduje się, że wykład ten wraz z przedmiotem „Statystyka bayesowska”, także opartym o oprogramowanie SAS, będzie podstawą do uzyskania certyfikatu korporacji SAS.</b>		
Wymagania wstępne i dodatkowe	Zaliczenie modułu (przedmiotu): Bazy danych		
Rodzaj i liczba godzin zajęć dydaktycznych wymagających bezpośredniego udziału nauczyciela akademickiego i studentów, gdy w danym module przewidziane są takie zajęcia	Łącznie: 60 godz. wykład: 30 godz., laboratorium: 30 godz.		
Liczba punktów ECTS przypisana modułowi	<b>6 pkt. ECTS</b>		
Bilans punktów ECTS	<p style="text-align: right;">Udział w wykładach – 30</p> <p style="text-align: right;">Analiza wybranych pozycji z literatury przedmiotu -20</p> <p style="text-align: right;">Praktyczne ćwiczenia w laboratorium – 30</p> <p style="text-align: right;">Przygotowanie do egzaminu i zaliczanie kolokwium - 20</p> <p style="text-align: right;">Udział w konsultacjach – 1</p>		

	łączy nakład pracy studenta - 102
Stosowane metody dydaktyczne	- Wykład z wykorzystaniem środków multimedialnych, Indywidualne konsultacje raz w tygodniu (2 godz. w tygodniu, 15 tygodni)
Metody sprawdzania i kryteria oceny efektów kształcenia uzyskanych przez studentów	Egzamin pisemny – testy egzaminacyjne są skonstruowane tak, by sprawdzić przewidziane dla przedmiotu efekty kształcenia
Forma i warunki zaliczenia modułu, w tym zasady dopuszczenia do egzaminu, zaliczenia, a także forma i warunki zaliczenia poszczególnych zajęć wchodzących w zakres danego modułu	Pozytywna ocena z egzaminu pisemnego – kryteria oceny podane przy rozpoczęciu zajęć. Skala ocen zgodna z Regulaminem Studiów UJ
Treści modułu kształcenia	<p>Treścią przedmiotu „Hurtownie danych i metody eksploracji danych” jest prezentacja podstawowego zakresu materiału dotyczącego budowy i eksploatacji hurtowni danych oraz metod eksploracji danych, w tym danych gromadzonych w hurtowniach danych.</p> <p><u>Wykład:</u></p> <ol style="list-style-type: none"> <li>1. <b>Hurtownie danych i ich zastosowania</b> <ol style="list-style-type: none"> <li>1.1. Koncepcja hurtowni danych</li> <li>1.2. Składniki hurtowni danych</li> <li>1.3. Wprowadzenie niejednorodnych danych do hurtowni danych</li> <li>1.4. Wyprowadzanie wielowymiarowych danych z hurtowni danych</li> <li>1.5. Architektura hurtowni danych.</li> <li>1.6. Zastosowania hurtowni danych</li> </ol> </li> <li>2. <b>Tworzenie hurtowni danych.</b> <ol style="list-style-type: none"> <li>2.1. Ekstrakcja i ujednoczenie danych.</li> <li>2.2. Agregacja i dostosowanie danych</li> <li>2.3. Optymalizacja zapytań</li> <li>2.4. Modelowanie i pomiar hurtowni danych</li> </ol> </li> <li>3. <b>Odświeżanie hurtowni danych</b> <ol style="list-style-type: none"> <li>3.1. Co to jest odświeżanie hurtowni danych</li> <li>3.2. Przyrostowa ekstrakcja danych</li> <li>3.3. Odświeżanie perspektyw</li> <li>3.4. Program (narzędzie) ETL</li> </ol> </li> <li>4. <b>Wielowymiarowe modele danych i agregacje.</b> <ol style="list-style-type: none"> <li>4.1. Model ROLAP</li> <li>4.2. Model MOLAP.</li> <li>4.3. Model HOLAP.</li> <li>4.4. Modele logiczne informacji wielowymiarowej</li> <li>4.5. Wybór właściwego formalizmu</li> </ol> </li> <li>5. <b>Metadane i jakość hurtowni danych</b> <ol style="list-style-type: none"> <li>5.1. Zarządzanie meta danymi w praktyce</li> <li>5.2. Model magazynu meta danych w metodologii DWQ</li> <li>5.3. Definiowanie jakości hurtowni danych</li> <li>5.4. Analiza jakości hurtowni danych (podejście QFD, podejście</li> </ol> </li> </ol>

GOM itp.)

5.5. Przykład szczegółowej analizy jakości danych

## **6. Projektowanie hurtowni danych z uwzględnieniem jakości**

6.1. Metodologie projektowania hurtowni danych

6.2. Metoda Kimballa projektowania hurtowni danych

6.3. Przykładowy projekt hurtowni danych

## **7. Metody eksploracji danych i zadania eksploracji danych.**

7.1. Definicja eksploracji danych

7.2. Opis

7.3. Estymacja

7.4. Predykcja

7.5. Klasyfikacja .

7.6. Grupowanie

7.9. Odkrywanie reguł.

## **8. Odkrywanie asocjacji**

8.1. Sformułowanie problemu

8.2. Typy reguł asocjacyjnych

8.3. Podstawowe algorytmy odkrywania binarnych reguł asocjacyjnych – alg. A-Prion, alg. FP-Growth.

8.4. Wielopoziomowe i wielowymiarowe reguły asocjacyjne.

8.5. Korelacje vs asocjacje

## **9. Klasyfikacja**

9.1. Sformułowany problem klasyfikacji

9.2. Metody klasyfikacji

9.3. Kryteria oceny metod klasyfikacji podziału.

9.4. Naiwny klasyfikator Bayesa, kNN

9.5. Dokładność klasyfikacji

## **10. Grupowanie**

10.1. Podział metod grupowania

10.2. Metody grupowania hierarchicznego.

10.3. Metody grupowania iteracyjno- optymalizacyjne.

10.4. Podstawowe algorytmy grupowania – alg. *k*-średnich, alg. *k*-medoidów

## **11. Eksploracja tekstu**

11.1. Wyszukiwanie dokumentów

11.2. Reprezentacja tekstu

11.3. Indeksowanie semantyczne SVD

11.4. Struktury danych

## **12. Eksploracja sieci Web**

12.1. Podstawowy algorytm rankingu stron

12.2. Algorytm Hubs&Authorities

12.3. Algorytm PageRank

12.4. Eksploracja logów

### Laboratorium:

Uzupełnieniem wykładu są ćwiczenia, których celem jest przeprowadzenie podstawowych metod i technik związanych z hurtowniami danych oraz eksploracji danych. Ćwiczenia laboratoryjne są prowadzone z użyciem systemów: WEKA, SAS Enterprise Miner 9.3, Rapid Miner, języka MatLab itp. celem laboratorium jest sprawdzenie efektywności poszczególnych metod i poznanie najważniejszych algorytmów używanych w eksploracji danych.

<p>Wykaz literatury podstawowej i uzupełniającej, obowiązującej do zaliczenia danego modułu</p>	<ol style="list-style-type: none"> <li>1. M. Jarke, M. Lenzerini, Y. Vassiliou, P. Vassiliadis, „Hurtownie danych. Podstawy organizacji i funkcjonowania”, WSiP, Warszawa 2003, ISBN 83-02-08850-1.</li> <li>2. C. Adamson, M. Venerable, „Data Warehouse Design Solutions, John Wiley and Sons, 1998.</li> <li>3. T. Chris, „Projektowanie hurtowni danych”, WNT, Warszawa 2005, ISBN 8320431506.</li> <li>4. D. Hand, H. Mannila, P. Smyth, „Eksploracja danych”, WNT, Warszawa 2005.</li> <li>5. J. Han, M. Kambor, „Data Mining. Concepts and Techniques. Second Edition”, The Morgan Kaufmann Publishers, 2006, ISBN 13: 978-1-55860-901-3.</li> <li>6. Z. Markov, D. T. Larose, „Eksploracja zasobów internetowych”, Wyd. Naukowe PWN, Warszawa, 2009, ISBN 978-83-01-15868-2</li> <li>7. D. Larose, „Metody i modele eksploracji danych”, Wyd. Naukowe PWN, Warszawa 2010, ISBN 978-83-01-15466-0.</li> <li>8. D. T. Larose, „Odkrywanie wiedzy z danych. Wprowadzenie do eksploracji danych”, Wyd. Naukowe PWN, Warszawa 2006, ISBN-13: 978-83-01-14836-2.</li> </ol>
<p>Wymiar, zasady i forma odbywania praktyk, w przypadku, gdy program kształcenia przewiduje praktyki</p>	