

KARTA OPISU MODUŁU KSZTAŁCENIA		
Nazwa modułu/przedmiotu Zaawansowane technologie przetwarzania danych		Kod 1010512331010510163
Kierunek studiów Informatyka	Profil kształcenia (ogólnoakademicki, praktyczny) ogólnoakademicki	Rok / Semestr 2 / 3
Ścieżka obieralności/specjalność Technologie przetwarzania danych	Przedmiot oferowany w języku: polski	Kurs (obligatoryjny/obieralny) obligatoryjny
Stopień studiów: II stopień	Forma studiów (stacjonarna/niestacjonarna) niestacjonarna	
Godziny Wykłady: 20 Ćwiczenia: - Laboratoria: 20 Projekty/seminaria: -		Liczba punktów 5
Status przedmiotu w programie studiów (podstawowy, kierunkowy, inny) (ogólnouczelniany, z innego kierunku) kierunkowy z danego kierunku		
Obszar(y) kształcenia i dziedzina(y) nauki i sztuki		Podział ECTS (liczba i %)
Odpowiedzialny za przedmiot / wykładowca:		
dr inż. Tomasz Koszłajda email: Tomasz.Koszłajda@cs.put.poznan.pl tel. 61 6652960 Wydział Informatyki ul. Piotrowo 3, 60-965 Poznań		dr hab. inż. Marek Wojciechowski email: Marek.Wojciechowski@cs.put.poznan.pl tel. 61 6652962 Wydział Informatyki ul. Piotrowo 3, 60-965 Poznań
Wymagania wstępne w zakresie wiedzy, umiejętności, kompetencji społecznych:		
1	Wiedza:	Student rozpoczynający ten przedmiot powinien posiadać podstawową wiedzę z systemów baz danych oraz programowania obiektowego.
2	Umiejętności:	Powinien posiadać umiejętność rozwiązywania podstawowych problemów występujących w analizie systemowej, projektowaniu i tworzeniu programów komputerowych, instalacji, konfigurowania i strojenia oprogramowania systemowego oraz umiejętność pozyskiwania informacji ze wskazanych źródeł. Powinien również rozumieć konieczność poszerzania swoich kompetencji i mieć gotowość do podjęcia współpracy w ramach zespołu.
3	Kompetencje społeczne	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:		
1. Przekazanie studentom podstawowej wiedzy z multimedialnych, przestrzennych i tekstowych baz danych, w zakresie nowych modeli danych dedykowanych do reprezentacji struktur i semantyki danych multimedialnych, obowiązujących standardów, wybranych produktów komercyjnych oferujących składowanie i przetwarzanie danych multimedialnych oraz rozwiązań systemowych związanych z wydajną implementacją takich produktów.		
2. Rozwijanie u studentów umiejętności rozwiązywania problemów związanych z modelowaniem, projektowaniem i konstrukcją programów przetwarzających duże i współdzielone repozytoria danych multimedialnych.		
Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia		
Wiedza:		
1. ma zaawansowaną i pogłębioną wiedzę z zakresu zaawansowanego przetwarzania danych oraz narzędzi, bibliotek i środowisk wykorzystywanych do tego celu - [K2st_W1]		
2. ma zaawansowaną wiedzę szczegółową dotyczącą wybranych zagadnień i standardów z zakresu przetwarzania danych tekstowych, przestrzennych i multimedialnych - [K2st_W3]		
3. ma zaawansowaną i szczegółową wiedzę o procesach zachodzących w cyklu życia aplikacji przetwarzających dane - [K2st_W5]		
4. zna zaawansowane metody, techniki i narzędzia stosowane przy tworzeniu aplikacji do zaawansowanego przetwarzania danych - [K2st_W6]		
Umiejętności:		

1. potrafi pozyskiwać informacje z dokumentacji dotyczącej technologii, bibliotek i narzędzi do przetwarzania danych (w języku polskim i angielskim), analizując ich silne i słabe strony i oceniając adekwatność w kontekście rozwiązywanego problemu - [K2st_U1]
2. potrafi wykorzystać metody eksperymentalne do formułowania i rozwiązywania zadań inżynierskich i prostych problemów badawczych z dziedziny przetwarzania danych - [K2st_U4]
3. potrafi przy analizie danych potrafi integrować wiedzę z dziedziny baz danych, multimediów, analizy języka naturalnego - [K2st_U5]
4. potrafi ocenić przydatność i możliwość wykorzystania nowych bibliotek i narzędzi do przetwarzania danych - [K2st_U6]
5. potrafi dokonać krytycznej analizy istniejących standardów, środowisk, narzędzi i bibliotek stosowanych do przetwarzania danych i wskazać możliwości ich ulepszenia - [K2st_U8]
6. potrafi ocenić przydatność metod i narzędzi służących do rozwiązania zadania inżynierskiego, uwzględniającego zaawansowane przetwarzanie danych, w tym dostrzec ograniczenia tych metod i narzędzi - [K2st_U9]
7. potrafi - stosując m.in. koncepcyjnie nowe metody - rozwiązywać złożone zadania informatyczne, dotyczące nietypowego przetwarzania danych, w tym zadania zawierające komponent badawczy - [K2st_U10]

Kompetencje społeczne:

1. rozumie, że w obszarze zaawansowanych technik przetwarzania złożonych danych wiedza i umiejętności bardzo szybko stają się przestarzałe - [K2st_K1]
2. rozumie znaczenie wykorzystywania najnowszych architektur, wzorców i technologii do przetwarzania złożonych danych - [K2st_K2]

Sposoby sprawdzenia efektów kształcenia

Efekty kształcenia przedstawione wyżej weryfikowane są w następujący sposób:

Ocena formująca:

a) w zakresie wykładów:

- aktywność w trakcie wykładów: szukanie odpowiedzi na pytania zadawane przez wykładowcę, krytyczne podejście do tłumaczenia wykładowców, zainteresowanie rozszerzeniem zakresu wykładów, znajdowanie błędów w materiałach wykładowych,

b) w zakresie laboratoriów:

- 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 charakterze problemowym, który składa się z 5 otwartych zadań, polegających na przykład na odtworzeniu działania wybranych algorytmów, liczbowej weryfikacji danej hipotezy, itp. Dla uzyskania oceny 3.0 wymagane jest uzyskanie co najmniej 50% punktów. W ocenie finalnej uwzględnia jest również ocena z laboratorium oraz aktywność w trakcie wykładów.

- omówienie wyników egzaminu,

b) w zakresie laboratoriów weryfikowanie założonych efektów kształcenia realizowane jest przez:

- ocenę przygotowania studenta do poszczególnych sesji zajęć laboratoryjnych (sprawdzian wejściowy) oraz ocenę umiejętności związanych z realizacją ćwiczeń laboratoryjnych,

- ocenianie ciągle, na każdych zajęciach (odpowiedzi ustne) - premiowanie przyrostu umiejętności posługiwania się poznanymi zasadami i metodami,

- ocenę realizacji ćwiczeń praktycznych,

- ocenę wiedzy i umiejętności związanych z realizacją zadań projektowych / laboratoryjnych poprzez 2 kolokwia w semestrze,

- ocenę i obronę przez studenta sprawozdania z realizacji projektu,

Uzyskiwanie punktów dodatkowych za aktywność podczas zajęć, a szczególnie za:

- efektywność zastosowania zdobytej wiedzy podczas rozwiązywania zadanego problemu,

- uwagi związane z udoskonaleniem materiałów dydaktycznych,

- wskazywanie trudności percepcyjnych studentów umożliwiające bieżące doskonalenia procesu dydaktycznego.

Treści programowe

Program wykładu obejmuje następujące zagadnienia:

1. Wprowadzenie do multimedialnych baz danych. Identyfikacja specyfiki danych multimedialnych. Ustalenie podstawowych problemów związanych z reprezentacją, składowaniem, wyszukiwaniem i przetwarzaniem danych multimedialnych.
2. Reprezentacja danych multimedialnych jako wektorów w przestrzeni wielowymiarowej. Specyfika wyszukiwania i przetwarzania danych o charakterystyce przestrzennej. Operacje charakterystyczne dla danych przestrzennych. Uzasadnienie potrzeby nowych rozwiązań systemowych gwarantujących wydajność składowania i wyszukiwania danych przestrzennych.
3. Wielowymiarowe struktury danych opisujące podział przestrzeni wielowymiarowej. Pliki kratowe, wielowymiarowe drzewa binarne, wielowymiarowe indeksy: kB-drzewa, hB-drzewa, drzewa ćwiartek (Quad Trees). Struktura, algorytmy wyszukiwania, algorytmy utrzymania i problemy związane ze stosowaniem wielowymiarowych struktur danych dzielących przestrzeń.
4. Wielowymiarowe struktury danych opisujące podział zbiorów obiektów przestrzennych. R-drzewa, R+-drzewa, R*-drzewa, krzywe wypełniające przestrzeń: Hilberta, Peano, itp, R-drzewa Hilberta. Definicja Minimalnych regionów pokrywających - MBR. Kryteria jakości R-drzew. Struktura, algorytmy wyszukiwania, algorytmy utrzymania i problemy związane ze stosowaniem różnych typów R-drzew. Kompresja R-drzew.
5. Zastosowania R-drzew dla rozwiązania problemu wyszukiwania najbliższego sąsiada NN. Algorytm wyszukiwania NN dla dużego zbioru danych przestrzennych. Zastosowanie R-drzew dla przyspieszenia wykonywania połączeń przestrzennych.
6. Specyfika przestrzeni o bardzo dużej liczbie wymiarów, kłótwa wielowymiarowości, utrata znaczenia NN. Przestrzenne struktury dla bardzo wielu wymiarów: X-drzewa, SS-drzewa, SR-drzewa, TV-drzewa oraz indeksy konstruowane za pomocą Pyramid-Technique.
7. Bazy dokumentów tekstowych. Specyfika wyszukiwania dokumentów tekstowych, miary jakości wyszukiwania. Reprezentacje dokumentów tekstowych. Algorytmy wyszukiwania wzorca w dokumentach tekstowych: Knutha-Morrisa-Pratta i Boyera-Moora. Miary odległości między dokumentami tekstowymi. Pliki odwrócone. Numeryczne sygnatury tekstów. Reprezentacja tekstów jako punktów w przestrzeni wielowymiarowej.
8. Architektury systemów multimedialnych baz danych. Modele danych dla multimedialnych baz danych. Klasyfikacje danych multimedialnych.
9. Duże obiekty w bazach danych. Typy danych standardu SQL przeznaczone do przechowywania dużych obiektów. Obsługa dużych obiektów w Oracle.
10. Składowanie, prezentacja i transmisja strumieniowa multimediiów. Striping i replikacja. Problemy strumieniowej transmisji danych multimedialnych wrażliwych na opóźnienia. Algorytmy i metody usprawniające transmisję strumieniową wideo.
11. Specyfika zapytań do multimedialnych baz danych. Wyszukiwanie w oparciu o zawartość (ang. content-based retrieval). Prototypowe języki zapytań do multimedialnych baz danych.
12. Standard SQL/MM: geneza i charakter standardu, szczegółowe omówienie specyfikacji Full Text, Spatial i Still Image. Przykłady zapytań opartych o typy danych standardu SQL/MM.
13. Rodzaje metadanych opisujących obiekty multimedialne. Przegląd standardów metadanych. Standard MPEG-7: geneza i podstawowe założenia, przewidywane zastosowania, format opisów, przegląd deskryptorów.
14. Przegląd systemów zarządzania bazą danych ogólnego przeznaczenia pod kątem wsparcia dla danych multimedialnych. Ewolucja podejścia do obsługi danych multimedialnych w bazach danych na przykładzie wiodących systemów zarządzania bazą danych.

Zajęcia laboratoryjne prowadzone są w formie piętnastu 2-godzinnych ćwiczeń, odbywających się w laboratorium. Ćwiczenia realizowane są przez studentów indywidualnie. Program laboratorium obejmuje następujące zagadnienia:

1. Przetwarzanie danych przestrzennych na przykładzie bazy danych Oracle ? podstawy
2. Przetwarzanie danych przestrzennych na przykładzie bazy danych Oracle ? przetwarzanie
3. Przetwarzanie danych przestrzennych na przykładzie bazy danych Oracle ? standard SQL/MM
4. Przetwarzanie danych przestrzennych na przykładzie bazy danych Oracle ? układy współrzędnych i LRS
5. Przetwarzanie danych przestrzennych na przykładzie bazy danych Oracle ? analiza sieci przestrzennych
6. Przetwarzanie dużych obiektów testowych na przykładzie bazy danych Oracle ? typy indeksów i odpowiadające im operatory, moduły indeksujące: składnica dokumentów, filtr.
7. Przetwarzanie dużych obiektów testowych na przykładzie bazy danych Oracle ? moduł podziału na sekcje, moduł podziału na leksemy, dodatkowe mechanizmy: listy słów wyłączonych, tezaursy; skonkatenowane indeksy pełno-tekstowe, typy klasyfikacji dokumentów tekstowych.
8. Duże obiekty binarne w Oracle. Wykorzystanie typów danych BLOB i BFILE. Obsługa obiektów binarnych procedurami i funkcjami pakietu DBMS_LOB. Ładowanie multimediiów do bazy danych z systemu plików.
9. Duże obiekty tekstowe w Oracle. Wykorzystanie typu danych CLOB i działanie na danych tego typu procedurami i funkcjami pakietu DBMS_LOB oraz podstawowymi funkcjami znakowymi SQL.
10. Obsługa dźwięków i obrazów w Oracle Multimedia. Wykorzystanie podstawowych typów danych Oracle Multimedia. Ładowanie treści binarnych do bazy danych z systemu plików i ich reprezentacja w formie obiektów Oracle Multimedia. Ekstrakcja meta danych i wykorzystanie ich w zapytaniach.
11. Wyszukiwanie obrazów w oparciu o zawartość poprzez typy danych standardu SQL/MM. Wykorzystanie typów danych SQL/MM Still Image z poziomu zapytań SQL i programów w języku PL/SQL.

Cześć wymienionych wyżej treści programowych realizowana jest w ramach pracy własnej studenta.

Metody dydaktyczne:

1. wykład: prezentacja multimedialna, prezentacja ilustrowana przykładami podawanymi na tablicy, rozwiązywanie zadań

2. ćwiczenia laboratoryjne: prezentacja multimedialna, ćwiczenia praktyczne przy komputerach, samodzielnie realizowane projekty, dyskusja

Literatura podstawowa:		
1. Principles of Multimedia Database Systems, V.S. Subrahmanian, Morgan Kaufmann, 1998		
2. Advanced Database Systems, Carlo Zaniolo, Morgan Kaufman, 1997, Part IV Spatial, Text and Multimedia Databases		
3. Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom, Systemy baz danych. Kompletny podręcznik. Wydanie II, 2011, Rozdział 14.4. Indeksy wielowymiarowe		
4. Dokumentacja systemu zarządzania bazą danych Oracle		
Literatura uzupełniająca:		
1. Managing and Mining Multimedia Databases, Bhavani Thuraisingham, CRC Press, 2001		
2. Distributed Multimedia Database Technologies Supported by MPEG-7 and MPEG-21, Harald Kosch, CRC Press, 2003		
3. Specyfikacja standardu SQL/MM		
4. Specyfikacja standardu MPEG-7		
5. Norbert Beckmann, Hans-Peter Knebel, Ralf Schneider, Bernhard Seege, The R*-tree: An Efficient and Robust Access Method for Points and Rectangles, Proceeding SIGMOD '90		
6. Antomn Guttman, R-trees. A Dynamic Index Structure for Spatial Searching Proceeding SIGMOD '84		
7. David B. Lomet, Betty Salzberg, The hB-tree: a multiattribute indexing method with good guaranteed performance, Readings in database systems (2nd ed.) Pages 136 ? 152		
8. Od interMedia do Multimedia - obsługa danych multimedialnych w Oracle 10g/11g, M. Wojciechowski, Materiały XIV konf. PLOUG, 2008		
9. Standard SQL/MM: SQL Multimedia and Application Packages, K. Jankiewicz, M. Wojciechowski, Materiały IX Seminarium PLOUG, 2004		
Bilans nakładu pracy przeciętnego studenta		
Czynność	Czas (godz.)	
1. udział w zajęciach laboratoryjnych	20	
2. udział w konsultacjach związanych z realizacją procesu kształcenia, w szczególności ćwiczeń laboratoryjnych (częściowo realizowane drogą elektroniczną)	4	
3. napisanie programu / programów, uruchomienie i weryfikacja (czas poza zajęciami laboratoryjnymi)	20	
4. przygotowanie do sprawdzianów	10	
5. udział w wykładach	20	
6. omówienie wyników egzaminu	2	
7. przygotowanie do egzaminu i obecność na egzaminie: 18 godz. + 2 godz.	20	
8. zapoznanie się ze wskazaną literaturą / materiałami dydaktycznymi (10 stron tekstu naukowego = 1 godz.), 200 stron	20	
Obciążenie pracą studenta		
forma aktywności	godzin	ECTS
Łączny nakład pracy	116	5
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	48	2
Zajęcia o charakterze praktycznym	40	2