



KARTA OPISU PRZEDMIOTU - SYLABUS

Nazwa przedmiotu

Zaawansowane technologie przetwarzania danych [S2Inf1-TPD>ZTPD]

Przedmiot

Kierunek studiów
Informatyka

Rok/Semestr
1/2

Studia w zakresie (specjalność)
Technologie przetwarzania danych

Profil studiów
ogólnoakademicki

Poziom studiów
drugiego stopnia

Język oferowanego przedmiotu
polski

Forma studiów
stacjonarne

Wymagalność
obligatoryjny

Liczba godzin

Wykład
30

Laboratorium
30

Inne
0

Ćwiczenia
0

Projekty/seminaria
0

Liczba punktów ECTS

5,00

Koordynatorzy

dr hab. inż. Marek Wojciechowski prof. PP
marek.wojciechowski@put.poznan.pl

Wykładowcy

Wymagania wstępne

Student rozpoczynający ten przedmiot powinien posiadać podstawową wiedzę z dziedziny systemów baz danych oraz obiektowego paradygmatu programowania. Powinien posiadać umiejętność rozwiązywania podstawowych problemów występujących w 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. 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 dotyczącej składowania i przetwarzania zbiorów danych o specyficznych typach danych: obrazy, dokumenty tekstowe, dokumenty semistrukturalne i dane przestrzenne (GIS). 2. Rozwijanie u studentów umiejętności rozwiązywania problemów związanych z modelowaniem, projektowaniem i konstrukcją programów przetwarzających duże i współdzielone repozytoria powyższych typów danych.

Przedmiotowe efekty uczenia się

Wiedza:

1. ma zaawansowaną i pogłębioną wiedzę z zakresu systemów informatycznych zarządzających danymi przestrzennymi, obrazami, dokumentami tekstowymi i semistrukturalnymi; (k2st_w1)
2. ma zaawansowaną wiedzę szczegółową dotyczącą obiektowo-relacyjnego modelu danych, metod wyszukiwania obrazów na podstawie ich zawartości (cbir), globalnych i lokalnych deskryptorów obrazów; baz dokumentów tekstowych i semistrukturalnych oraz systemów gis; (k2st_w3)
3. ma wiedzę o trendach rozwojowych i najistotniejszych nowych osiągnięciach w zakresie technik przetwarzania danych tekstowych, semistrukturalnych, przestrzennych i multimedialnych (k2st_w4)
4. ma wiedzę o procesach zachodzących w cyklu życia systemów baz danych; (k2st_w5)
5. zna zaawansowane metody, techniki i narzędzia stosowane w wypadku obsługi danych multimedialnych, takie jak: histogramy kolorów, tekstur i kształtów dla reprezentacji zawartości obrazów. (k2st_w6)

Umiejętności:

1. potrafi pozyskiwać informacje z literatury, baz danych integrować je, dokonywać ich interpretacji i krytycznej oceny; (k2st_u1)
2. potrafi integrować wiedzę z różnych obszarów informatyki: języków programowania, grafiki komputerowej i internetu; (k2st_u5)
3. potrafi ocenić przydatność i możliwość wykorzystania nowych osiągnięć oraz nowych produktów informatycznych (k2st_u6)
4. potrafi dokonać krytycznej analizy istniejących rozwiązań technicznych oraz zaproponować ich ulepszenia, np. przez rozszerzenia modelu obiektowego; (k2st_u8)
5. potrafi ocenić przydatność metod i narzędzi służących do składowania i wyszukiwania danych różnych typów; (k2st_u9)

Kompetencje społeczne:

1. rozumie, że w informatyce wiedza i umiejętności bardzo szybko stają się przestarzałe, na przykładzie ewolucji multimedialnych baz danych lub obiektowych baz danych; (k2st_k1)
2. rozumie znaczenie wykorzystywania najnowszej wiedzy z zakresu informatyki w rozwiązywaniu problemów badawczych i praktycznych, np w dziedzinie rozpoznawania obrazów w samochodach autonomicznych. (k2st_k2)

Metody weryfikacji efektów uczenia się i kryteria oceny

Efekty uczenia się przedstawione wyżej weryfikowane są w następujący sposób:

Ocena formująca:

a) w zakresie wykładów:

- obecności i 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 przygotowania do realizacji zadań,
- na podstawie dyskusji na temat materiału prezentowanego na wykładach

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 kilku 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 aktywność w trakcie wykładów.

- omówienie wyników egzaminu,

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

- rozliczenie z realizacji ćwiczeń laboratoryjnych,

- ocenę i obronę przez studenta 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 modułu obejmuje następujące zagadnienia:

1. Nowe dziedziny zastosowań baz danych i ich specyfika.
2. Obiektowo-relacyjne bazy danych.
3. Przechowywanie i wyszukiwanie obrazów.
4. Bazy dokumentów tekstowych. Wyszukiwanie pełnotekstowe.
5. Wielowymiarowe struktury danych i indeksy.
6. Przechowywanie i przetwarzanie w bazach danych dużych obiektów binarnych (BLOB) i tekstowych (CLOB).
7. Przestrzenne bazy danych.
8. Standardy dotyczące przetwarzania danych tekstowych, przestrzennych i multimedialnych.
9. Technologie XML.

Tematyka zajęć

Program wykładów obejmuje następujące zagadnienia:

1. Nowe dziedziny zastosowań baz danych i ich specyfika. Nieadekwatność klasycznych systemów baz danych do rozwiązywania problemów w nowych dziedzinach zastosowań. Potrzeba znalezienia nowych generacji systemów baz danych, opracowania nowych modeli danych i nowych rozwiązań systemowych.
2. Obiektowo-relacyjne bazy danych; nowe konstrukcje: klasy, dziedziczenie klas, obiektowy typ danych, konstruktory złożonych typów danych; systemowy identyfikator obiektów; referencje między obiektami, polimorficzne kolekcje obiektów; język definiowania, zapytań i przetwarzania obiektów w bazie danych.
3. Przechowywanie i wyszukiwanie obrazów. Wyszukiwanie obrazów za pomocą ich zawartości (CBIR). Deskryptory obrazów: deskryptory kolorów, tekstury i kształtów. Lokalne deskryptory obrazów: punkty charakterystyczne (punkty zainteresowania). Algorytmy identyfikacji i deskryptory punktów charakterystycznych.
4. 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.
5. Wielowymiarowe struktury danych. Wielowymiarowe indeksy: R-drzewa, R+-drzewa, R*-drzewa i R-drzewa Hilberta. Wyszukiwanie najbliższych sąsiadów za pomocą R-drzew. Struktury danych i algorytmy dla dużej liczby wymiarów.
6. Przechowywanie i przetwarzanie w bazach danych dużych obiektów binarnych (BLOB) i tekstowych (CLOB). Zalety i wady przechowywania dużych obiektów w bazie danych i w systemie plików.
7. Przestrzenne bazy danych: reprezentacja danych przestrzennych w bazach danych, indeksowanie przestrzennych baz danych, zapytania do przestrzennych baz danych z wykorzystaniem relacji przestrzennych.
8. Standardy dotyczące przetwarzania danych tekstowych, przestrzennych i multimedialnych: SQL/MM, MPEG-7.
9. Technologie XML: przestrzenie nazw XML, standard XPath, arkusze stylów XSL, transformacje XSL, schematy XML (XML Schema), język zapytań XQuery, standard SQL/XML.

Program laboratoriów obejmuje następujące zagadnienia:

1. Obiektowo-relacyjny model danych w Oracle: typy obiektowe, sposoby składowania obiektów, dziedziczenie, kolekcje.
2. Przechowywanie i przetwarzanie dużych obiektów binarnych (BLOB) i tekstowych (CLOB) w Oracle.
3. Przechowywanie i przetwarzanie danych przestrzennych w Oracle przy użyciu opcji Oracle Spatial: typ SDO_GEOMETRY, indeksy przestrzenne, zależności przestrzenne, typy danych SQL/MM, układy LRS.
4. Indeksowanie i przeszukiwanie danych tekstowych (ang. full text) przy użyciu baz danych SQL (na przykładzie Oracle Text) i środowisk opartych na bibliotece Lucene.
5. Przetwarzanie XML: walidacja, transformacje, wyszukiwanie.

Metody dydaktyczne

Wykład: prezentacja multimedialna, ilustrowana przykładami podawanymi na tablicy.

Ćwiczenia laboratoryjne: prezentacja multimedialna prezentacja ilustrowana przykładami podawanymi na tablicy oraz wykonanie zadań podanych przez prowadzącego - ćwiczenia praktyczne.

Literatura

Podstawowa

1. Advanced Database Systems, Carlo Zaniolo, Morgan Kaufman, 1997, Part IV Spatial, Text and Multimedia Databases
 2. Beginning XML, 4th Edition, David Hunter, Jeff Rafter, Joe Fawcett, Eric van der Vlist, Danny Ayers, Wrox, 2007
 3. Principles of Multimedia Database Systems, V.S. Subrahmanian, Morgan Kaufmann, 1998
 4. Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom, Systemy baz danych. Kompletny podręcznik. Wydanie II, 2011, Rozdział 14.4. Indeksy wielowymiarowe
 5. Dokumentacja systemu zarządzania bazą danych Oracle
 6. Dokumentacja Apache Solr
- 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 Knegel, Ralf Schneider, Bernhard Seege, The R*-tree: An Efficient and Robust Access Method for Points and Rectangles, Proceeding SIGMOD '90
 6. Anton 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. Krzysztof Jankiewicz, Marek Wojciechowski, Standard SQL/MM: SQL Multimedia and Application Packages, Materiały IX Seminarium PLOUG, 2004

Bilans nakładu pracy przeciętnego studenta

	Godzin	ECTS
Łączny nakład pracy	125	5,00
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	60	2,50
Praca własna studenta (studia literaturowe, przygotowanie do zajęć laboratoryjnych/ćwiczeń, przygotowanie do kolokwium/egzaminu, wykonanie projektu)	65	2,50