COURSE DESCRIPTION CARD - SYLLABUS

Course name
Advanced data processing technologies

Course

Field of study
Computing

Year/Semester
1/2

Area of study (specialization)
Data Processing Technologies

Profile of study
general academic

Level of study
Second-cycle studies

Course offered in
Polish

Form of study
full-time

Requirements
elective

Number of hours

Lecture	Laboratory classes	Other (e.g. online)
30	30	
Tutorials	Projects/seminars	

Number of credit points

5

Lecturers

Responsible for the course/lecturer:
Tomasz Koszlajda, PhD
email: Tomasz.Koszlajda@cs.put.poznan.pl
phone: 61 6652960
Faculty of Computing and Telecommunications
address: Piotrowo 2, 60-965 Poznań

Responsible for the course/lecturer:
Marek Wojciechowski, PhD, Dr. Habil.
email: Marek.Wojciechowski@cs.put.poznan.pl
phone: 61 6652962
Faculty of Computing and Telecommunications
address: Piotrowo 2, 60-965 Poznań

Prerequisites

A student beginning this course should have basic knowledge of database systems and object-oriented programming paradigm.

They should have the ability to solve basic problems occurring in the design and development of computer programs, installation, configuration and tuning of system software and the ability to obtain information from specified sources. They should also understand the need to broaden their competences and be ready to cooperate within the team.

Moreover, in the scope of social competences they must present such attitudes as honesty, responsibility, perseverance, cognitive curiosity, creativity, personal culture, respect for other people.

Course objective

1. Provide students with basic knowledge about the storage and processing of data sets with specific data types: images, text documents, semi-structured data, and spatial data (GIS).

2. Develop students' ability to solve problems related to modeling, designing and developing applications that process large and shared repositories of the aforementioned data types.

Course-related learning outcomes

Knowledge

1. has advanced and in-depth knowledge of information systems managing spatial data, images, text documents and semi-structured data; (K2st_W1)

2. has advanced detailed knowledge of object-relational data model, content-based image retrieval (CBIR) methods, global and local image descriptors; textual and semi-structured document databases and GIS systems; (K2st_W3)

3. has knowledge of development trends and the most significant new developments in textual, semi-structured, spatial and multimedia data processing techniques (K2st_W4)

4. has knowledge of the processes occurring in the life cycle of database systems; (K2st_W5)

5. knows advanced methods, techniques and tools used in the case of multimedia data handling, such as color, texture and shape histograms for image content representation. (K2st_W6)

Skills

1. is able to extract information from literature and databases, integrate it, interpret it and critically evaluate it; (K2st_U1)

2. is able to integrate knowledge from different areas of computer science: programming languages, computer graphics, and the Internet; (K2st_U5)

3. is able to assess the usefulness and usability of new developments and new IT products (K2st_U6)

4. is able to critically analyze existing technical solutions and propose improvements to them, e.g. through object model extensions; (K2st_U8)

5. is able to assess the suitability of methods and tools for storing and retrieving different types of data; (K2st_U9)

Social competences

1. understands that in computer science knowledge and skills are rapidly becoming obsolete, as illustrated by the evolution of multimedia or object-oriented databases; (K2st_K1)

2. understands the importance of using the latest IT knowledge in solving research and practical problems, e.g. in the field of image recognition in autonomous cars. (K2st_K2)

Methods for verifying learning outcomes and assessment criteria

Learning outcomes presented above are verified as follows:

Formative evaluation:

(a) as for lectures:

- attendance and activity during lectures: answering questions asked by the lecturer, critical approach to the lecturer's explanations, interest in extending the scope of lectures, finding errors in lecture materials

(b) as for laboratories:

- on the basis of an assessment of preparation for the tasks,
- on the basis of a discussion on the material presented during the lectures

Summative evaluation:

a) as for lectures, the verification of the assumed educational results is carried out by:

- assessment of knowledge and skills demonstrated at a written problem-oriented examination, consisting of several open tasks, such as recreating the operation of selected algorithms, numerical verification of a given hypothesis, etc. In order to obtain the grade 3.0 (E) obtaining at least 50% of available points is required. Activity during lectures also contributes to the final grade.
- discussion of examination results

(b) as for laboratories the verification of the training results achieved shall be carried out by:

- verification of the realization of laboratory exercises
- the evaluation and defense of the project

Obtaining additional points for activity during classes, especially for:

- effectiveness of applying the acquired knowledge while solving a given problem
- remarks regarding potential improvement of teaching materials
- poining out the students' perception difficulties, enabling ongoing improvement of the didactic process

Programme content

The programme of lectures includes the following topics:

1. New fields of application of databases and their specific nature. Inadequacy of traditional database systems to solve problems in new application areas. Need to develop new generations of database systems, new data models and new system solutions.

2. Object-relational databases; new constructs: classes, class inheritance, object data type, constructors of complex data types; system object identifier; references between objects, polymorphic object collections; language for defining, querying and processing objects in a database.

3. Storage and search of images. Content-based image retrieval (CBIR). Image descriptors: color, texture and shape descriptors. Local image descriptors: characteristic points (points of interest). Identification algorithms and descriptors of characteristic points.

4. Text document databases. Specific nature of text document search. Search quality measures. Representations of text documents. Algorithms of pattern search in text documents: Knuth-Morris-Pratt and Boyer-Moor. Measures of distance between text documents. Inverted files. Numeric text signatures. Representation of texts as points in a multidimensional space.

5. Multidimensional data structures. Multidimensional indexes: R-trees, R+-trees, R*-trees and Hilbert R-trees. Searching for nearest neighbors with R-trees. Data structures and algorithms for a large number of dimensions.

6. Storage and processing of large binary (BLOB) and textual (CLOB) objects in databases. Advantages and disadvantages of storing large objects in a database and in the file system.

7. Spatial databases: representation of spatial data in databases, spatial database indexing, spatial database queries using spatial relationships.

8. Standards for text, spatial and multimedia data processing: SQL/MM, MPEG-7.

9. XML technologies: XML namespaces, XPath standard, XSL stylesheets, XSL transformations, XML schema, XQuery query language, SQL/XML standard.

The programme of the laboratories includes the following topics:

1. Object-relational data model in Oracle: object types, object storage methods, inheritance, collections.

2. Storing and processing large binary (BLOB) and text (CLOB) objects in Oracle.

3. Storing and processing spatial data in Oracle using Oracle Spatial: SDO_GEOMETRY type, spatial indexes, spatial relationships, SQL/MM data types, LRS reference systems.

4. Indexing and searching full-text data using SQL databases (using Oracle Text as an example) and environments based on the Lucene library.

5. XML processing: validation, transformations, searching.

Teaching methods

Lecture: multimedia presentation, illustrated with examples given on the board.

Laboratory classes: multimedia presentation illustrated by examples given on the board; completion of tasks given by the instructor (practical exercises).

Bibliography

Basic

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, Database Systems: The Complete Book (2nd Edition), 2011, Chapter 14.4. Multidimensional indexes

5. Oracle DBMS documentation

6. Apache Solr documentation

Additional

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. SQL/MM standard specifiation

4. MPEG-7 standard specifiation

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. 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. Krzysztof Jankiewicz, Marek Wojciechowski, Standard SQL/MM: SQL Multimedia and Application Packages, Materiały IX Seminarium PLOUG, 2004 (in Polish)

Breakdown of average student's workload

	Hours	ECTS
Total workload	125	5,0
Classes requiring direct contact with the teacher	60	2,5
Student's own work (literature studies, preparation for laboratory classes, preparation for exam, project preparation) [1]	65	2,5

[1] delete or add other activities as appropriate