DOCTORAL SCHOOL OF INFORMATICS
COMPLEX EXAM SUBJECT

Spatial Databases and Data Mining (main subject)

The exam focuses on the management, processing, and analysis of geospatial data in database and programming environments. Students are expected to demonstrate theoretical understanding and practical skills in handling vector and raster data, using spatial databases (PostgreSQL with PostGIS), and applying geoinformatics data mining techniques. Students should be able to manage spatial data, including data types, spatial indexing, map projections, and topological concepts. They must demonstrate competence in using built-in geoprocessing functions, importing and exporting spatial data, and integrating databases with web-based applications. The exam also covers geoinformatics data mining, including retrieving data from sources such as OpenStreetMap, processing vector and raster data using Python libraries (GDAL/OGR, Shapely, GeoPy, etc.), working with GTFS and other geospatial data formats, converting between formats (e.g., text files to GeoJSON or KML), and visualizing geospatial data. Students are required to integrate theoretical knowledge and practical skills to apply modern geoinformatics tools effectively in research and applied contexts.

1. Fundamentals of database management: relational databases, records, fields, key fields, table relationships, indexing (B-tree, R-tree indexes), client–server systems, and storage engines. Describe a database management system.
2. Structure and management of spatial data in a PostgreSQL/PostGIS environment. WKT and WKB formats. Data types. Spatial indexing. Handling map projections. Other functions that return geometric information.
3. Functions that return the dimensions of vector features. Distances, measurements, and enclosed angles.
4. Use of geometry-creating functions. Bounding geometries. Intersection, difference, symmetric difference, union, clip, buffer zones. Shortest line.
5. Geoprocessing tools (analysis of spatial relationships): intersects, crosses, overlaps, contains, etc.
6. Topological data structures (introduction to the topology module). Advantages of line simplification in a topological data structure.
7. Handling raster data in databases in analytical workflows.
8. SQL in GIS software. Describe the possibilities for establishing database connections, performing database operations, and importing and exporting vector and raster data to and from databases.
9. Data mining in geoinformatics. Describe the classical steps of data mining. Present the same process in geoinformatics through a real-world problem.
10. Querying, cleaning, and processing geospatial data from various sources. Possible preprocessing steps using scripts. Use of Python libraries (GDAL/OGR, Shapely, Pandas, GeoPandas, GeoPy, Matplotlib, etc.) for vector and raster data.
11. Transfer and transformation of spatial data between different formats (e.g., GeoJSON, KML, database files, Shapefile, GPKG, etc.).
12. Integration of spatial data into web applications and services.
13. Effective presentation of geospatial data for analysis, research, and decision-making. Information.

14. Introduction to the OpenStreetMap spatial database. Data structure, map features, and feature groups. Download options. Use of map data. Data quality.

Literature:

- Ungvári Zsuzsanna: Data mining in geoinformatics. Lecture notes. 2025. https://www.inf.elte.hu/dstore/document/227925/ungv%C3%A1ri_zsuzsanna_data%20mining.pdf
- Ungvári Zsuzsanna: Spatial Databases. Lecture notes. 2025. https://www.inf.elte.hu/dstore/document/227926/ungv%C3%A1ri_zsuzsanna_spatial_databases_jegyzet_vegleges.pdf
- PostGIS Documentation: https://postgis.net/documentation/
- Nagy G.: Spatial databases online lecture notes 1-6. chapters.
  https://regi.tankonyvtar.hu/hu/tartalom/tamop425/0027_SDO1/index.html
  https://regi.tankonyvtar.hu/hu/tartalom/tamop425/0027_SDO2/index.html
  https://regi.tankonyvtar.hu/hu/tartalom/tamop425/0027_SDO3/index.html
  https://regi.tankonyvtar.hu/hu/tartalom/tamop425/0027_SDO4/index.html
  https://regi.tankonyvtar.hu/hu/tartalom/tamop425/0027_SDO5/index.html
  https://regi.tankonyvtar.hu/hu/tartalom/tamop425/0027_SDO6/index.html

- OpenStreetMap WIKI: https://wiki.openstreetmap.org/wiki/Main_Page, 2020
- Pang-Ning Tan, Michael Steinbach, Vipin Kumar: Introduction to Data Mining. Pearson Education Inc., 2019 ISBN: 9780133128901, 0133128903. https://www-users.cs.umn.edu/~kumar001/dmbook/index.php#chapters
- PostgreSQL Documentation: https://www.postgresql.org/docs/
- Ian H. Witten, Eibe Frank, Mark A. Hall, Christopher J. Pal: Data mining. Practical Machine Learning Tools and Techniques. ELSEVIER SCIENCE & TECHNOLOGY. 2017. ISBN 9780128042915 https://www.cs.waikato.ac.nz/~ml/weka/book.html
- Obe.E Regina: PostGIS in Action. Shelter Island, Manning, 2015. ISBN: 9781935182269
- Arsanjani, Zipf, Mooney, Helbich (eds.): OpenStreetMap in GIScience: Experiences, Research, and Applications. Springer, 2015. ISBN: 9783319142807