



DATABASE SYSTEMS

Case studies, privacy and security



BUDAPESTI MŰSZAKI
ÉS GAZDASÁGTUDOMÁNYI EGYETEM
Építőmérnöki Kar - építőmérnöki képzés 1782 óta

Fotogrammetria és Térinformatika Tanszék

Bence Molnár

2022.03.10.

AGENDA

- Homework experiences
- Case study
 - *Online statistics*
- Privacy and security
- Case study II.
 - *WebDLT*

Homework experiences



BUDAPESTI MŰSZAKI
ÉS GAZDASÁGTUDOMÁNYI EGYETEM

Építőmérnöki Kar - építőmérnöki képzés 1782 óta

Fotogrammetria és Térinformatika Tanszék

NEW HW PHASE OPENS NEXT WEEK

- 32 uploaded homework (40 students)
- Reviewed
 - *Interesting topics - <https://abr.fmt.bme.hu/listprojects>*
 - *Checklist violations and lecturer feedback is visible below your upload.*
 - *You can upload anew version anytime, I'll give feedback repeatedly*
- Lottery
 - *Same faculty and specialization*
 - *Same category, same difficulty (I cannot guarantee, but try my best)*
- Statistics - <https://abr.fmt.bme.hu/results>

CHECKLIST

- Is it possible to derive Output information from Input data?
- **Does every input dataset logically connect?**
- **Is it possible to define four relations based on specification?**
- Have you included at least 100 rows in the submitted dataset?
- If raw data is available through multiple files, please compress them into a ZIP file. In case You upload a new version, please include all files again!

GUIDELINES

- Validate official e-mail address in Neptun system
- You can upload a new version any time, but only last one is visible, so make sure that last one includes everything.
- If you decided/forced to not finishing the class, please inform lecturers ASAP.
- Messages
 - *There is a separate page for each project*
 - *Comments are colored based on different roles in project*
 - *Messaging with lecturers is not possible here, use project fields instead*
 - *Do not keep participants waiting; reply ASAP*
- Start working on tasks as soon as possible, please consider, that other participants may not available on last week and they cannot answer to you.
- Is it a good plan to do the job instead of others? Even if it might be faster, I do not recommend it.
- Please keep anonymity and field requirements.
- Please visit checklist and task requirement list before submitting homework

Handling raw data



BUDAPESTI MŰSZAKI
ÉS GAZDASÁGTUDOMÁNYI EGYETEM
Építőmérnöki Kar - építőmérnöki képzés 1782 óta

Fotogrammetria és Térinformatika Tanszék

GENERATING DATA

Good for

- Pilot projects (or for homework)
- Testing – unexpected data
- Excel
 - *RAND()*
 - *RANDBETWEEN()*
- Online sites for data generation
- Obtained from various sources

READING RAW DATA

- Excel
 - CSV
 - Separator
 - Quotation mark vs apostrophe
 - Character encoding – special characters
 - End of Line (EOL)
 - Empty fields
 - PDF
 - Copy - Paste
 - Text to columns
- OCR
 - *Optical character recognition*

USEFUL FUNCTIONS IN EXCEL

- Ctrl+F
- Filter
- VLOOKUP()
 - *Ordered and unordered list*
 - *Compound key?*
- HLOOKUP()
- CONCATENATE() / &
- TRIM()
- LEFT()
- RIGHT()
- FIND()
- REPT()
- TEXT()
- TEXT(B6,REPT("0",C5))

YOUTUBE TUTORIALS

- OCR:
<http://www.youtube.com/watch?v=mR7E7-VAWTQ>
- VLOOKUP:
<http://www.youtube.com/watch?v=VYOSUiRGmuA>
- Text to Columns:
<http://www.youtube.com/watch?v=K7urf2cKPF8>

Case study

Online statistics – Homework site



BUDAPESTI MŰSZAKI
ÉS GAZDASÁGTUDOMÁNYI EGYETEM

Építőmérnöki Kar - építőmérnöki képzés 1782 óta

Fotogrammetria és Térinformatika Tanszék

PERSONAL STATISTICS

- Counter for your clicks on site
 - Number of submitted HW versions
 - Logon locations
 - Your browser type
-
- HW results
 - Test results
 - Final result

Results

Result of T1 and P2 are not grades but categories. Final grade will be calculated based on the P3 and Final test!

T1: D kategória

P2: D kategória

P3: 2

Test: 37

Final test result: 37 (4)

Test minimum requirement: 25

Final grade: 3

Personal statistics

Click made by you on site: 20518

T1 variants: 16

T2 variants: 7

T3 variants: 14

Locations where you logged in: Budapest, Pécs, Sopron

Your browsers: Android Browser, Chrome, Firefox



AGGREGATED STATISTICS

- Online data acquisition – each click related to

- *Time*
- *Visited page (e.g. uploading HW)*
- *Location*
- *Person*

Full statistic (on 2015. 01. 29)

Number of students: 168 [5] [2] [82] [74] [5]
Number of submitted 1st homeworks: 161 (601 version)
Number of corrected 1. homeworks: 161
Number of submitted 2nd homeworks: 156 (410 versions)
Number of corrected 2nd homeworks: 156
Number of submitted 3rd homeworks: 156 (290 versions)
Corrected 3rd homeworks: 156
Number of messages: 751
Number of corrected tests: 0 min: max: average:
Binary data: 356.14 MB
Total site hits: 117069 hit
Hits by students: 87534 hit
Hits by tutors: 21278 hit

- Analyzing all clicks on site will end up an interesting conclusions.
- <https://abr.fmt.bme.hu/results?year=hallgato>

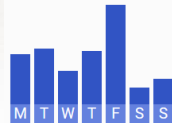
TIME BASED ANALYTICS

- Hourly aggregation – When do we sleep?
- Weekday aggregation – Do we consider weekends?
- Daily aggregation – Can we recognize the deadline?
- Task based aggregation – Does habits change during semester?
- Lecturer activity

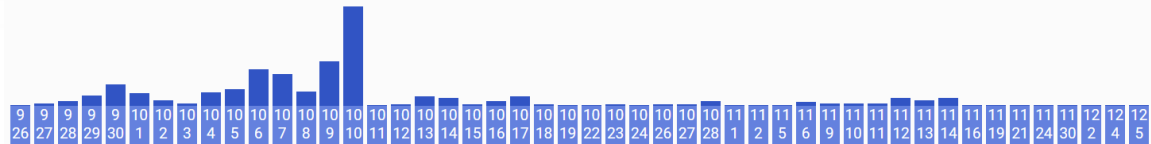
1st homework submissions daily distribution:



1st homework submissions weekly distribution:



1st homework submissions dates:



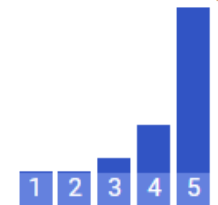
UNDERSTANDING RESULTS

- Test results – was that a difficult test?
- Comparing test groups – were they balanced?
- Which has higher impact on final result: test or HW?
- Why does anyone give up the class?

ZH (B csoport):

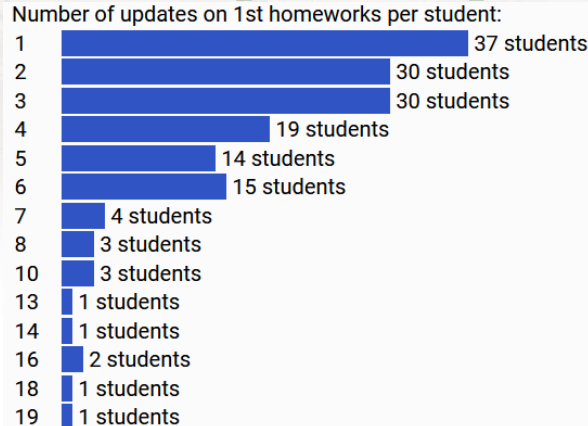


Feladat eredmények:



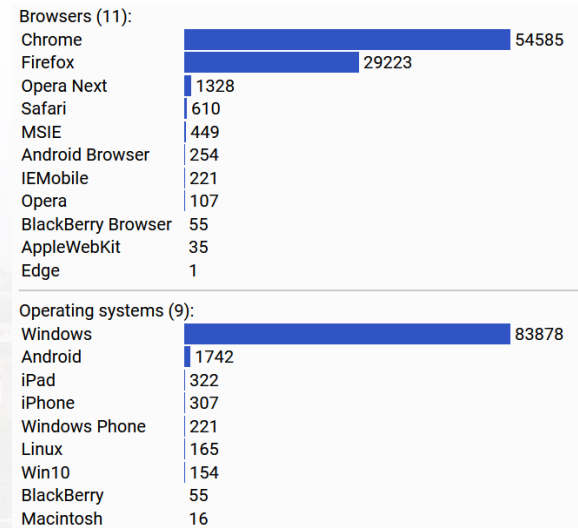
TASK REPORTS

- How many resubmission was required?
 - *Is the task definition clear?*
 - *Were students focused on the task?*
- How many clicks are required to pass the course?

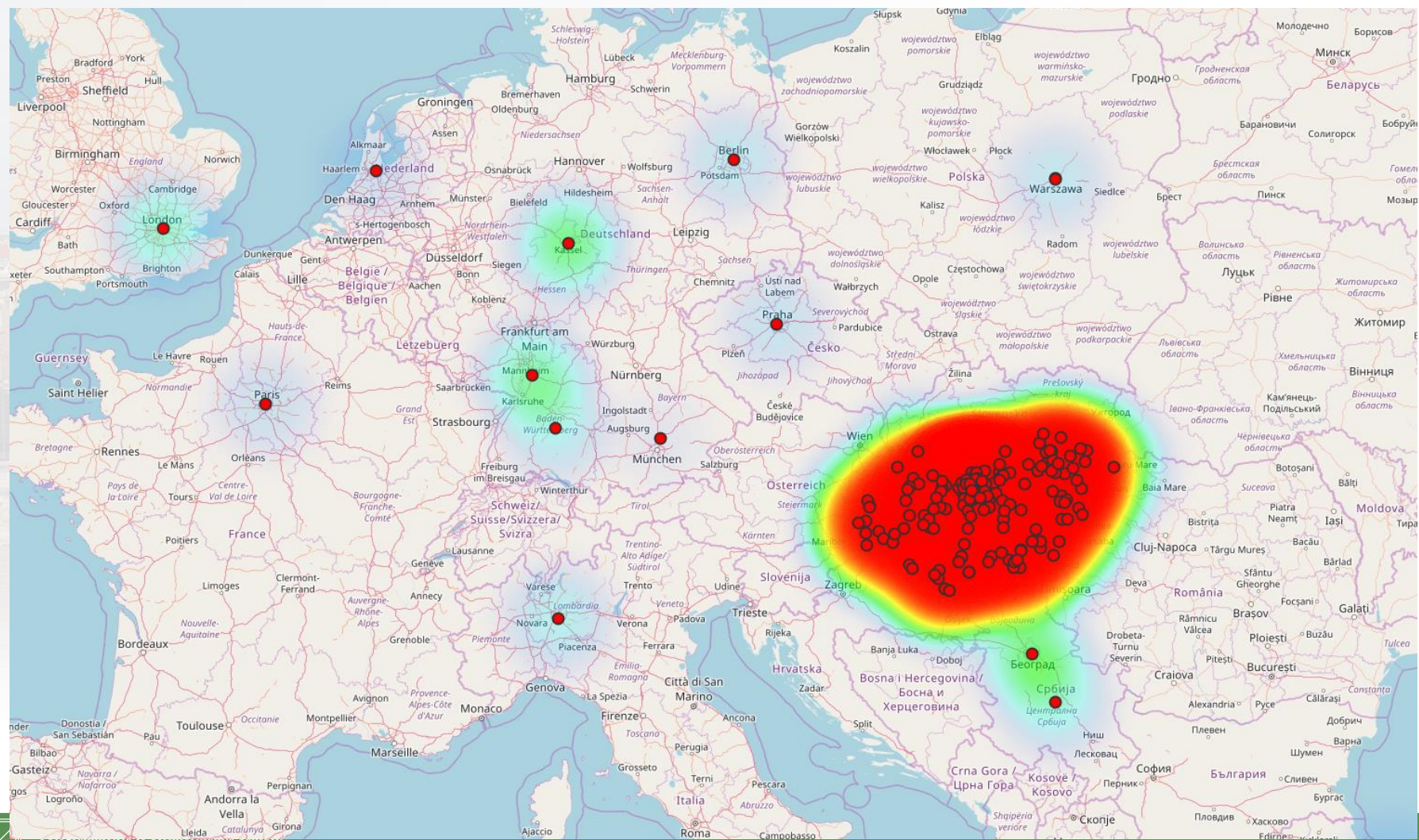


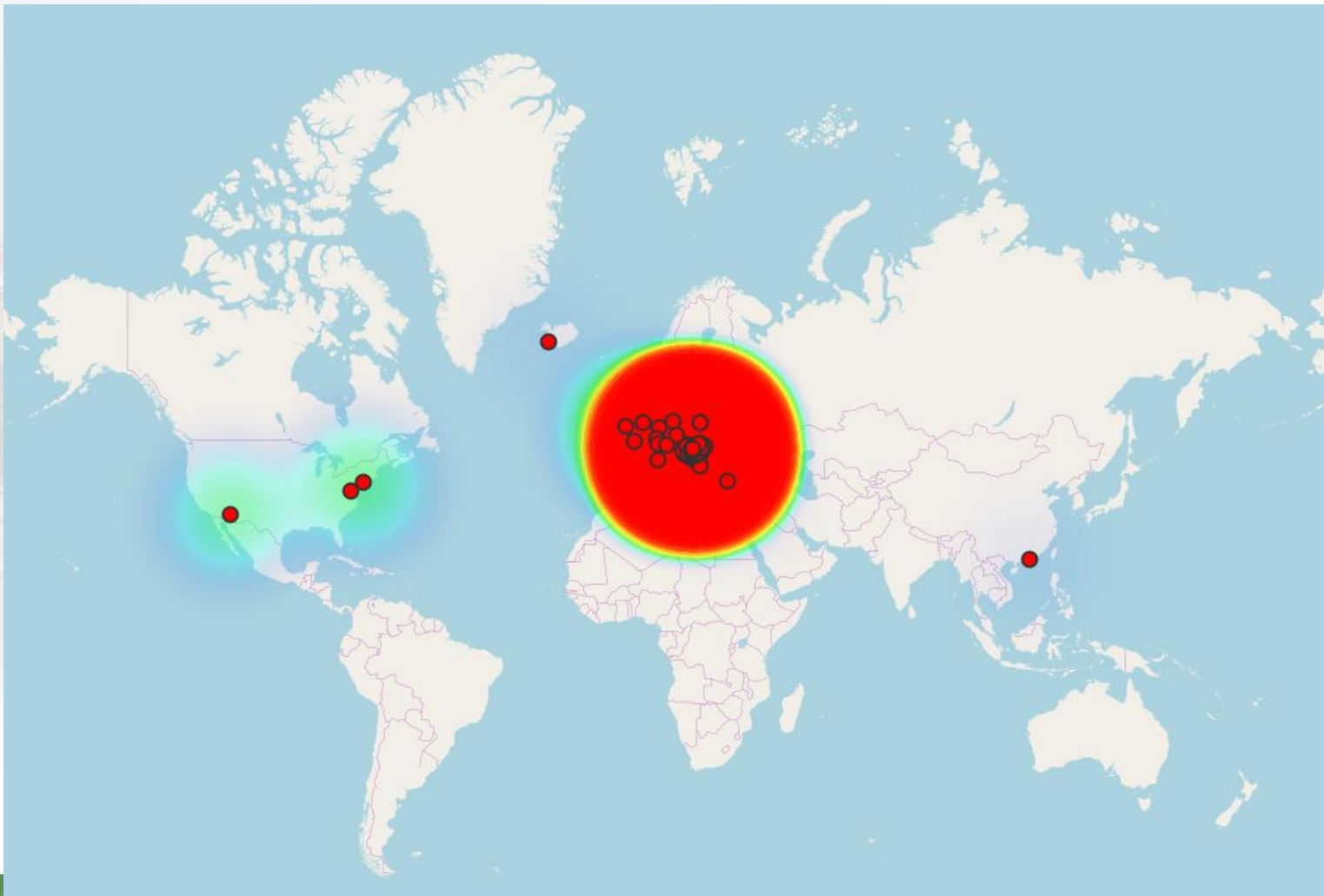
EXTENDED DATA ANALYTICS

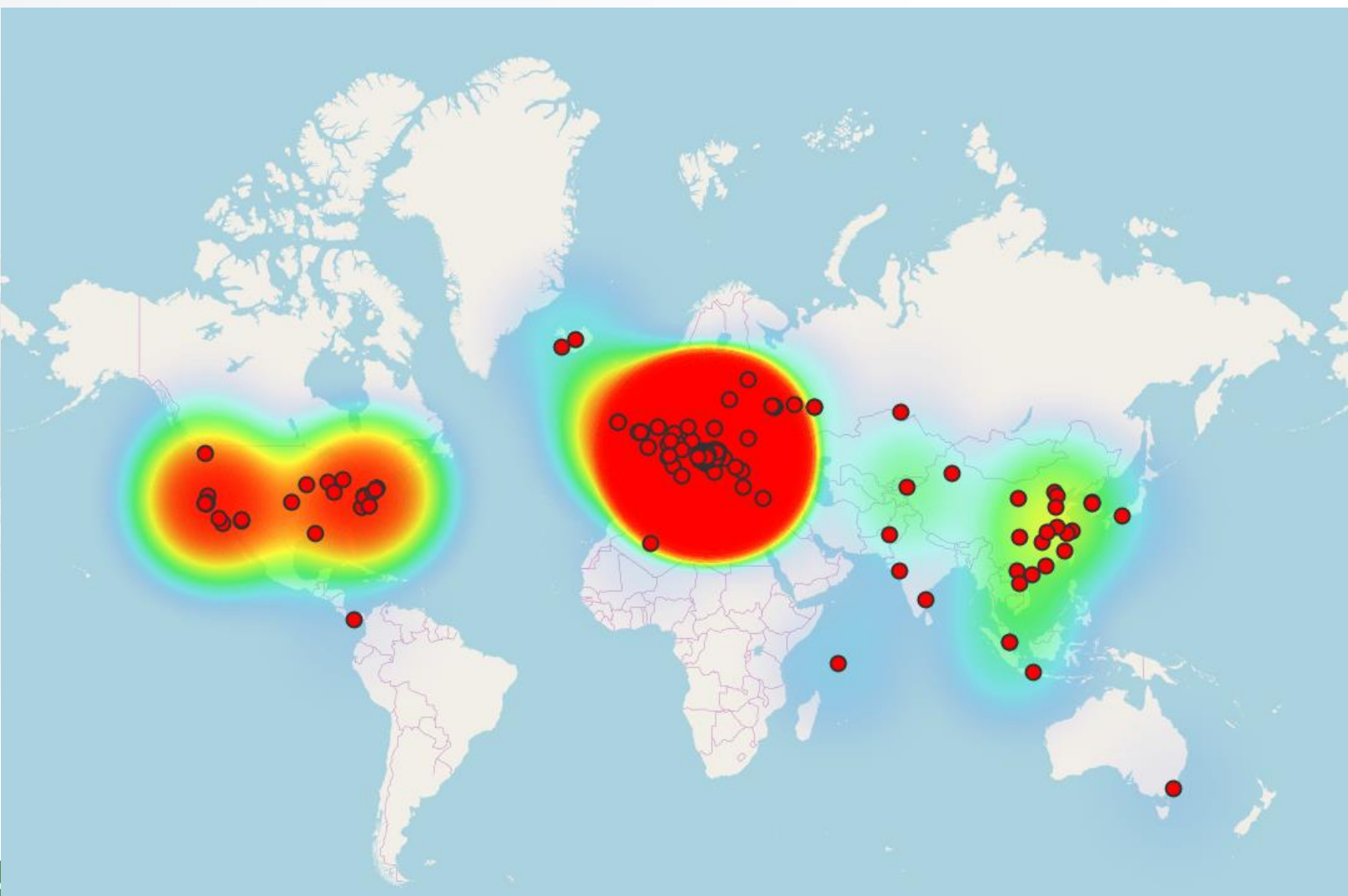
- Popular browsers
- Popular operating systems
- Login location
 - *Each communication on internet is made between IP addresses*
 - *There is a database which maps IP addresses and location*
 - It is not highly accurate – some BME IP addresses are registered to Eastern train station
 - VPN – moves your communication behind a different IP address
 - *Visitors not logged in – e.g. search bots (Google)*











GEOLOCATIONAL ANALYTICS

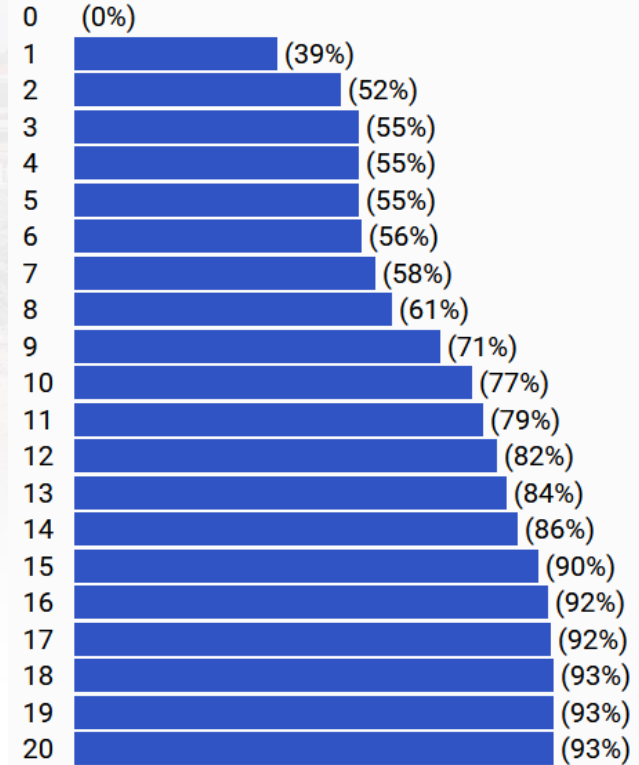
- Agglomeration
- Traveling students
- VPN
- Search bots

VIEWING TEST RESULTS

- How information is spreading?

Viewing test results:

Number of tests:153 Results view count: 150



INDIVIDUALS ARE NOT IDENTIFIABLE

- Is it shocking how data describes you habits?
- Generally we do not track individuals
- Not single clicks are analyzed

We prefer to analyze a big set of humans and clicks

REPORTING OPPORTUNITIES

- HW statistics
 - *Number of submitted, reviewed homework*
 - *Classification: average, min, max, lottery*
- Test statistics
 - *Number of reviewed test*
 - *Grading: average, min, max*
 - *A/B group: count, difficulty*
- **Correlation of HW and test results**
- **Correlation of HW submission time and quality**
- Time based statistics
- **Address, gender, dormitory roommates**
- GPS, IP based statistics
- Platform, browser statistics

UNDER THE HOOD

- MariaDB relational database, as backend
- Linux operating system
- Apache webserver
- PHP interpreter
- HTML, CSS & JavaScript

ABR - Design



BUDAPESTI MŰSZAKI
ÉS GAZDASÁGTUDOMÁNYI EGYETEM

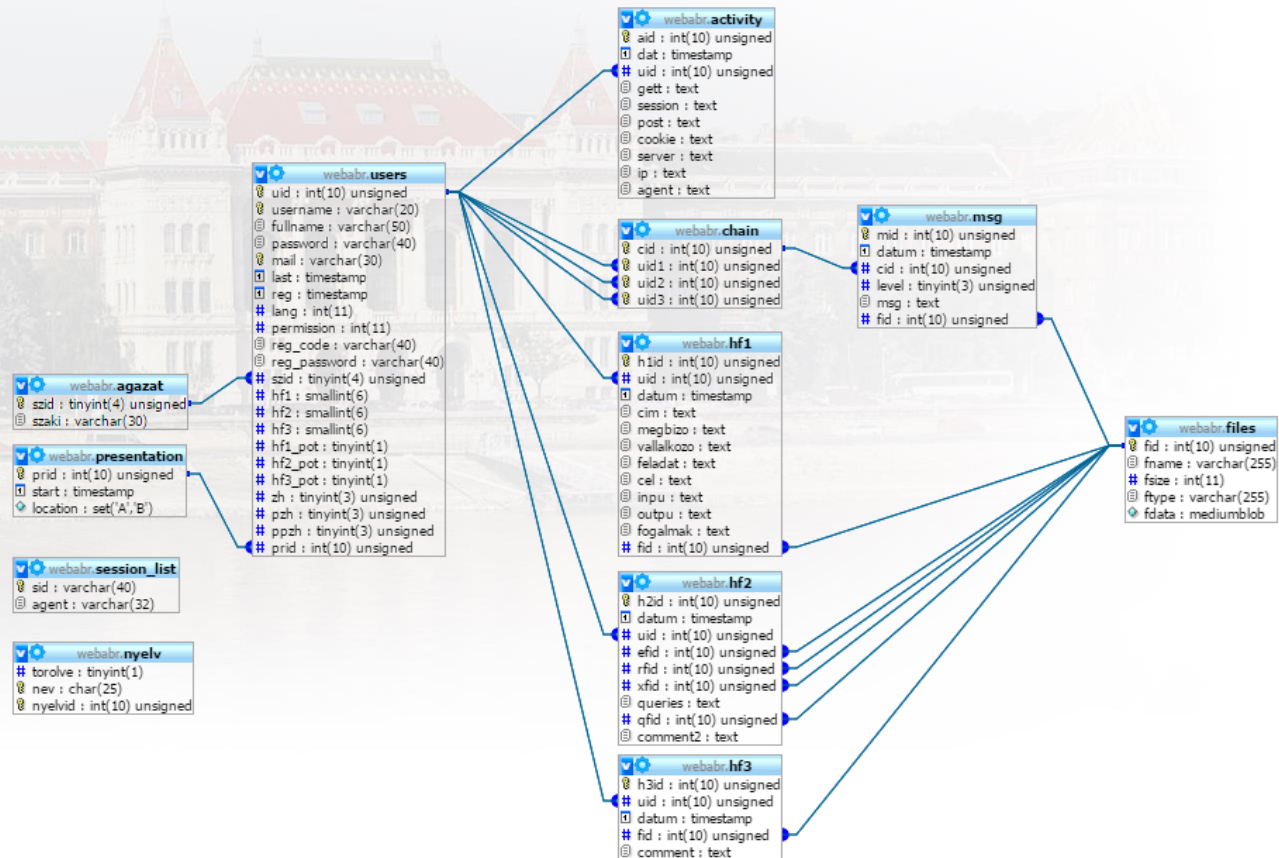
Építőmérnöki Kar - építőmérnöki képzés 1782 óta

Fotogrammetria és Térinformatika Tanszék

SPECIFICATION

- Goal: Site for managing homework submissions and review
- Input
 - *Tasks and messages*
 - *Lecturer tasks (e.g. reviewing)*
 - *User data and activity*
- Output
 - *Displaying projects*
 - *Generating statistics*

RELATIONAL SCHEMA DIAGRAM



Data security and privacy



BUDAPESTI MŰSZAKI
ÉS GAZDASÁGTUDOMÁNYI EGYETEM

Építőmérnöki Kar - építőmérnöki képzés 1782 óta

Fotogrammetria és Térinformatika Tanszék

KEY POINTS

- For reports, we need data
 - *How to obtain? What is the source?*
- Reports are valuable
- Privacy – GDPR
 - *ISP (internet service provider)*
 - *3rd-party*
 - *Man in the middle*
 - Free wifi
 - HTTP/HTTPS
 - *Using same password for multiple systems/sites?*
 - In case of an incident they will have access to all other system as well
 - Sharing accounts between systems

WebDLT



BUDAPESTI MŰSZAKI
ÉS GAZDASÁGTUDOMÁNYI EGYETEM

Építőmérnöki Kar - építőmérnöki képzés 1782 óta

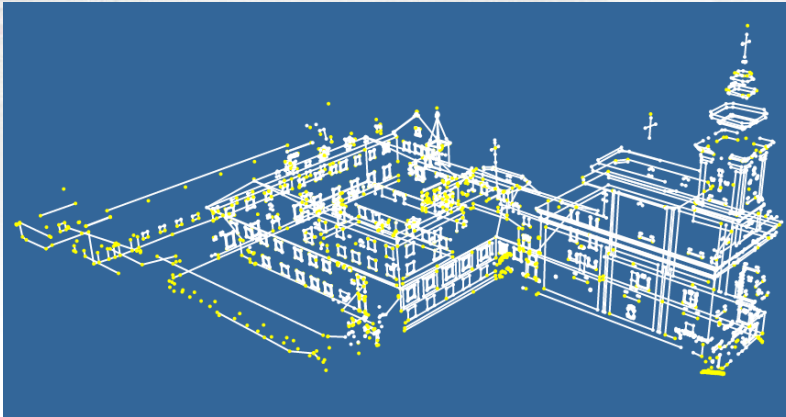
Fotogrammetria és Térinformatika Tanszék

WEBDLT

- Web based photogrammetry
 - *If I can identify same point on multiple photos, I can calculate 3D coordinates of the point*
 - *Storing image coordinates and other metadata in DB*
 - *MariaDB DBMS + PHP + HTML + CSS + JavaScript*
- <https://dlt.fmt.bme.hu>
 - *test / test*
 - *Open project: 10.19*
 - *Digitalization*
 - *Results*



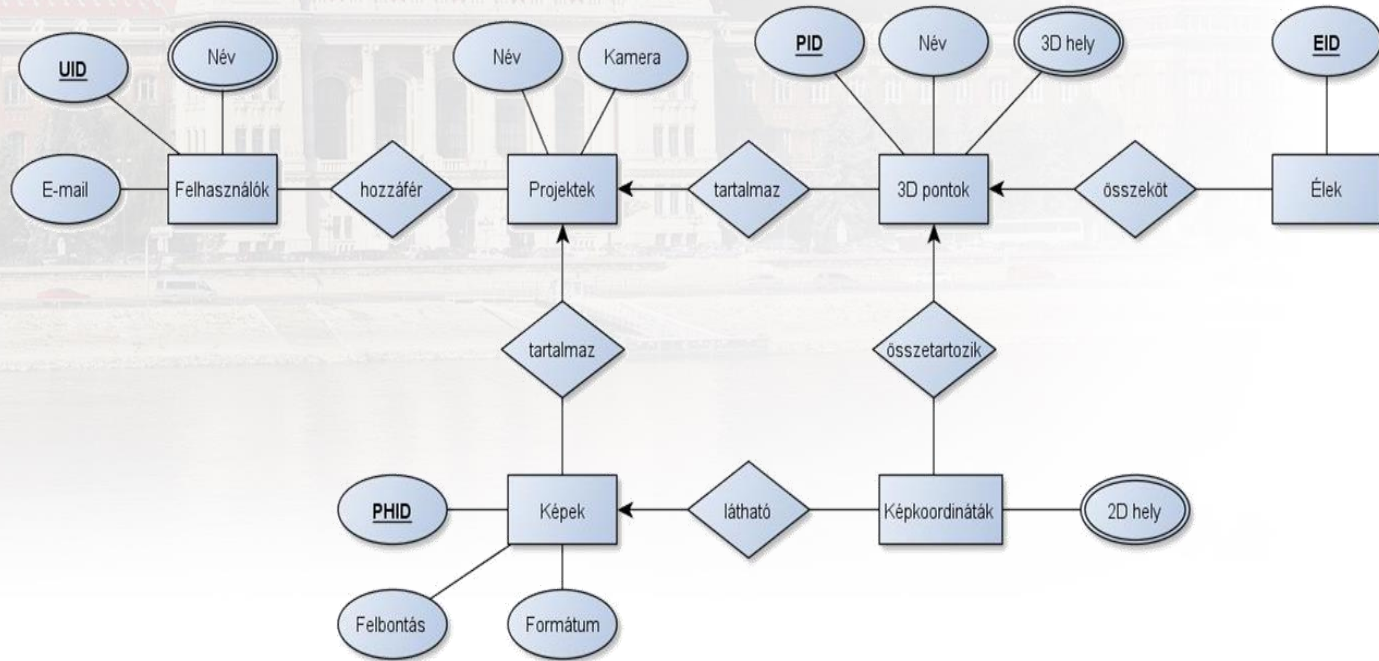
SAMPLE PROJECT



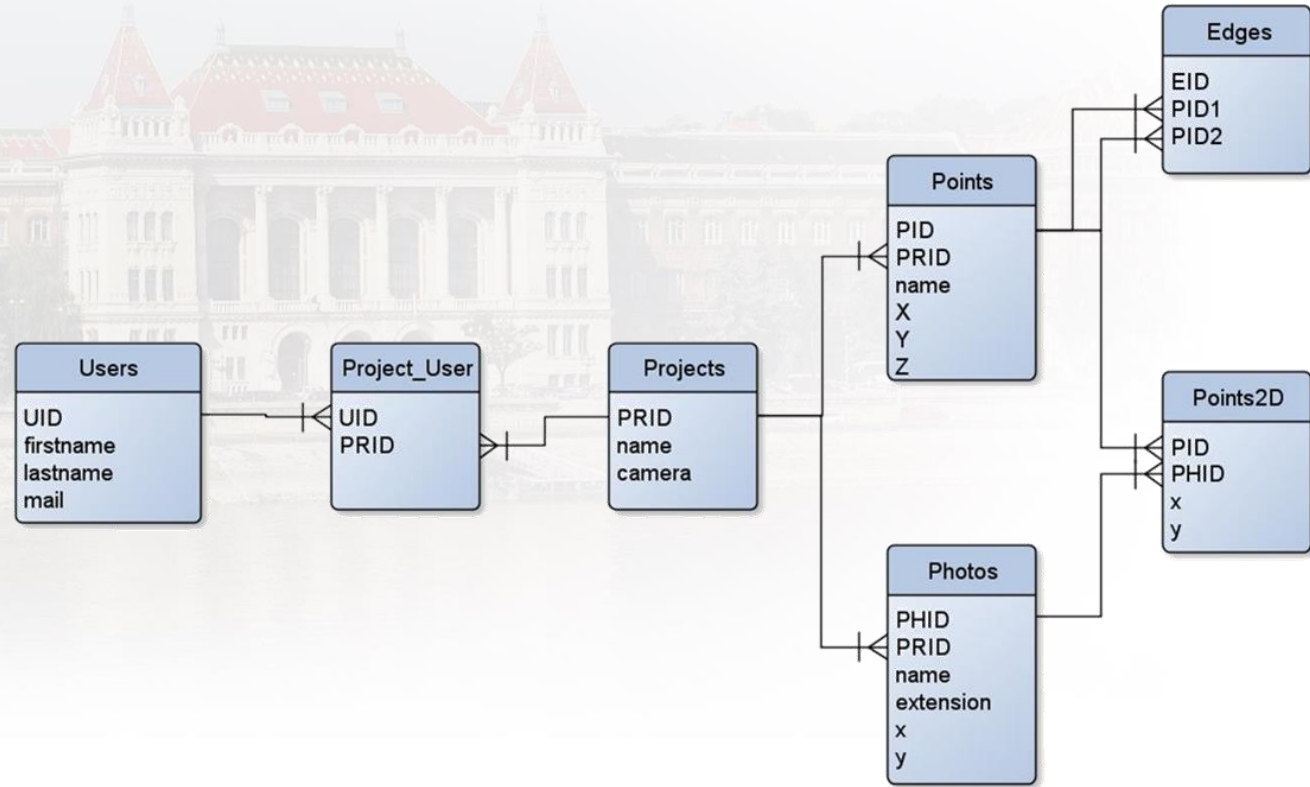
SPECIFICATION

- Input:
 - *User data*
 - *Images and image coordinates (2D)*
 - *Calculation results*
- Output:
 - *Number of points in a project*
 - *3D points in a project*
 - *Quality analysis of the project*

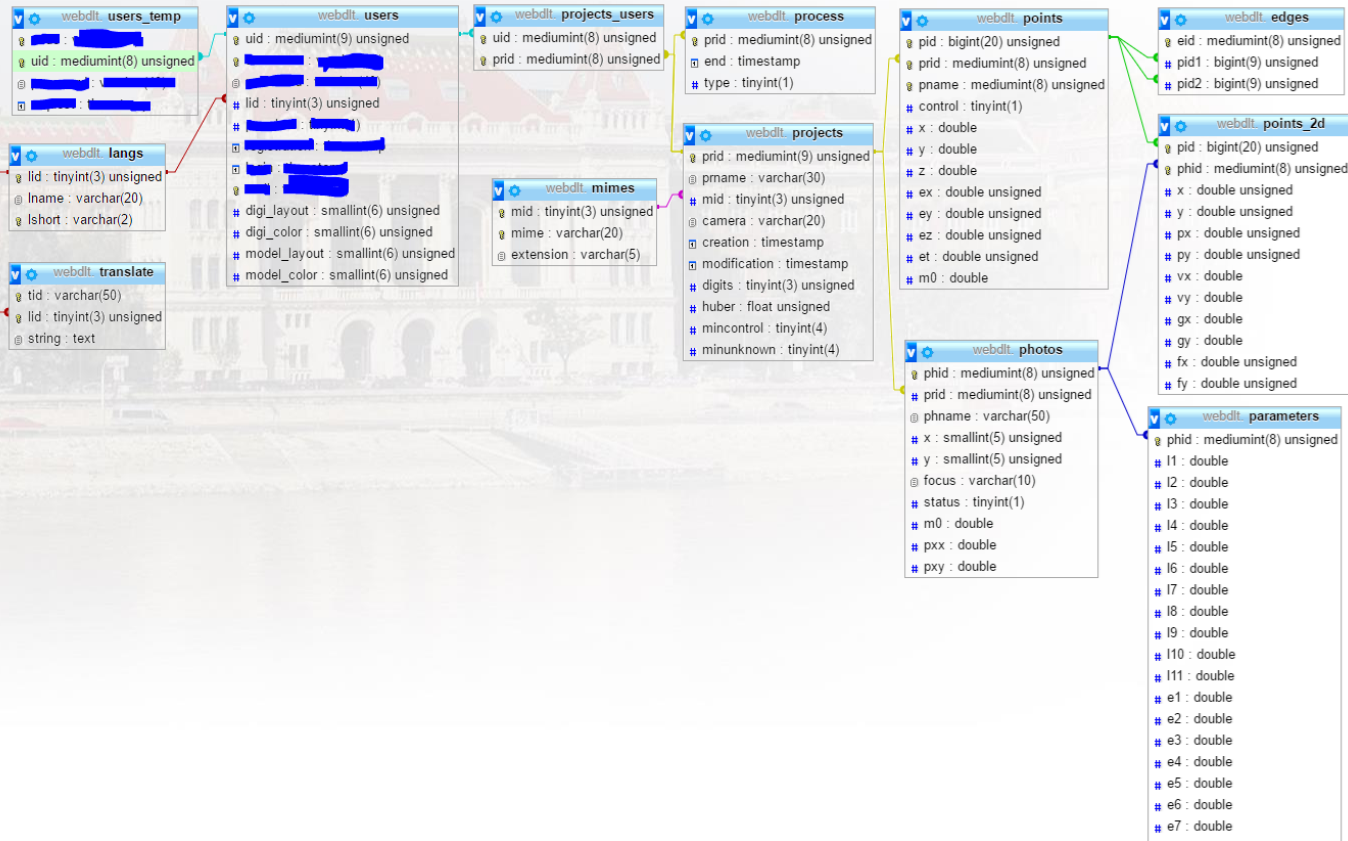
E/R DIAGRAM



SIMPLIFIED RELATIONAL SCHEMA DIAGRAM



FULL RELATIONAL SCHEMA DIAGRAM



phid	prid	phname	x	y	focus
1	1	DSC00164.JPG	3888	2592	160/10
2	1	DSC00838.JPG	3888	2592	160/10
3	1	DSC00840.JPG	3888	2592	160/10
4	1	DSC00841.JPG	3888	2592	160/10
5	2	L1060539.JPG	3136	2352	100/10
6	2	L1060540.JPG	3136	2352	100/10
7	2	L1060535.JPG	3136	2352	90/10
8	2	L1060538.JPG	3136	2352	90/10
9	2	L1060559.JPG	3136	2352	90/10
10	2	L1060785.JPG	3136	2352	100/10
11	2	L1060546.JPG	3136	2352	110/10

pid	pnid	x	y
1	1	39.0158	220.6818
1	2	462.796	115.4278
1	3	464.7067	162.783
1	4	306.1757	321.0843
2	1	739.3551	361.3168
2	2	1067.1689	382.3495
2	3	1100.3428	357.0253
2	4	893.4193	414.478
3	1	999.1262	113.7045
3	2	1530.98354253	105.641731486
3	3	1668.9564	216.7662
3	4	1113.3456	184.4889
4	2	2500.1023	355.4395
4	3	2568.3146	407.1178
4	4	2192.1705	215.9834
5	1	454.5572	588.0144
5	2	775.1694	536.793
5	3	818.9275	577.9722
5	4	655.5164	621.38

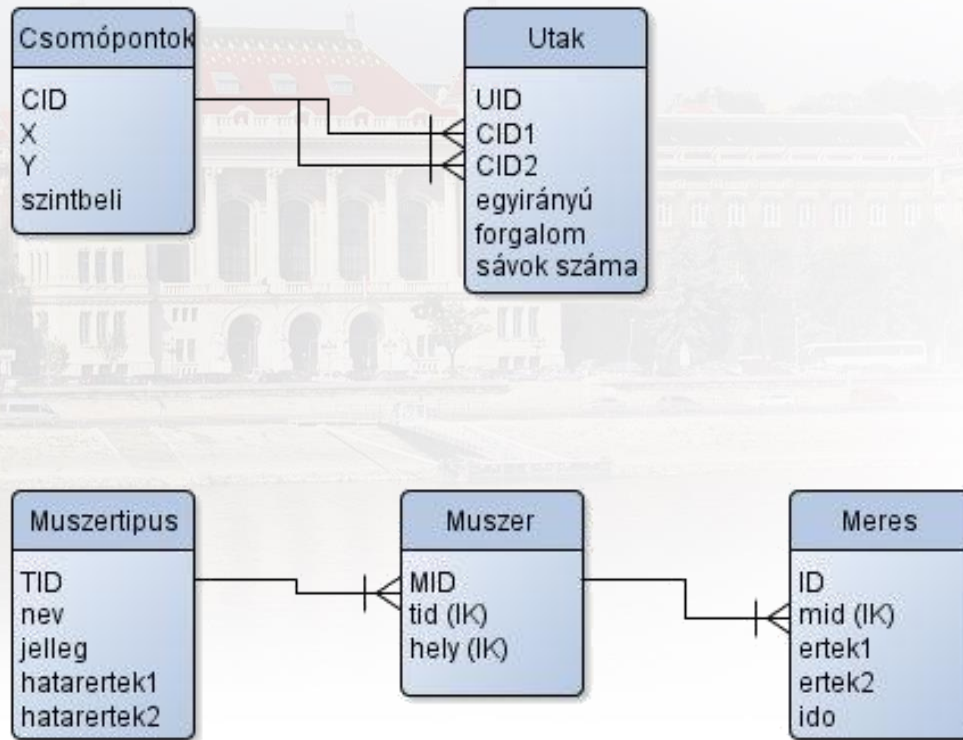
pid	prid	pname	control	x	y	z	ex	ey	ez
1	1	1	1	37.0928	270.9316	60.5645	NULL	NULL	NULL
2	1	2	1	96.6709	271.2785	19.8864	NULL	NULL	NULL
3	1	4	1	155.314	270.4152	70.7968	NULL	NULL	NULL
4	1	7	1	276.3766	271.4792	40.1353	NULL	NULL	NULL
5	1	8	1	66.573	241.3614	25.873	NULL	NULL	NULL
6	1	10	1	276.8236	241.776	50.2347	NULL	NULL	NULL
7	1	12	1	156.7464	211.5336	18.9318	NULL	NULL	NULL
8	1	14	1	97.5291	182.0065	34.018	NULL	NULL	NULL
9	1	17	1	396.9045	151.6384	25.402	NULL	NULL	NULL
10	1	20	1	337.1254	62.9023	45.9568	NULL	NULL	NULL
11	1	22	1	216.322	33.2538	80.7864	NULL	NULL	NULL
12	2	1	1	377.0196	451.5317	196.171	NULL	NULL	NULL
13	2	2	1	377.0205	451.4743	205.2246	NULL	NULL	NULL
14	2	3	0	368.76	448.67	205.2	0.012464	0.010419	0.0086125
15	2	4	0	368.7	448.82	196.11	0.019143	0.050543	0.017216
16	2	6	0	368.07	448.41	205.21	0.010043	0.0086532	0.0070856
17	2	7	0	367.79	448.32	205.2	0.010297	0.0089777	0.0073242
18	2	8	0	367.73	448.37	196.07	0.038841	0.015537	0.0089256
19	2	9	0	368.01	448.45	196.08	0.024589	0.010961	0.00715
20	2	10	1	367.1206	448.0557	205.2165	NULL	NULL	NULL
21	2	11	0	360.99	445.94	205.2	0.023323	0.020619	0.018001
22	2	12	0	361	445.95	196.06	0.01842	0.014727	0.0078193
23	2	13	1	367.0897	448.104	196.1022	NULL	NULL	NULL



QUERIES

- Which projects are accessible for user with ID 10?
 - *SELECT prid, pname FROM projects NATURAL JOIN project_users WHERE uid=10;*
- Is point 100 marked on photo 200?
 - *SELECT points_2d.pid FROM points_2d NATURAL JOIN points WHERE points_2d.pid=100 AND phid=200;*
- What is the starting point of edge with ID 100?
 - *SELECT * FROM edges JOIN points ON edges.pid1=points.pid WHERE eid=100;*

IDEAS



CONCLUSION

- Homework experiences
- Case study
 - *Online statistics*
- Privacy and security
- Case study II.
 - *WebDLT*



Köszönöm a figyelmet!

Kérdések?



BUDAPESTI MŰSZAKI
ÉS GAZDASÁGTUDOMÁNYI EGYETEM

Építőmérnöki Kar - építőmérnöki képzés 1782 óta

Fotogrammetria és Térinformatika Tanszék