KATH: A No-Coding Data Processing Aid for Genetic Researchers

Kaunas University of Technology:

Dr. Kazimieras Bagdonas
Ignas Sabaliauskas
Dainius Kirsnauskas
Gabrielius Salyga
Kajus Černiauskas
Mantvydas Deltuva
Nojus Sajauskas
Paulius Preikša
Vladyslav Levchenko

Massachusetts Eye and Ear, Harvard University (Genetics), USA

Dr. Kinga M. Bujakowska
Dr. Eglė Galdikaitė-Brazienė

UAB Genomika

Dr. Lukas Žemaitis
Ignas Galminas

University of Pisa

Dr. Daniele Mazzei
Dr. Roberto Figliè
Dr. Giuseppe Prencipe

The problem

The following issues have been identified in discussions with researchers from Harvard University:

1. Complexity of the DNA analysis software tools;
2. Lack of programming skills among scientists, even world-class scientists manually refactor and analyze data;

Programmers lack domain knowledge in genetics; severe lack of bioinformaticians;

The KATH – A No-Coding Data Processing Aid for Genetic Researchers project seeks to develop a cutting-edge platform to revolutionize the work of genetics researchers by providing an integrated, automated, and AI-driven toolset.

The system will consist of three primary components: a database of pathogenic DNA mutations, an integration system for DNA analysis tools, and an AI assistant for natural language processing and command generation.

The database will aggregate and standardize data from publicly available sources like LOVD, ClinVar, and gnomAD. The integration system will connect this data to various DNA analysis tools, such as CADD, REVEL, and SpliceAI, allowing researchers to perform complex analyses and visualize the results through a user-friendly interface. The AI assistant, powered by a Large Language Model (LLM), will enable users to interact with the system through natural language, making advanced genetic analysis more accessible and efficient.

Current Status

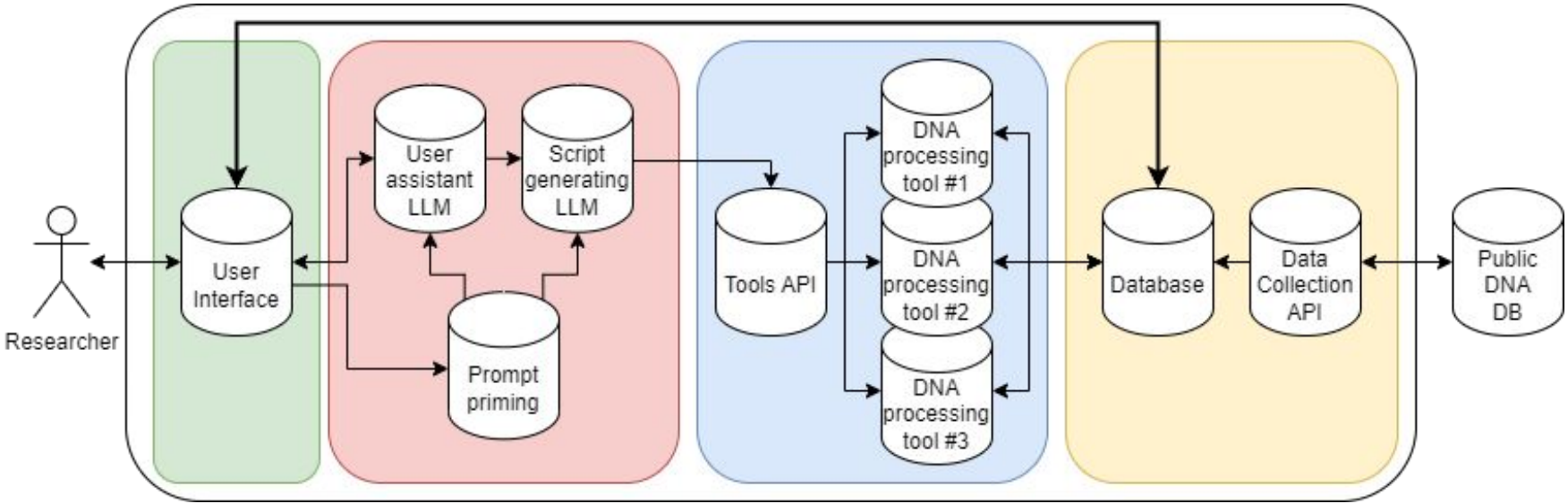
Since December 2023, a prototype of the system has been developed by a team of undergraduate students at KTU. This prototype can gather and refactor DNA data, integrate several DNA analysis tools, and provide a basic web-based UI for result visualization.

Initial AI assistant tests using public LLM models have been partially successful, and the prototype has received positive recognition at academic and innovation conferences.

On September 2024 a prototype version of the KATH system has been delivered to HU researchers at Massachusetts Eye and Ear, Harvard Medical School for testing, evaluation and feedback.

The solution

KATH: A No-Coding Data Processing Aid for Genetic Researchers

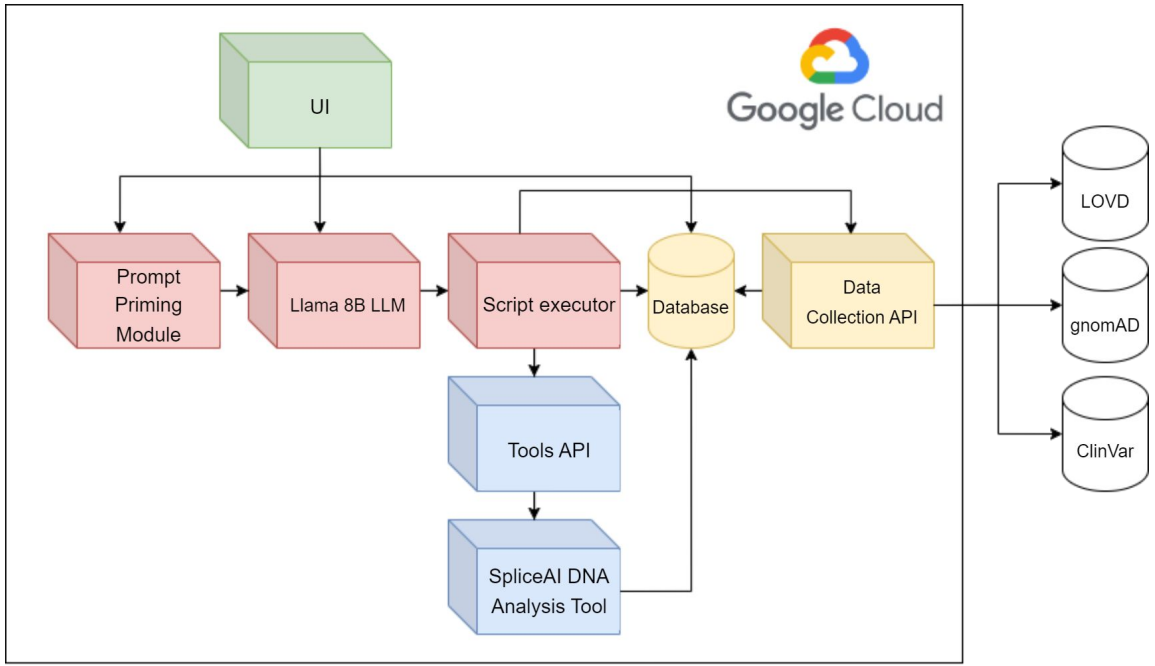


TRL 4 Proof of Concept Implementation

Analysis and R&D has been performed with the guidance of Harvard researchers;

TRL 4 prototype has been implemented as a cloud-based solution;

Results have been published and presented at the international scientific student conference IVUS 2024 and Technorama 2024.



Project budget

	KATH budget				Requirements
Funding level	**Partner**	**Financing, €**	**Budget, €**	**Budget share**	
100%	KTU	€150,000.00	€49,999.00	17%	
66.67%	UAB Genomika	€50,000.00	€150,001.00	**50%**	>= 50%
100%	UNIPI	€100,000.00	€100,000.00	33%	
LT	300,000	Budget LT, €	€200,000.00	**67%**	
IT	400,000	Budget IT, €	€100,000.00	**33%**	>=30%
	700,000	**Total, €**	**€300,000.00**		

Project scope

Project Objectives:

1. Develop a Comprehensive Genetic Database Module (*UAB Genomika & KTU*)
2. Create an Advanced Integration System (*UAB Genomika & KTU*)
3. Develop an AI Assistant for Natural Language Processing (*University of Pisa*)
4. Deliver a Minimum Viable Product (MVP) (*UAB Genomika*)

Work Packages

WP1: System Development (Database, Integration System, AI Assistant)

WP2: Testing and Validation

WP3: Commercialization Strategy and Market Entry

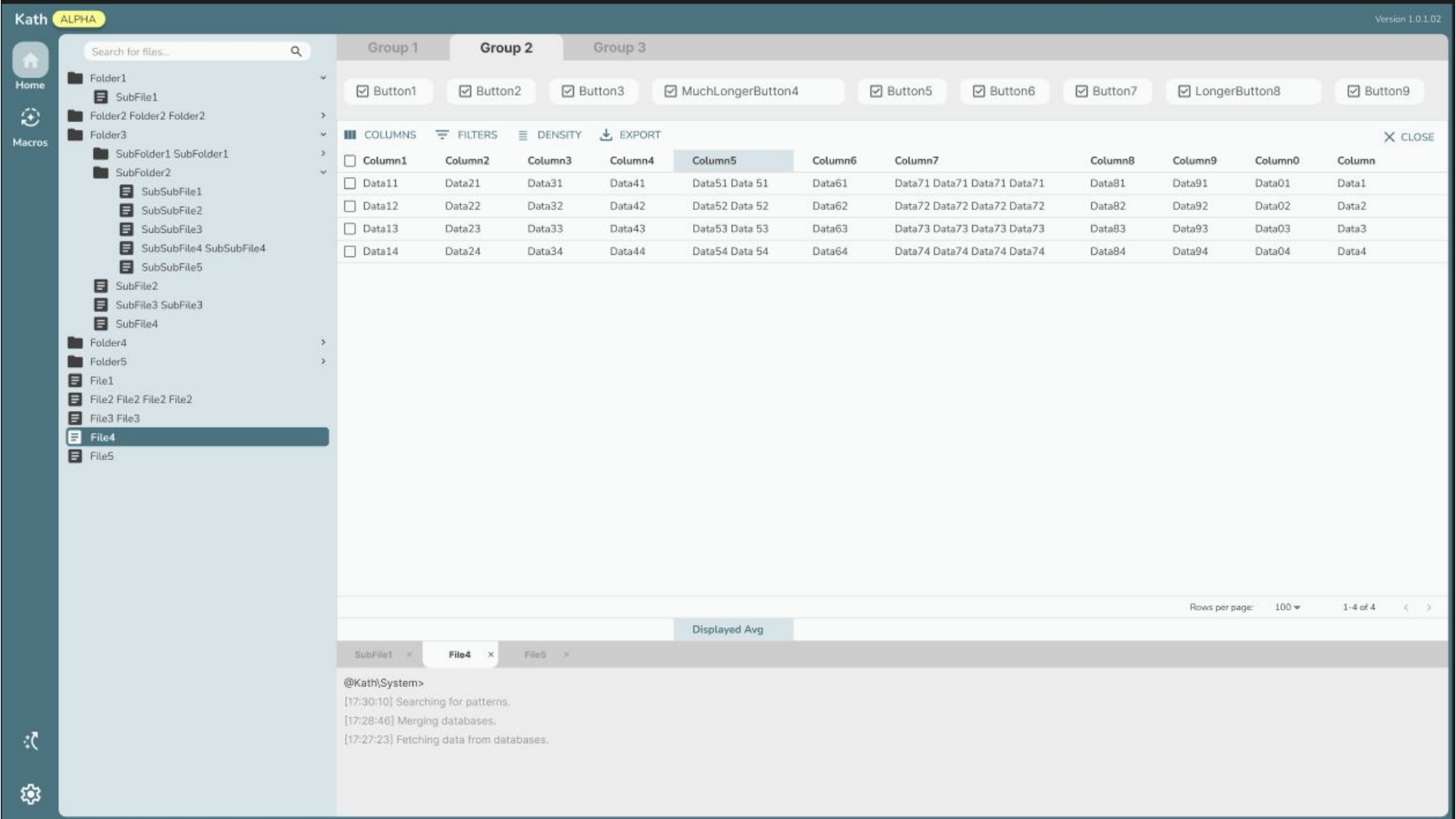
WP4: Project Management and Coordination

2024 Q3 Deliverables

KATH *alpha* version:

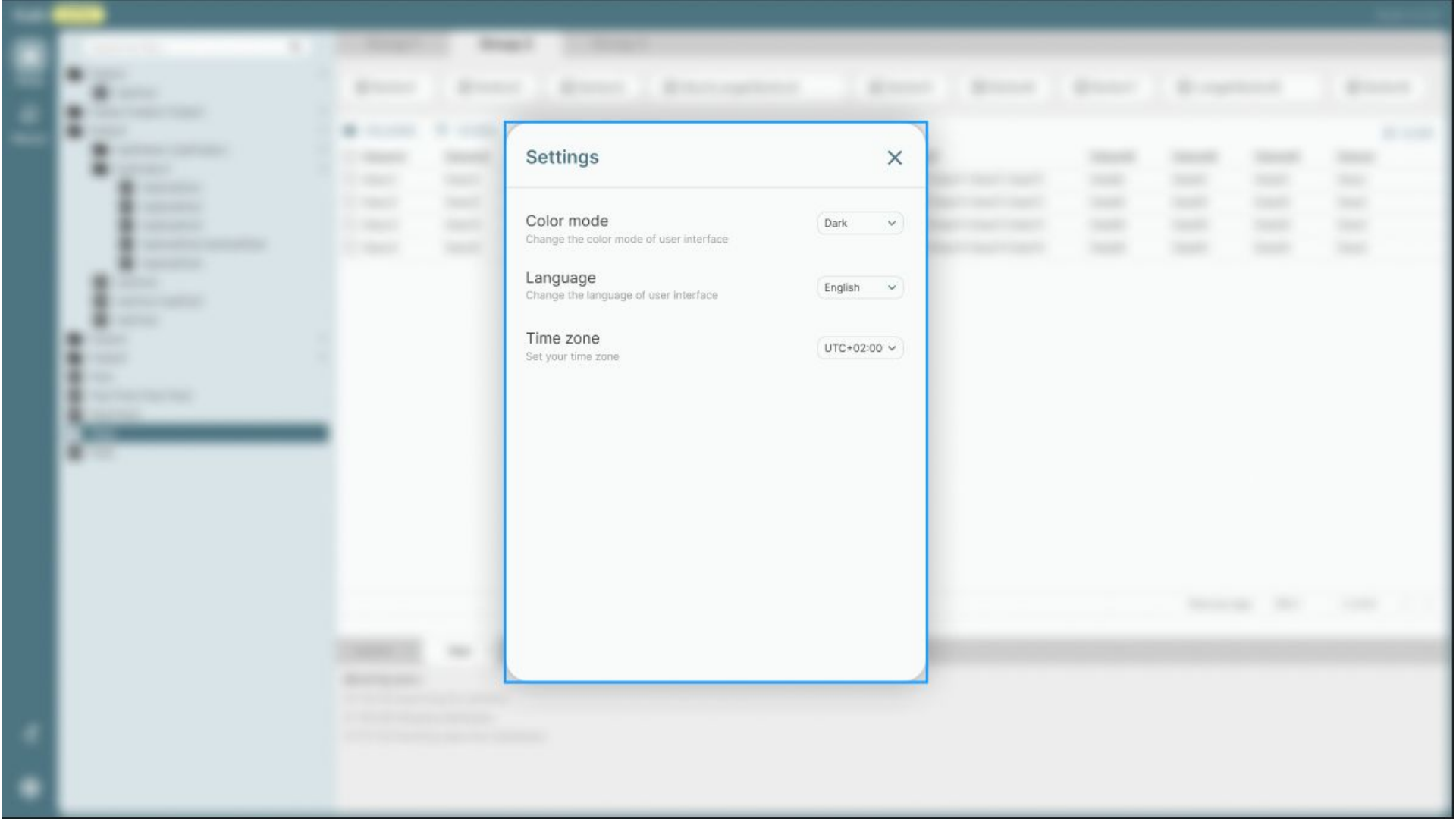
System architecture: https://miro.com/app/board/uXjVKlU1KV8=/

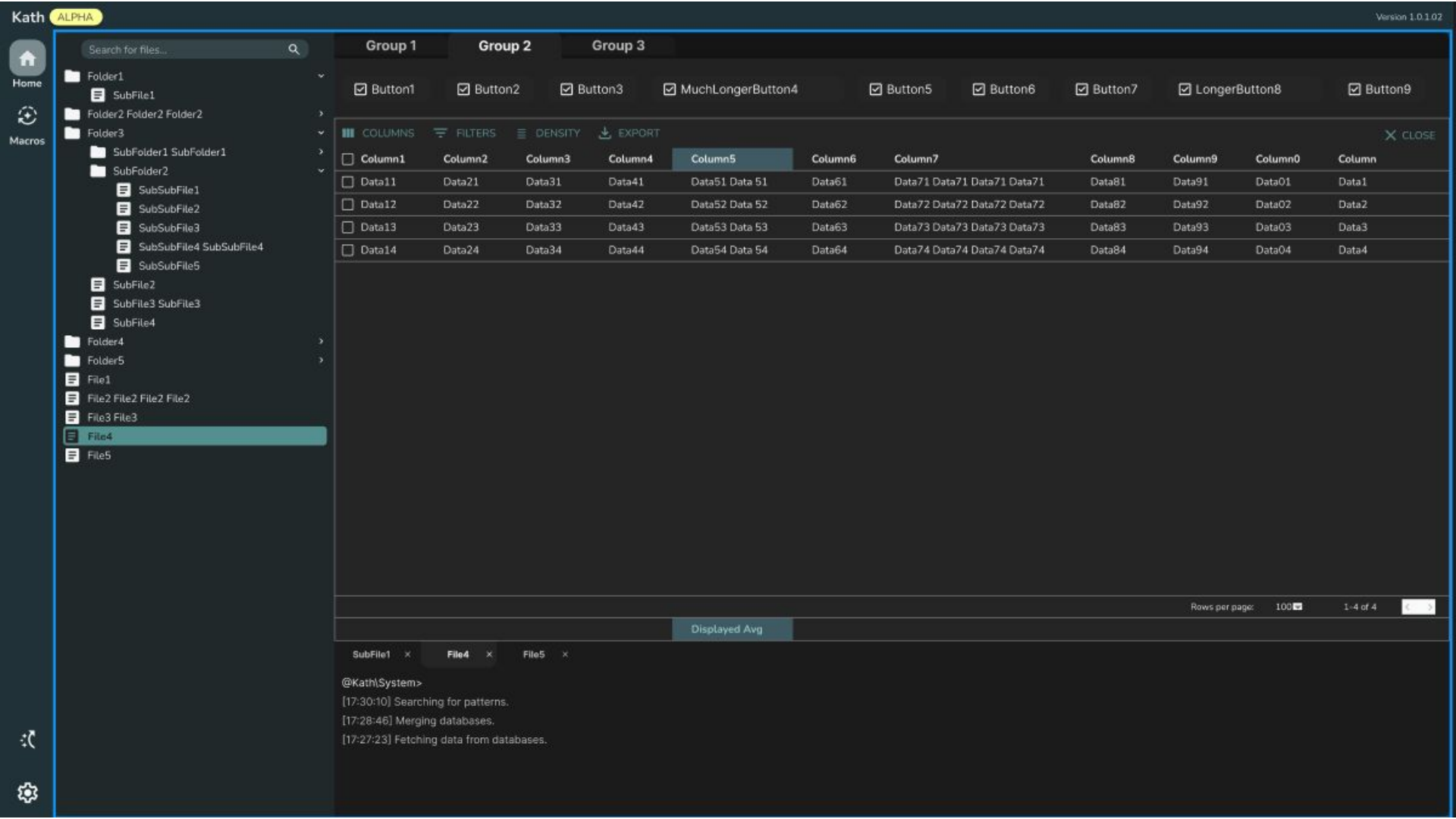
System UI: https://www.figma.com/design/UjO2G4w318OyNBuNI4Ox25/project-kath?node-id=0-1&t=PrnxLSUw8bYSJAzj-1


Kath
ALPHA
Version 1.0.1.02
Home
Macros
Search for files...
Folder1
SubFile1
Folder2 Folder2 Folder2
Folder3
SubFolder1 SubFolder1
SubFolder2
SubSubFile1
SubSubFile2
SubSubFile3
SubSubFile4 SubSubFile4
SubSubFile5
SubFile2
SubFile3 SubFile3
SubFile4
Folder4
Folder5
File1
File2 File2 File2 File2
File3 File3
File4
File5
Group 1
Group 2
Group 3
Button1
Button2
Button3
MuchLongerButton4
Button5
Button6
Button7
LongerButton8
Button9
COLUMNS
FILTERS
DENSITY
EXPORT
CLOSE
Column1 Column2 Column3 Column4 Column5 Column6 Column7 Column8 Column9 Column0 Column
Data11 Data21 Data31 Data41 Data51 Data 51 Data61 Data71 Data71 Data71 Data71 Data81 Data91 Data01 Data1
Data12 Data22 Data32 Data42 Data52 Data 52 Data62 Data72 Data72 Data72 Data72 Data82 Data92 Data02 Data2
Data13 Data23 Data33 Data43 Data53 Data 53 Data63 Data73 Data73 Data73 Data73 Data83 Data93 Data03 Data3
Data14 Data24 Data34 Data44 Data54 Data 54 Data64 Data74 Data74 Data74 Data74 Data84 Data94 Data04 Data4
Rows per page: 100
1-4 of 4
Displayed Avg
SubFile1
File4
File5
@Kath\System>
[17:30:10] Searching for patterns.
[17:28:46] Merging databases.
[17:27:23] Fetching data from databases.

Shortcuts

Shortcut Description	Windows	Mac
Select the entire page	Ctrl + A	Cmd + A
Copy selected information	Ctrl + C	Cmd + C
Cut selected information	Ctrl + X	Cmd + X
Paste copied information	Ctrl + V	Cmd + V
Open a file	Ctrl + O	Cmd + O
Save a file	Ctrl + S	Cmd + S
Undo the last operation	Ctrl + Z	Cmd + Z
Redo the last operation	Ctrl + Y	Shift + Cmd + Z
Find text in the current document/page	Ctrl + F	Cmd + F
Task Manager	Ctrl + Shift + Esc	Alt + Cmd + Esc
Application Switcher	Alt + Tab	Cmd + Tab
Run an Application	Win + R	Cmd + Space
Minimize the current window	Win + ↓	Cmd + M
Minimize all windows	Win + M	Alt + Cmd + M
Jump to the application's menubar	F10	Shift + Cmd + /
Take a screenshot	Print Screen	Shift + Cmd + 3
Open a new browser tab	Ctrl + T	Cmd + T
Close the current browser tab	Ctrl + W	Cmd + W
Open a previously closed tab	Ctrl + Shift + T	Shift + Cmd + T
Focus and select web browser's address bar	Ctrl + L	Cmd + L


Settings
Color mode
Change the color mode of user interface
Dark
Language
Change the language of user interface
English
Time zone
Set your time zone
UTC+02:00


Kath
ALPHA
Version 1.0.1.02
Home
Macros
Search for files...
Folder1
SubFile1
Folder2 Folder2 Folder2
Folder3
SubFolder1 SubFolder1
SubFolder2
SubSubFile1
SubSubFile2
SubSubFile3
SubSubFile4 SubSubFile4
SubSubFile5
SubFile2
SubFile3 SubFile3
SubFile4
Folder4
Folder5
File1
File2 File2 File2 File2
File3 File3
File4
File5
Group 1
Group 2
Group 3
Button1
Button2
Button3
MuchLongerButton4
Button5
Button6
Button7
LongerButton8
Button9
COLUMNS
FILTERS
DENSITY
EXPORT
CLOSE
Column1 Column2 Column3 Column4 Column5 Column6 Column7 Column8 Column9 Column0 Column
Data11 Data21 Data31 Data41 Data51 Data 51 Data61 Data71 Data71 Data71 Data71 Data81 Data91 Data01 Data1
Data12 Data22 Data32 Data42 Data52 Data 52 Data62 Data72 Data72 Data72 Data72 Data82 Data92 Data02 Data2
Data13 Data23 Data33 Data43 Data53 Data 53 Data63 Data73 Data73 Data73 Data73 Data83 Data93 Data03 Data3
Data14 Data24 Data34 Data44 Data54 Data 54 Data64 Data74 Data74 Data74 Data74 Data84 Data94 Data04 Data4
Rows per page: 100
1-4 of 4
Displayed Avg
SubFile1
File4
File5
@Kath\System>
[17:30:10] Searching for patterns.
[17:28:46] Merging databases.
[17:27:23] Fetching data from databases.

Roadmap

1. Students will continue the development of the KATH system through September - December 2024
2. At least one B.Sc. final project will address issues of KATH
3. Researchers at HU will perform validation and testing of the KATH *alpha* version and provide feedback over September - December 2024
4. If granted, funding will start at Q1 2025
5. MVP development and validation will take pace during 2025