

SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE

Fakulta elektrotechniky a informatiky

Tímový projekt

Projektová dokumentácia

SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE

Fakulta elektrotechniky a informatiky

Tímový projekt

Projektová dokumentácia

Študijný program: Aplikovaná informatika
Predmet: TP1, TP2 – Tímový projekt 1, Tímový projekt 2
Vedúci tímového projektu: Ing. Stanislav Marochok

2025

**Bc. Michal Balogh
Bc. Juraj Hušek
Bc. Ján Osadský
Bc. Fridrich Molnár
Bc. Zoltán Renczes**



ZADANIE TÍMOVÉHO PROJEKTU

Vedúci tímového projektu: Ing. Stanislav Marochok

Názov práce: **Collaborative Platform for Digitisation, Annotation, and Decryption of Historical Ciphers**

Jazyk, v ktorom sa práca vypracuje: slovenský jazyk

Špecifikácia zadania:

The aim of this project is to design and implement a full-stack web application for the processing of historical handwritten and printed encrypted texts together with handwritten cipher key documents. These sources include cipher keys, fully encrypted texts, semi-encrypted texts, and mixed pages where both keys and texts appear together. The application should provide researchers and students with modern tools for digital preservation, pre-processing, annotation, recognition, reconstruction, symbol mapping, and decryption, all within a single collaborative platform.

The system is expected to cover the entire workflow from uploading digitized images to creating machine-readable editions that can be directly applied to decipher encrypted texts. It should allow multi-level annotation of regions, entries, and symbols, enabling users to highlight areas of interest, classify symbols, reconstruct substitution tables, and link plaintext with ciphertext. To support this functionality, the processing pipeline will combine pre-processing methods such as normalization, binarization, deskewing, segmentation, and noise reduction with both AI-based and heuristic approaches. The AI modules should include handwriting recognition, symbol detection, and table reconstruction, while the non-AI modules should rely on contour analysis, morphological operations, and rule-based reconstruction. A unified symbol mapping system will normalize all detected objects into a shared database of codes, ensuring that digitalized cipher keys can be applied seamlessly to digitalized encrypted texts.

Beyond the technical pipeline, the platform will also include user management with different roles, shared workspaces for projects, and document sharing features. A credit and attribution system will ensure that contributions such as annotations, reconstructions, and decryptions are transparently recorded. Users will also interact with a system of virtual points that they earn by contributing work on the platform and can later spend on advanced actions such as launching automatic symbol detection or AI-based table reconstruction. The decryption module will connect the different parts of the system by allowing reconstructed cipher keys to be applied directly to encrypted passages with support for frequency analysis, automatic suggestions, and manual corrections.

The final outcome of the project will be a cloud-deployed, production-ready application that integrates pre-processing, annotation, recognition, reconstruction, decryption, and collaborative features, including user management, document sharing, credits, and virtual points. It will also include both AI and non-AI methods, a demonstrator dataset of annotated pages, and comprehensive documentation.

1. Study the domain – research encrypted texts, cipher keys, AI and non-AI methods, and preprocessing techniques.
2. Design the system – define requirements, architecture, user management, sharing, credits, and virtual points.
3. Plan the project – prepare timeline, milestones, assign roles, set up tools.
4. Model the system – create UML diagrams and database schemas.
5. Implement the application – build backend, frontend, preprocessing, AI and non-AI modules, reconstruction, user management, sharing, credits, and virtual points.
6. Test the system – perform unit, integration, and system tests; compare AI and non-AI methods; evaluate reconstruction and decryption on synthetic and historical data.
7. Write documentation – prepare user guides, technical documentation, and the final report.

Abstrakt

Cieľom projektu bol návrh a implementácia plnohodnotnej full-stack webovej aplikácie určenej na spracovanie historických šifrovaných textov a dokumentov obsahujúcich šifrové kľúče. Aplikácia pokrývala celý pracovný postup od nahratia digitalizovaných obrazov cez predspracovanie, anotáciu, rozpoznávanie a rekonštrukciu až po samotné dešifrovanie. Navrhnutý systém spájal metódy umelej inteligencie (rozpoznávanie rukopisu, detekcia symbolov, rekonštrukcia tabuliek) s heuristickými prístupmi (morfologické operácie, kontúrová analýza, pravidlové algoritmy) a poskytoval používateľom jednotnú databázu symbolov pre mapovanie a aplikáciu šifrových kľúčov. Platforma zároveň podporovala kolaboráciu, správu používateľov s rôznymi rolami, zdieľanie dokumentov a systém kreditov a bodov za príspevky. Výsledkom projektu bola nasadená cloudová aplikácia, ktorá poskytla moderné nástroje pre výskum, digitalizáciu a dešifrovanie historických šifrovaných materiálov a zároveň vytvorila priestor na spoločnú prácu výskumníkov a študentov.

Kľúčové slová

historické šifry, spracovanie obrazu, digitálna anotácia, kolaboratívna platforma, webová aplikácia

Abstract

The aim of the project was to design and implement a comprehensive full-stack web application for processing historical encrypted texts and documents containing cipher keys. The application covered the entire workflow from uploading digitized images through preprocessing, annotation, recognition, and reconstruction up to the actual decryption. The designed system combined artificial intelligence methods (handwriting recognition, symbol detection, table reconstruction) with heuristic approaches (morphological operations, contour analysis, rule-based algorithms) and provided users with a unified symbol database for mapping and applying cipher keys. The platform also supported collaboration, user management with different roles, document sharing, and a system of credits and points for contributions. The outcome of the project was a cloud-deployed application that offered modern tools for research, digitization, and decryption of historical encrypted materials, while also creating a collaborative environment for researchers and students.

Keywords

historical ciphers, image processing, digital annotation, collaborative platform, web application

Obsah

Úvod	10
1 Zázpisnice zo stretnutí	11
1.1 Zimný semester 2025/26	11
1.1.1 Úvodné stretnutie tímu – 25. septembra 2025	11
Záver	14

Zoznam značiek a skratiek

Zoznam algoritmov

Zoznam výpisov kódov

Úvod

1 Zápisnice zo stretnutí

1.1 Zimný semester 2025/26

1.1.1 Úvodné stretnutie tímu – 25. septembra 2025

Dátum: 25. septembra 2025

Čas: 12:00 – 13:00

Umiestnenie: Slovenská technická univerzita v Bratislave, Fakulta elektrotechniky a informatiky

Miestnosť: A208, FEI STU

Zapisovateľ: Bc. Juraj Hušek

Účastníci:

- Bc. Michal Balogh – ML/AI Engineer
- Bc. Juraj Hušek – Backend Developer
- Bc. Ján Osadský – Backend Developer
- Bc. Fridrich Molnár – Frontend Developer
- Bc. Zoltán Renczes – Frontend Developer

Vedúci tímového projektu: Ing. Stanislav Marochok

Program stretnutia:

1. Oboznámenie sa s projektom
2. Rozdelenie rolí v tíme
3. Organizačné pokyny
4. Diskusia o technológiách a funkcionalitách
5. Plánovanie ďalších krokov

Diskusia a rozhodnutia:

1. Oboznámenie sa s projektom Vedúci projektu podrobne predstavil zadanie. Cieľom je vyvinúť plnohodnotnú webovú aplikáciu, ktorá umožní digitalizáciu, anotáciu, rozpoznávanie a dešifrovanie historických šifrovaných textov. Diskutovalo sa o výskumnom potenciáli aj praktickom prínose riešenia. Tím sa zhodol na potrebe modulárneho prístupu a postupného zavádzania funkcionalít.

2. Rozdelenie rolí Tím sa dohodol na nasledovnom rozdelení úloh:

- ML/AI Engineer: Bc. Michal Balogh
- Backend Developers: Bc. Juraj Hušek, Bc. Ján Osadský
- Frontend Developers: Bc. Fridrich Molnár, Bc. Zoltán Renczes

Rozdelenie vychádza z predchádzajúcich skúseností členov a zodpovedá potrebám projektu.

3. Organizačné pokyny Stretnutia budú pravidelne každý štvrtok o 12:00 v miestnosti A208. GitHub bude využívaný ako spoločný repozitár zdrojového kódu webovej aplikácie, webovej stránky projektu a AI experimentov. Komunikácia ohľadom projektu bude prebiehať prezenčne a na platforme Discord. Tím sa zhodol, že je potrebné zaviesť pravidelné konzultácie s vedúcim tímového projektu, aby sa predišlo nejasnostiam a priebežne sa riešili otázky, problémy a možné ďalšie funkcionality.

4. Diskusia o technológiách a funkcionalitách Diskutovalo sa o základnom návrhu architektúry a výbere technológií:

- Backend: Python FastAPI alebo Flask
- Frontend: React
- Databáza: PostgreSQL alebo MySQL
- AI/ML: Python (OpenCV, TensorFlow alebo PyTorch)
- Deployment: Docker a Raspberry Pi 5 pre webovú aplikáciu, GitHub Pages pre informačnú stránku projektu

Diskusia o úvodnej funkcionalite: Diskutovali sme o návrhu základných krokov úvodnej funkcionality aplikácie, ktorá by mala umožniť používateľovi nahrať dokument do systému. Ide o pracovný návrh, ktorý nie je finálny a bude ešte spresňovaný v nasledujúcich konzultáciách:

1. Nahratie dokumentu a klasifikácia – systém rozpozná, či ide o textový dokument. Ak dokument nebude textového charakteru, spracovanie sa zastaví.
2. Preprocessing – plánované operácie: deskewing, binarizácia a odstránenie šumu, pričom používateľ bude mať možnosť nastavovať parametre.
3. Detekcia obsahu – identifikácia tabuliek, textových oblastí alebo šifrovaných úsekov.

4. Priradenie symbolov – pokus o mapovanie šifrovaných častí na znaky uložené v databáze a ponuka možných riešení používateľovi.
5. Možnosť manuálnych zásahov – každý krok bude realizovateľný aj manuálne, neurónové siete budú slúžiť len ako asistent.

5. Úlohy do nasledujúceho stretnutia

- Všetci členovia: detailne si preštudovať zadanie projektu.
- Bc. Michal Balogh: založiť GitHub projekt, vytvoriť príslušné repozitáre a pripraviť základnú verziu webovej stránky projektu.
- Bc. Juraj Hušek, Bc. Michal Balogh, Bc. Ján Osadský: vypracovať a navrhnuť use case diagramy webovej aplikácie.
- Bc. Fridrich Molnár a Bc. Zoltán Renczes: preskúmať možnosti a konfiguračné kroky potrebné na nasadenie aplikácie na Raspberry Pi server.

Nasledujúce stretnutie:

Dátum: 2. októbra 2025

Čas: 12:00

Umiestnenie: Slovenská technická univerzita v Bratislave, Fakulta elektrotechniky a informatiky

Miestnosť: A208, FEI STU

Poznámky:

- Projekt má potenciál priniesť významné výsledky z pohľadu výskumu aj praktického využitia.
- Tím sa zhodol, že na jeho realizáciu je vhodné postupovať agilným spôsobom, aby bolo možné priebežne reagovať na nové požiadavky a zistenia.
- Zdôraznená bola potreba pravidelných konzultácií s vedúcim tímového projektu, ktoré umožnia včasnú identifikáciu a riešenie problémov a ďalších funkcionalít systému.

Záver