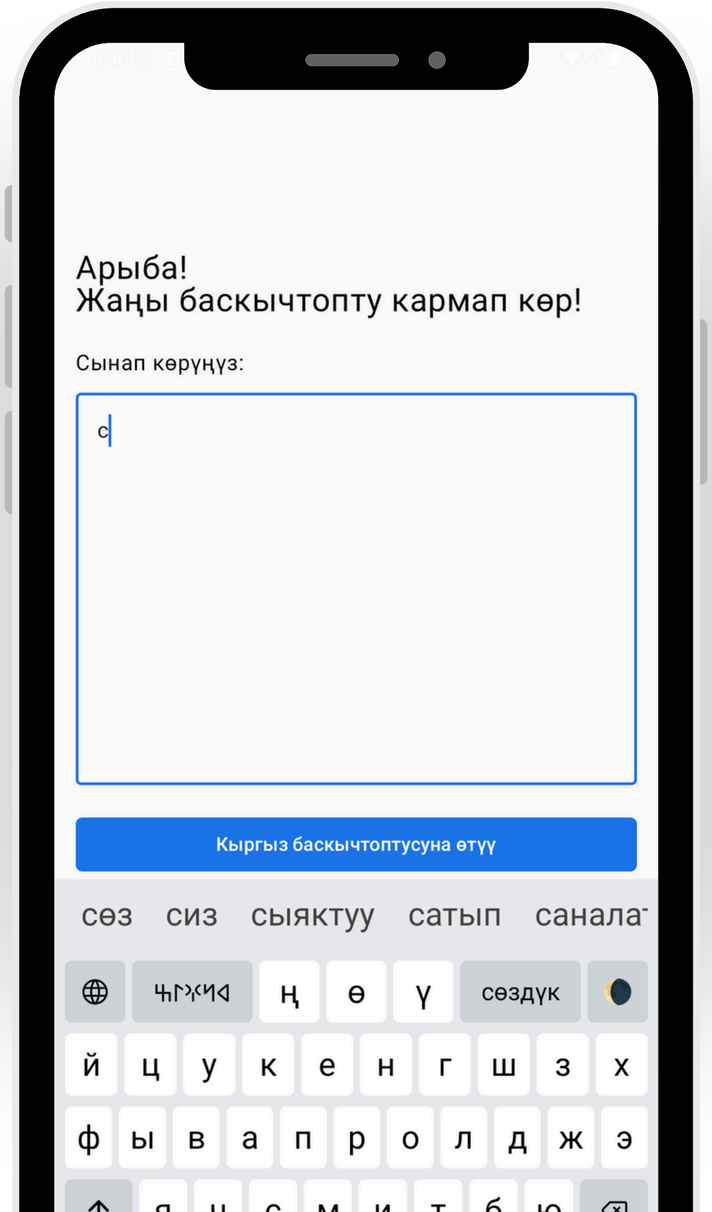
Kyrgyz Keyboard

Team Members:

Ademi Abdykerimova
Nikita Paniukhin
Pavel Boubel

Mentor:

Ivan Komarov



Domain Introduction

What is the domain?

Minority-language tools for underserved communities

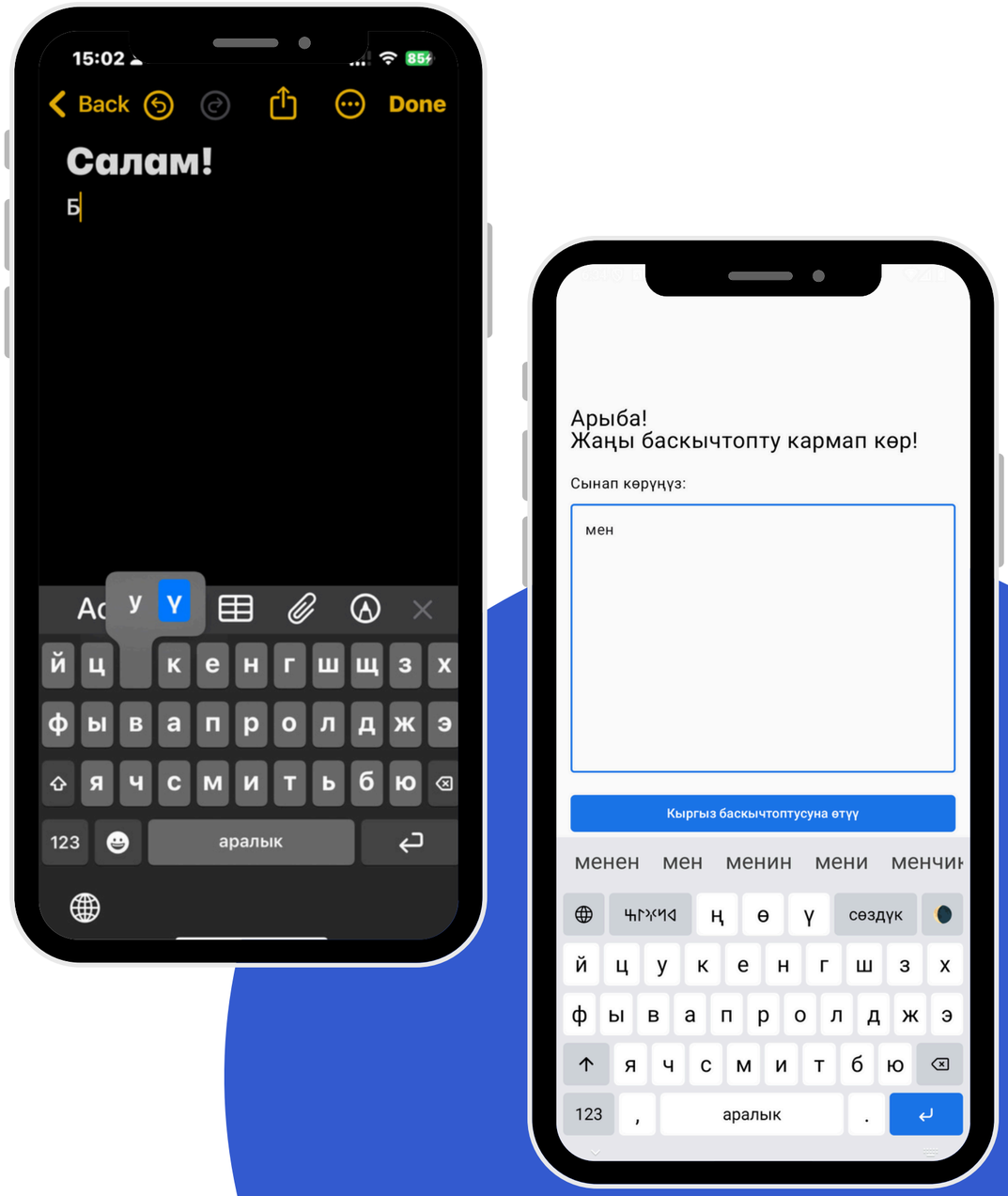
Why is this important?

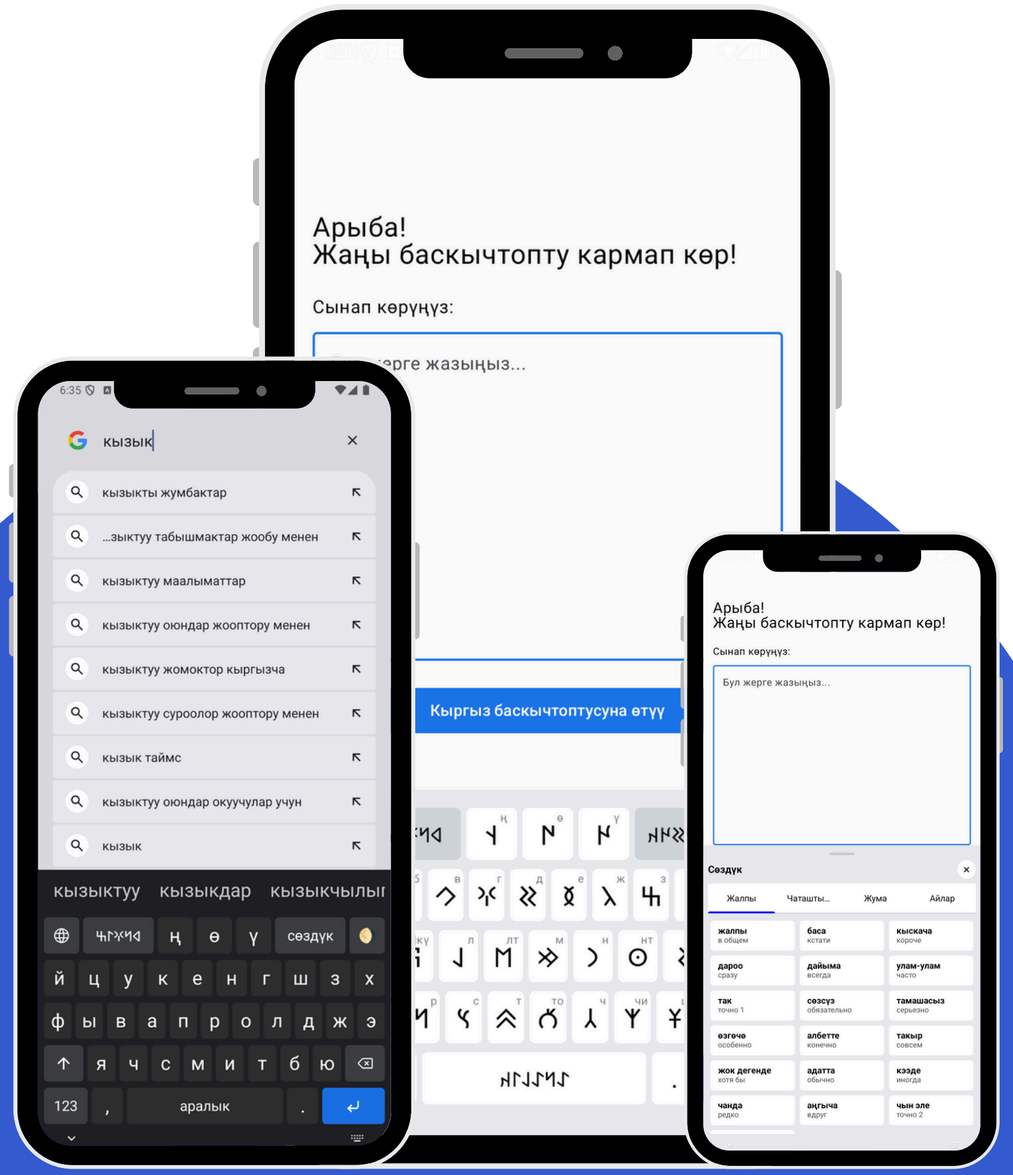
Existing Kyrgyz keyboards lack optimized layouts, predictive text, or script-switching (Cyrillic/Latin)

Who benefits?

Empowers students, elders, professionals to type naturally.
Preserves the Kyrgyz language in the digital space

First open-source keyboard tailored for Kyrgyz





Welcome To Our Application

Bridging digital gaps for Kyrgyz speakers

Optimized Layout

All Kyrgyz letters (ң, ө, ү) accessible without long-press — type faster

Word Predictions

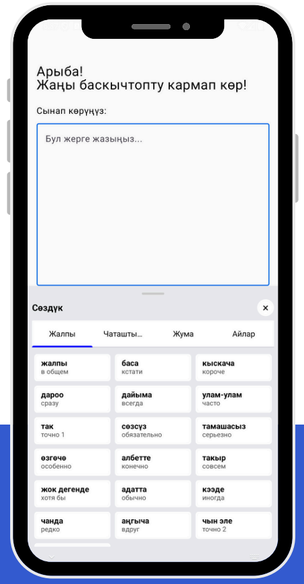
Reduces typos

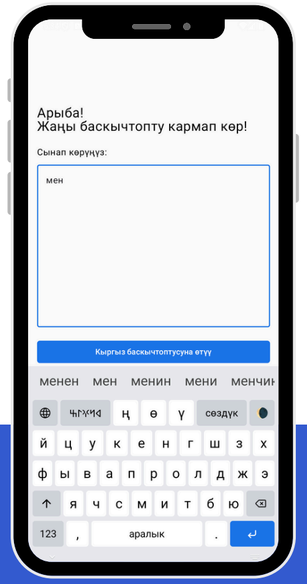
Real-time fixes for agglutinative word forms

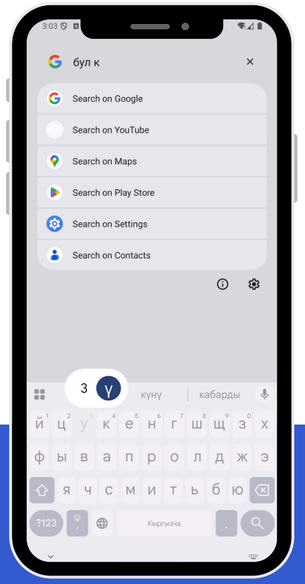
User-Centric Design

Extra layouts and symbols to encourage engagement

Analogs & Comparison







Features	Our Keyboard	Gboard (Google)	iOS Keyboard	Samsung Keyboard
Kyrgyz Layout	✓ ң ө ү	Long-press	Long-press	Long-press
Predictive Text for Kyrgyz	✓ Statistical model	Limited (basic)	✗	✗
Dialect Support	✓	✗	✗	✗

Technologies used

Kotlin for Android development

Jetpack Compose for modern UI

MVVM architectural pattern

StateFlow for state management

Coroutines for async operations

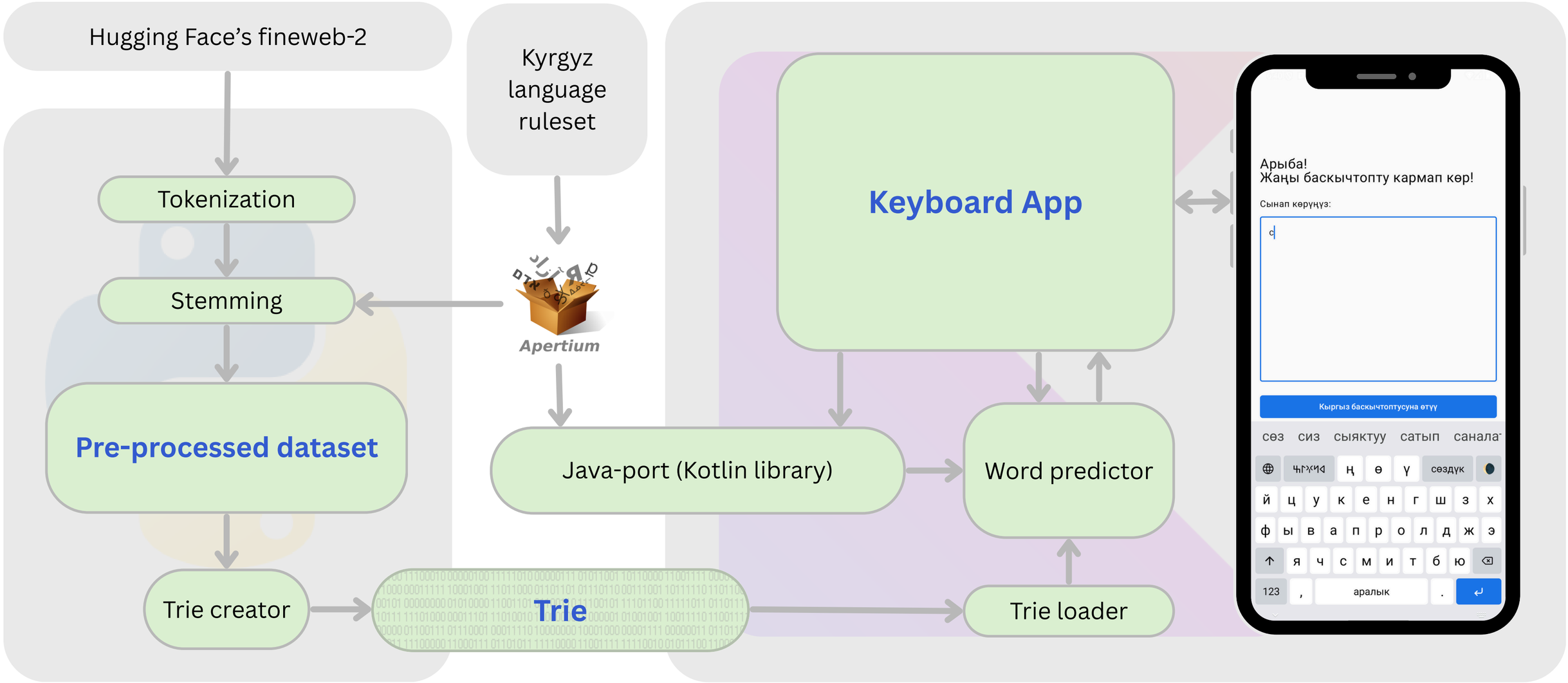
Python for highly parallel data processing

Hugging Face's **fineweb-2** dataset
for Kyrgyz language corpus

Apertium and **Helsinki Finite State Toolkit**
for morphological analysis and stemming

Trie structures for faster processing

Architecture

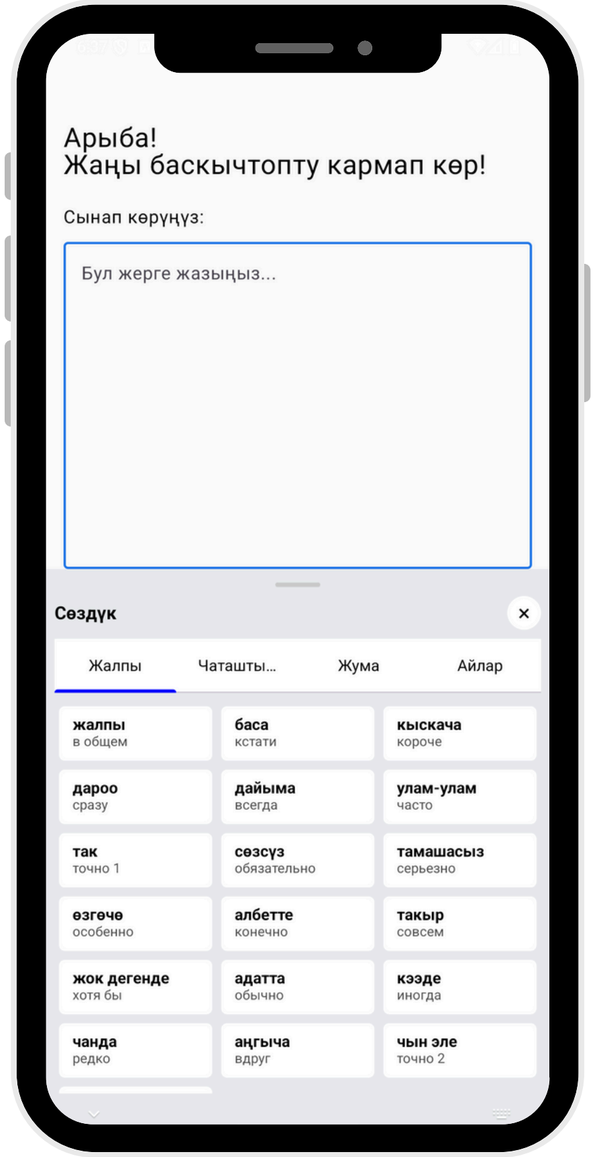


Ademi

Android App Development

Main app UI & Keyboard System

UI with Jetpack Compose
Custom keyboard implementation
Activation & permission flow
System integration & switching
Implemented built-in dictionary



Nikita

Complete Backend Design and Implementation

Pre-processing

Found and tokenized the dataset using a custom-built tokenizer
Created a **huge highly concurrent preprocessing pipeline**

Apertium

Researched Apertium and integrated it into the preprocessing pipeline
Compiled and linked legacy Apertium with Kyrgyz morphological data for use on Java-based platforms

Trie

Designed a binary data structure for efficient word storage and retrieval
Implemented its creation, (de-)serialization, and usage in both Python and Kotlin

---

The Invisible Work That Makes It All Possible

Worked A LOT on performance optimization, which resulted in a very high processing speed and a low trie size
Ensured clean and efficient code with manual reviews (supported by CI), automated deployment via CD
Leveraged a high-memory server (>100 GB RAM) to support trie construction

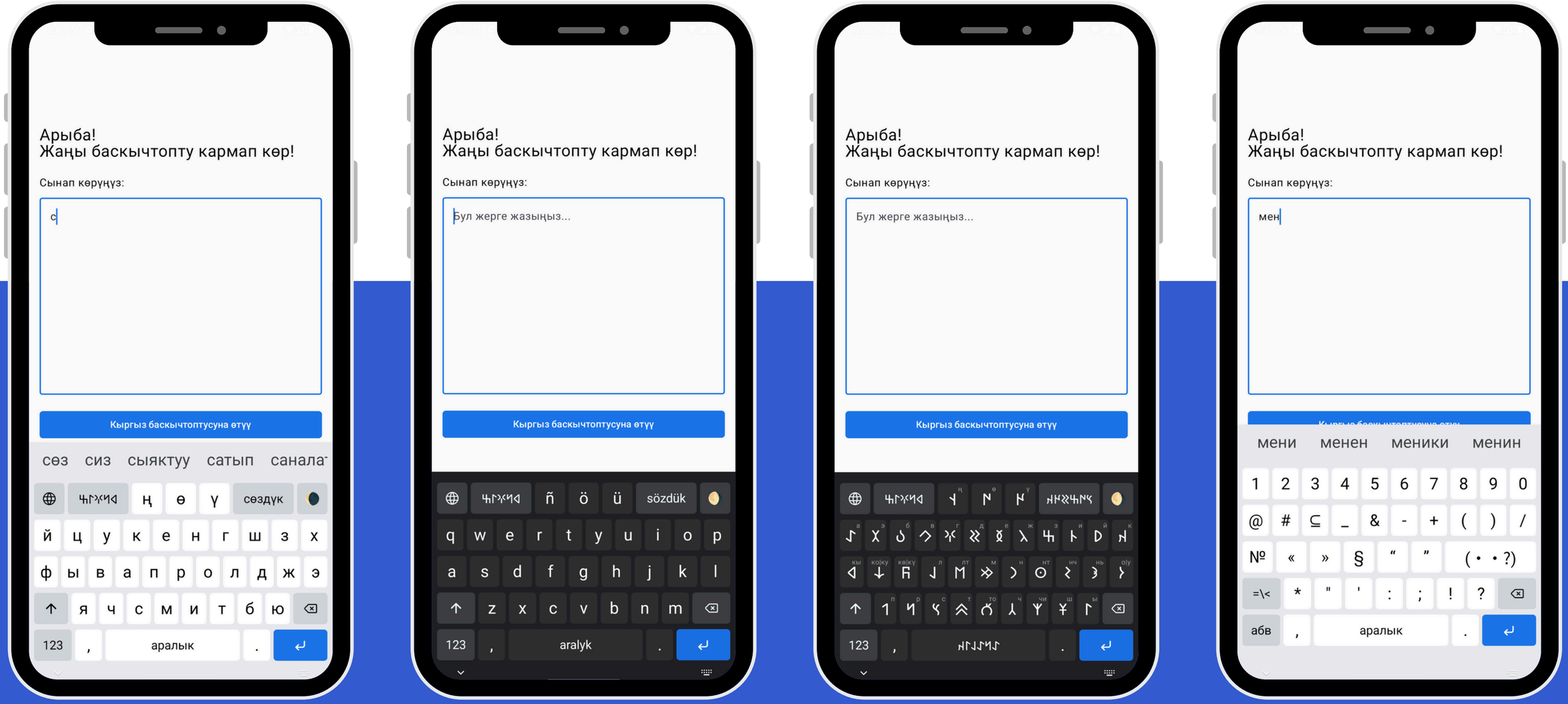
Pavel

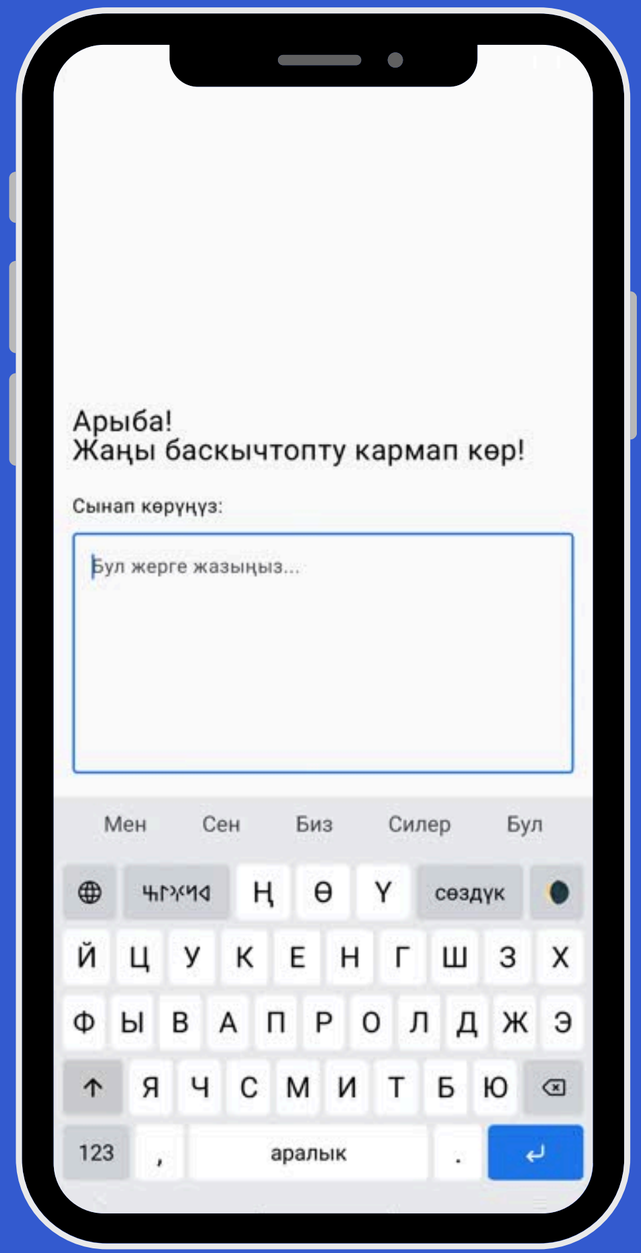
Assistance

Valuable feedback

Provided feedback
Participated in weekly meetings
Attempted early versions of trie and trie loader
Tested app for bugs using Android Studio emulator

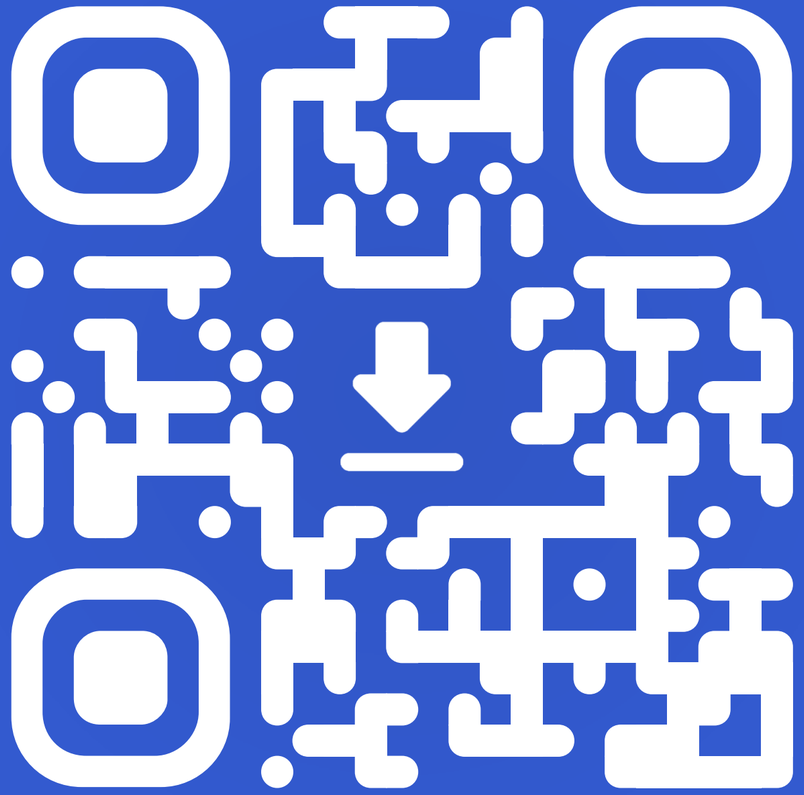
Screenshots




Арыба!
Жаңы баскычтопту кармап көр!
Сынап көрүңүз:
Бул жерге жазыңыз...
Мен
Сен
Биз
Силер
Бул
Ң
Ө
Ү
сөздүк
Й Ц У К Е Н Г Ш З Х
Ф Ы В А П Р О Л Д Ж Э
Я Ч С М И Т Б Ю
123
,
аралык
.

Demo

https://github.com/Kyrgyz-Keyboard



https://kyr.npanuhin.me/Kyrgyz-Keyboard.apk