Ministry of Education and Science of the Republic of Kazakhstan
Suleyman Demirel University

Zhomart Rsaliyev

The e-book reader platform development in Kazakh language

A thesis submitted for the degree of
Bachelor in Computer Systems and Software
(degree code: 5B070400)

Kaskelen, 2018

Ministry of Education and Science of the Republic of Kazakhstan
Suleyman Demirel University
Faculty of Engineering and Natural Sciences

The e-book reader platform development in Kazakh language

A thesis submitted for the degree of
Bachelor in Computer Systems and Software
(degree code: 5B070400)

Author: **Zhomart Rsaliyev**

Supervisor: **Associate Professor, Doctor of Technical Sciences Lyazat Naizabayeva**

Dean of the faculty:
Assist. Prof. Meirambek Zhaparov

Kaskelen, 2018

Abstract

Nowadays, the Republic of Kazakhstan is going to make historical step forward by switching the country's alphabet from Cyrillic to Latin script. The government announced a law as a part of modernization and development effort until 2025 year. It is clear for all to see that these type of changes will affect culture and ideology of nation. If we take a transition as a flying object, so my mission is to ease the landing to the ground by developing an application platform, which will collect e-book readers into one place.

In this thesis, a multi-script e-book reader application for Android OS is presented. The application takes advantage e-books initial "usability" to manage the difficulties in understanding the new alphabet by combining the initial context with the Latin script and to provide the feedback of transition.

Abstract

Nowadays, the Republic of Kazakhstan is going to make historical step forward by switching the country's alphabet from Cyrillic to Latin script. The government announced a law as a part of modernization and development effort until 2025 year. It is clear for all to see that these type of changes will affect culture and ideology of nation. If we take a transition as a flying object, so my mission is to ease the landing to the ground by developing an application platform, which will collect e-book readers into one place.

In this thesis, a multi-script e-book reader application for Android OS is presented. The application takes advantage e-books initial "usability" to manage the difficulties in understanding the new alphabet by combining the initial context with the Latin script and to provide the feedback of transition.

Abstract

Nowadays, the Republic of Kazakhstan is going to make historical step forward by switching the country's alphabet from Cyrillic to Latin script. The government announced a law as a part of modernization and development effort until 2025 year. It is clear for all to see that these type of changes will affect culture and ideology of nation. If we take a transition as a flying object, so my mission is to ease the landing to the ground by developing an application platform, which will collect e-book readers into one place.

In this thesis, a multi-script e-book reader application for Android OS is presented. The application takes advantage e-books initial "usability" to manage the difficulties in understanding the new alphabet by combining the initial context with the Latin script and to provide the feedback of transition.

Аңдатпа

Қазіргі таңда, заман талабы бойынша Қазақстан Республикасы ұлттық ана тілін латиница әліпбиіне көшіру жайлы ақпарат берді. Мемлекет басшысы еліміздің әлеуметтік жаңғырту мен даму жолында 2025 жылға дейінгі өту кестесін бекітті. Осыған ұқсас тарихи өзгерістер халқымыздың мәдениеті мен көзқарасын өзгертетіні айдан анық. Сол себепті, халықтың бөлшегі ретінде менің мақсатым - жаңа әліпбиге ауысу таңында қолғабыс ете білу.

Дипломдық жұмыс Android операциялық жүйесіне арнайы жасалған бағдарламаны ұсынады. Бұл бағдарлама - келе жатқан өзгерістерді жеңілдету және содан алынған пікірлерді жинау мақсатымен жасалған электрондық кітаптарды оқуға арналған құрал.

Аннотация

В настоящее время Республика Казахстан на пороге исторического перехода с казахского алфавита на латиницу. Глава государства подписал указ о переводе, как часть модернизации и на пути развитии страны до 2025 года. Разумеется, что масштабы изменении такого типа повлияют на культуру и идеологию нации. Поэтому, чтобы избежать последствия перехода, хочу облегчить адаптацию с помощью приложения.

В этой дипломной работе представляю удобное приложение для операционной системы Android, которое объединит читателей электронных книг. Приложение придает ясность в понимании языков и дает возможность получать отзыв от пользователей.

Contents

Chapter 1

Introduction

"Ана тілін ұмытқан адам өз халқының, өткенінен де, бүгінінен де, болашағынан да қол үзеді."

- Ғабит Мүсірепов

Everyone knows that the language is the heritage, which is given from generation to generation, but nobody knows the real power. Language is the bridge between nations and religions. Therefore, every nation respects and appreciates mother-tongue, because human language is specific from all other known animal form of communication. Language makes individual to be able to express emotions and thoughts. Thus, the language helps to pass on the torch to the next generation by two ways, spoken and written, respectively. Firstly, the most ancient way of transfer of knowledge in the history of mankind, spoken way, is not reliable, because as time goes it could twist the meaning of words, even no one can guarantee that the text will soon be forgotten. After invention of progressive way of recording information by our ancestors, people have begun to use for the letter almost all on what it is possible to write: clay tablets, tree bark, sheets of metal. So secondly, the written way is more reliable, because it is transferred by newspapers, magazines, books and so on, which is proven over time.

Book is written by individual to share experiences gained; advice not to repeat the same mistakes; emotions and how to handle them; dream about the future and give freedom for creating; imagine unreal things to implement any kind of stuff. Although, digital age shows that new way of delivering information is more secure and more comfortable. So that, electronic books are becoming more popular and on the way of replacing paper based books from 2000s.

The Republic of Kazakhstan on the verge of the beginning of new era and these difficult times will get us closer to other Turkic speaking countries. In 2017, the President announced the country's language will steadily switch from Cyrillic to Latin alphabet until 2025 year [1]. The government is planning to spend a thousand millions of tenge to teach teachers and students, to translate official documents, etc. Regardless, it is obvious that it is political demonstration

and promotion to nationalism. On the other hand, it was the dictates of the present time, because from the beginning Turkish flesh and blood influenced every person by insensible degrees. In addition, many examples can refresh minds, like international projects used Latin letters to introduce own product for countries the other side of sea. Moreover, "mind modernization" of native nation needed new script to open the doors of the information age together and collect minority groups of population into one community all over the world.

In order to handle a discussed situation with the highest possible result, this thesis advances an application for users to allow them reading electronic books. Users understand the possibilities that comes widely in daily basis and enter the universe of digital wave. Growth in interest among e-books have also changed the standards of daily used extensions of books. Today, in this rivalry between extensions Electronic Publication(ePub) is the fast growing type and has already found own level through becoming the widely accepted standard. With the help of ePub users can easily approach to unlimited numbers of books in various languages in special readers. As such, ePub format meets requirements of those who have an unstoppable lust for reading.

1.1 Motivation

The transformation of the way of accessing and reading a book changed in last decades. Electronic books become more attractive instrument among readers as long as smart phones could store a huge amount of data like bookshelves that takes a lot of place and impossible to carry all of them. Consumers prefer e-books for these well-known reasons:

- Easy to carry in a single device in spite of the quantity

- Accessible wherever and whenever

- Supports different languages/scripts

- Preview and gain information in application

- Readable in low light or even total darkness

List of the reasons described on the top answer the main question. Firstly, creating a space for users where they can experience the Latin alphabet as foremost reader and allow them to share their feedback. Because they are the general audience and each individual could get a chance to influence as a part of system. Secondly, youth generation is a key for the future and categorized books for the youngsters is the number one mission. Thirdly, independent writers who does not have support will be able to get a ready audition to present new written books. If all three categories of users get together at one platform, every person could get what they want.

1.2   Purpose of Thesis

The purpose of a thesis is to provide with maximum information about the development of an e-book reader application, which is capable to display catalogues of books in two alphabets. Appearance of the layout is built in such a way that it helps to achieve major goals as expected: remodeling the process of acquisition of product and improving an usability of the new script through learning e-books.

Platform focuses firstly on users, it means application made for users to face their needs and it is called "Oqyrman"(translated as "reader", from Kazakh). Oqyrman offers special advantages, which are not reachable in other services and has some good features like:

Multi-script: It is not the unreal breakthrough among functionality, but Oqyrman has already implemented this feature. In future plans, Oqyrman will be able to display two screens for comfortable comparison between versions. However, it is the main tricky part of application that is fully free to use and will get new feelings.

Content-rich: A fresh stylized catalogue with amazing collections of books done with the love and taste. Whatever the application has many features, however, the content takes a main role and it is the base that it would be hard to describe without.

Comfortable: Smart phones become popular because of their smooth controls by fingers and the term "scrolling" took a special place in our culture of using technologies. Gestures are revolution and it would be mistake if developers do not use in their practices. Gesture abilities that could adjust desired standards of brightness, font size, font style and so on, are included and ready to be used.

Universal: ePub format of e-book is universal and it can be used in other e-book readers without any problem whether a user prefers using their own known application rather than Oqyrman.

Favourite: User can store selected bookmarks and quotes right in his beloved device, where every information is possible to be obtained whenever. Favorite books are adjustable for creating book collections with the last saved pages as playlists in music player applications.

User experience and feedback guide developers for adding new features. Application creating progress shows developers new ways of implementing functions to reach the perfect product at all. Feedback is available at Play Market.

1.3   Thesis Outline

The first chapter is Introduction chapter. It gives short insight into the work done. Claims and explains why exactly this thesis theme and this area have been chosen. In Chapter Background Research, we review related work and formulate the problem to solve in next chapters. It gives, as its name "background", background image and history for a reader, also it is connected with previous chapter meaningfully. Chapter Technologies displays all technologies that is used in application development process and defines with specific book terms. They are all official definitions taken from other sources. Technologies consists of major three parts: ePub, Django and Android OS. Each section and subsection draws a general picture of what is happening. Chapter Software is describing the solution to the problem and reveals final result. And in the last, Conclusion and Future work chapters are concluding parts of the thesis, where everyone could find for himself necessary information and final output.

Chapter 2

Background Research

This chapter discusses e-Reader market situation as a whole object in three sections. Slow progress of content-making in Kazakhstan shows poor result, it could be noticed when you start to search something you need and could not get a proper response. The research describes a problem and gives alternative suggestion to improve further scenario. Generally, chapter is associated with multi-script language and submits the solution against the current arguments.

2.1   Latin Alphabet

This section describes the full chronology of the situation in particular aspects. The original alphabet on Cyrillic has been accepted in 1940 and contains 42 letters: 33 letters of the Russian alphabet and 9 specific letters of Kazakh.

In 2017, the President, Nursultan Nazarbayev, declared a decree of transition of the Kazakh alphabet from Cyrillic to Latin script, then this news had been reached to the ears of media of European and North American countries. Newspapers and online resources tried to bring whole story to the readers.

First version of alphabet was the worst which is happened with this change, because it was consisted of 25 letters and 8 combined letters or digraphs. In opposition, there were

№	Написание	Звук	№	Написание	Звук
1	A a	[а]	17	N' n'	[ң], [нг]
2	A' a'	[ә]	18	O o	[о]
3	B b	[б]	19	O' o'	[ө]
4	D d	[д]	20	P p	[п]
5	E e	[е]	21	Q q	[қ]
6	F f	[ф]	22	R r	[р]
7	G g	[г]	23	S s	[с]
8	G' g'	[ғ]	24	S' s'	[ш]
9	H h	[х], [h]	25	C' c'	[ч]
10	I i	[і]	26	T t	[т]
11	I' i'	[и], [й]	27	U u	[ұ]
12	J j	[ж]	28	U' u'	[ү]
13	K k	[к]	29	V v	[в]
14	L l	[л]	30	Y y	[ы]
15	M m	[м]	31	Y' y'	[у]
16	N n	[н]	32	Z z	[з]

Figure 2.1: Oct.26, 2017 decree

raised furious reaction from community and government decided to show new version by citing "the first step is always the hardest".

Second version officially offered a version with replacing digraphs to letters with apostrophes after a month(Figure 2.1) [2]. But the population was violently indignant on all platforms, saying that it is insufferably difficult to read words with apostrophes in applications and websites. So, the time showed that it was second disaster happened, because it was impossible to write or type while entering the internet addresses and different forms of letters with apostrophes was not good idea. With such Latin text volume increases, the speed of set is slowed down, the convenience of perception decreases. Apostrophes visually break off words and a whole picture of the text. From technical aspect the word with an apostrophe isn't complete. As an example, Uzbekistan which uses the alphabet since 1995, similar to second variant. Till recent times in Google it was impossible to look for fully the Uzbek texts - the search engine considered an apostrophe a rupture of words.

Anyway, developers found their gold mine and started to present their own products and succeeded in gaining popularity in short time. After, it took waves of increasing negative comments and it forced the government to make their third attempt while linguists were trying to make their own individual variants of new script.

Indignation of Kazakhstan citizens have been heard. Third attempt had better variant than before, because it was closer to other Turkic languages and of course, got warm welcome from audition. New script consists of 32 letters and 6 new letters with acute accent. In addition, it has combinations of letters that sound as one, they are not included for the alphabet because of their origin, Russian language. But also this option allegedly demands completion.

Meanwhile pupils of one of schools of Petropavlovsk have developed the Latin Translator Dictionary [3] mobile application for transfer of the Kazakh words to Latin graphics. The program will work for mobile phones with the Android operational system. Schoolgirls have chosen it not accidentally. They have conducted a research and have defined that 70 percent of Kazakhstan citizens use Android.

№	Написание	Звук	№	Написание	Звук
1	A a	[a]	17	Ń ń	[ң]
2	Á á	[ә]	18	O o	[o]
3	B b	[б]	19	Ó ó	[ө]
4	D d	[д]	20	P p	[п]
5	E e	[e]	21	Q q	[қ]
6	F f	[ф]	22	R r	[p]
7	G g	[г]	23	S s	[c]
8	Ǵ ǵ	[ғ]	24	T t	[т]
9	H h	[x], [h]	25	U u	[ұ]
10	I i	[i]	26	Ú ú	[ү]
11	I ı	[и], [й]	27	V v	[в]
12	J j	[ж]	28	Y y	[ы]
13	K k	[к]	29	Ý ý	[у]
14	L l	[л]	30	Z z	[з]
15	M m	[м]	31	Sh sh	[ш]
16	N n	[н]	32	Ch ch	[ч]

Figure 2.2: Feb.19, 2018 decree

Approximately in half of a year, people and governance, together, made progressive actions. It is a good sign that government is trying to make language comfortable for users and listening to their residents. But, it is not the finish, because, as long as time goes, practice will show which direction is right, which is not.

To sum up, adaptation to the new script will take a long time if citizens do not act as one organism. The future challenges are waiting for and this research will help Kazakh language.

2.2 The Advantages of Latin alphabet

Kazakhstan is to change its official alphabet for the third time in less than 100 years in what is seen in part as symbolic move to underline its independence [4]. The new alphabet is becoming social phenomenon that is governed by the needs of globalization and cultural openness. The great transition is an important step toward establishing the new digital country. Due to transition, content-making in Kazakhstan will be improved and would catch up with the others.

Secondly, similarity of the alphabet with Western world would interest them for investment and partnership with native businessmen. Close relations will increase development in country and influence other partners.

2.3 E-books

Today, it is hard to imagine world without technology and its gadgets. It is entered into our life so easy, such that it becomes clear of new peaks are going to be reached. After smart phones, Kindle readers, and even tablets are invented, paper based books transferred into electronic gadgets. Almost every human being has smart phones in his pocket and each of them spend their time with this tool. It means that there is no problem to contact with user whenever you want. That's why e-book is one of the alternatives of communicating with terminal user.

An electronic book (or e-book) is a book publication made available in digital form, consisting of text, images, or both, readable on the flat-panel display of computers or other electronic devices [5]. Although sometimes defined as "an electronic version of a printed book", some e-books exist without a printed equivalent. Commercially produced and selling e-books are usually intended to be read on dedicated e-reader devices. However, almost any sophisticated computer device that features a controllable viewing screen can also be used to read e-books, including desktop computers, laptops, tablets and smart phones.

However, lack of resources in Kazakh stops the whole process. In order to solve this problem let's discuss it in next section.

(a) Bookmate   (b) Oyan

Figure 2.3: Two e-book readers

2.4 The e-Reader market

Poor market of e-books in Kazakh leaves a lot to be desired. So, let's make a comparative table that shows differences between e-book reader applications:

Resource	Application	Website	Database	Price	Amount of books
kitap.kz	web-based	yes	wikibilim*	Free	10,000
Oyan	yes	no	wikibilim*	Free	10,000
Bookmate	yes	yes	bookmate	9,99 £	~4000*

1. WikiBilim Public Foundation is a nonprofit organization operating in the Republic of Kazakhstan. The organization was established for developing and promoting online educational content in the Kazakh language. "Wiki" - means content generated by users, "Bilim" means knowledge [6].

2. Unknown quantity of books.

As the table shows, books translated into Kazakh language can be downloaded/bought from these three sources. Firstly, kitap.kz is a website and it has main advantage as a database from WikiBilim and they are partners. As we know, kitap.kz has approximately 10,000 books and this number is increasing day by day. But, the main disadvantage that it has no mobile application, which makes it weak competitor between sources. However, simple application, iOS and Android, Oyan has been created with the database of kitap.kz by Zero to One Labs. Nonetheless, it has many bugs and disadvantages: cannot save the last page of

read book, app searches incorrectly and shows a few result, has little functionality. The main problem of these two sources that they simple and sharp mistakes in words, which makes feel uncomfortable and unreadable. Thirdly, Bookmate is the powerful application, also iOS and Android, which is done by professionals and no doubt that it is the best among all. But, foreign market for them affects with small database of publications in spite of their attempt by teaming with Kcell, popular cellular communication operator placed in Kazakhstan. Lastly, sales in country show that native people is not ready to pay money for reading books.

All in all, after some analysis we can understand that mobile application is good decision than website and Android/iOS users are almost equal. Then, the main moment - is the price, because customers usually choose cheaper variant, in other words "free content rather than paying". As we can see, all of the resources do not support independent writers, which is not critical, but improper and dishonest against the content creators.

2.5   Alternatives

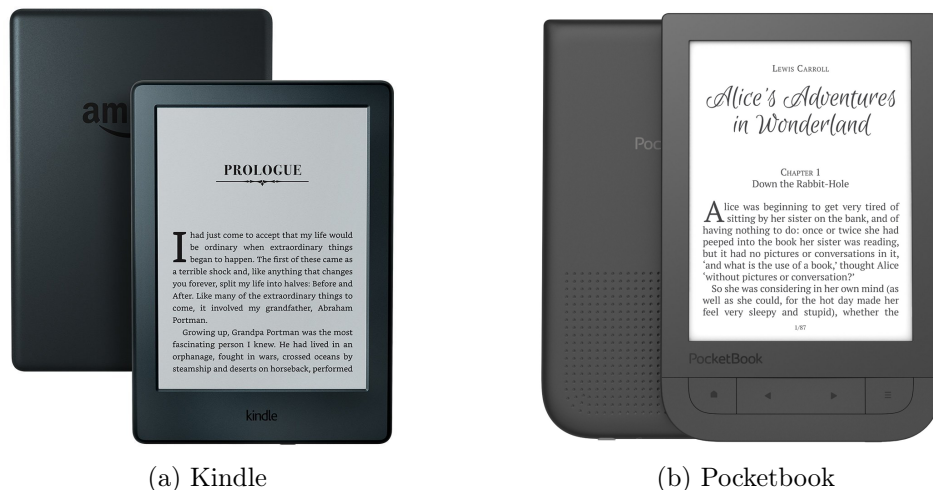(a) Kindle   (b) Pocketbook

Figure 2.4: The most popular e-book readers

A growing number of gadgets turned over people's consciousness towards new inventions in the Information Age. It is impossible to imagine that how mankind lived before and it is clear that it is just beginning. It should be noted that comfortableness among technologies takes a main role and as time goes computers are turning into smaller and smaller versions. This progress led us into hand held and portable devices. There are many examples for these devices, for instance, mobile electronic devices, e-book devices and laptops. First two alternatives are in advance of laptops, because day-and-night usage statistics overcome personal computers' usage.

Diagram shows us that mobile phone users grow with unreal numbers and the forecast of 2019 guesses that it will hit 5 billions of users. Samsung and Apple are
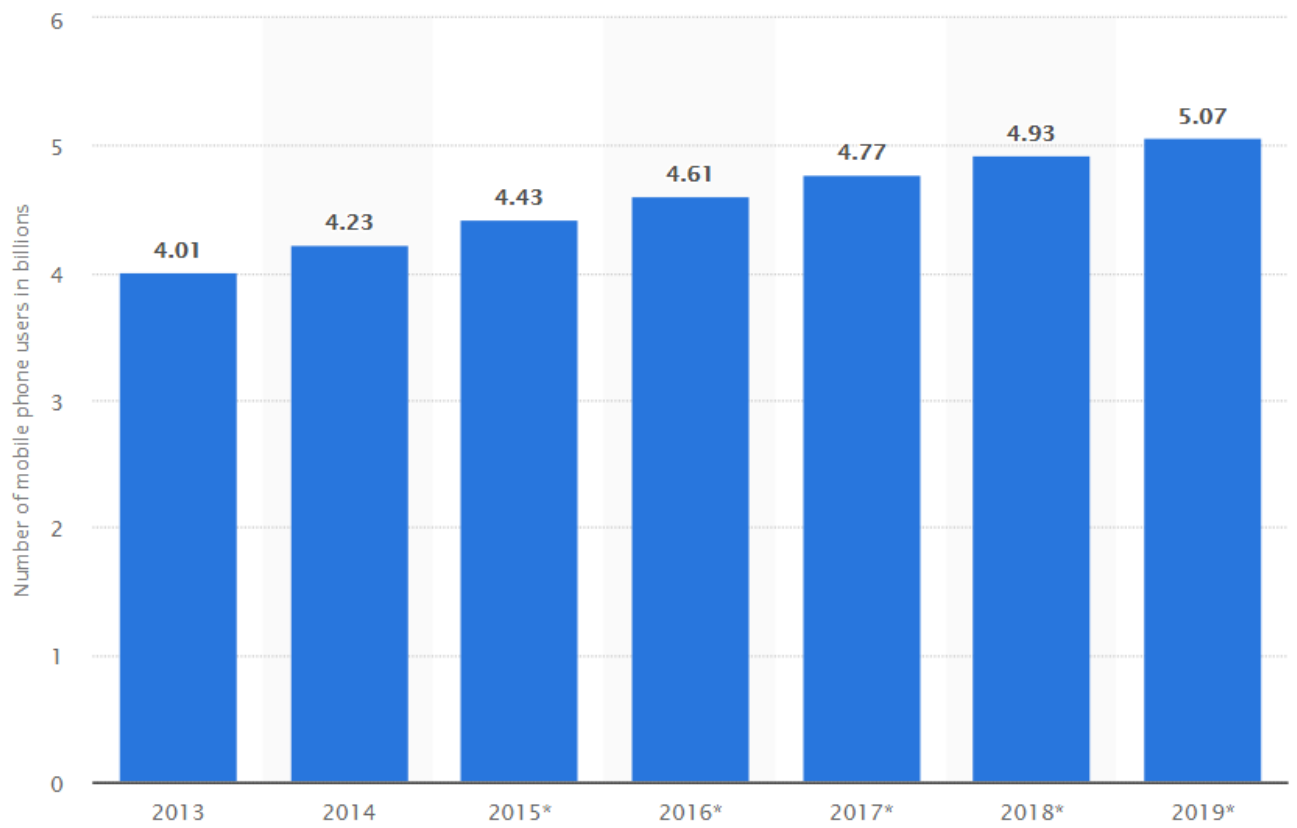
Figure 2.5: Number of mobile phone users worldwide from 2013 to 2019 (in billions)

leading smart phone vendors, with about 18 percent of the market share each [7]. From this statistics, it is clear that mobile phones are more popular than tablet and e-book reader users.

Chapter 3

Technologies

3.1 Choosing Technologies

In this section thesis puzzles out technological part of application. Programming languages and libraries for them are developing very fast, so this element is extremely important. As a content, electronic book, ePub is a good option, because it is universal and whole market sees this format as a standard. Following ePub section will explain completely why exactly is the most favorable decision.

After book extension investigation technologies section describes coding part, which includes back-end and front-end. As a back-end has been chosen Django web framework with the implementation of JSON data for retrieving details and links of files. Then, as a front-end Android OS is more interesting and accessible platform in each aspect. The text below defines with the entire information.

3.2 ePub

EPUB is an e-book file format with the extension **.epub** that can be downloaded and read on devices like smartphones, tablets, computers, or e-readers [8]. It became officially standard in 2007 and it is supported by the largest number of hardware readers. After some study between formats, advantages of ePub took a interest and chosen for the project. The initial version - Open eBook Publication Structure or "OEB" has been conceived in 1999, ePub release - 2007. Advantages of this format:

- Cross-platform

- Adjustable styles

- Like a HTML web page

- Stores bookmarks/notes

- Supports audio/video

3.2.1 ePub Structure

ePub is archived set of files that contains HTML media, CSS style sheet, metadata and audio or video could be added. Also, it could contain specific font styles.

```
--ZIP Container--
mimetype
META-INF/
   container.xml
OEBPS/
   content.opf
   chapter1.xhtml
   ch1-pic.png
   css/
      style.css
      myfont.otf
   toc.ncx
```

Figure 3.1: ePub structure example

1. **mimetype** - is an ASCII file that stores information which type of file format it is, usually "application/epub+zip"

The first file in the archive must be the mimetype file. It must be unencrypted and uncompressed so that non-ZIP utilities can read the mimetype. This file provides a more reliable way for applications to identify the mimetype of the file than just the .epub extension.

2. **META-INF** - folder that directs to xml file which is called container

File container.xml must have this following code:

```
<container xmlns="urn:oasis:names:tc:opendocument:xmlns:container" version="1.0"
   ><rootfiles><rootfile full-path="OEBPS/content.opf" media-type="application/
   oebps-package+xml"/></rootfiles></container>.
```

It explains the main structure of ePub and points to contents of the book, the OPF file.

3. **OEBPS** - folder that contains full information which is needed to construct ePub

The other files (OPF, NCX, XHTML, CSS and images files) are traditionally put in a directory named OEBPS.

4. **content** - does function of a body and contains metadata, manifest, spine and guide

5. **chapter** - content designed as HTML

6. **toc.ncx** - navigation

Table of contents of the book

7. **other files** - media, style and font

3.2.2 Publication

The ePUB container must contain:

- At least one content document

- One navigation document

- One package document listing all publication resources. This file should use the file extension .opf. It contains metadata, a manifest, fallback chains, bindings, and a spine. This is an ordered sequence of ID references defining the default reading order.

The ePUB container may contain:

- style sheets

- PLS Documents

- media overlay documents

3.2.3 Creating an ePub file

```python
def replace(text):
    dic = {"а" : "a", "ә" : "á", "б" : "b", "в" : "v", "г" : "g",\
           "ғ" : "ǵ", "д" : "d", "е" : "e", "ё" : "e", "ж" : "j",\
           "з" : "z", "и" : "ı", "й" : "ı", "к" : "k", "қ" : "q",\
           "л" : "l", "м" : "m", "н" : "n", "ң" : "ń", "о" : "o",\
           "ө" : "ó", "п" : "p", "р" : "r", "с" : "s", "т" : "t",\
           "у" : "ý", "ұ" : "u", "ү" : "ý", "ф" : "f", "х" : "h",\
           "h" : "h", "ц" : "ts", "ч" : "ch", "ш" : "sh", "щ" : "sh",\
           "ъ" : "", "ы" : "y", "і" : "i", "ь" : "", "э" : "e",\
           "ю" : "ıý", "я" : "ıa", \
           "А" : "A", "Ә" : "Á", "Б" : "B", "В" : "V", "Г" : "G",\
           "Ғ" : "Ǵ", "Д" : "D", "Е" : "E", "Ё" : "E", "Ж" : "J",\
           "З" : "Z", "И" : "I", "Й" : "I", "К" : "K", "Қ" : "Q",\
           "Л" : "L", "М" : "M", "Н" : "N", "Ң" : "Ń", "О" : "O",\
           "Ө" : "Ó", "П" : "P", "Р" : "R", "С" : "S", "Т" : "T",\
           "У" : "Ý", "Ұ" : "U", "Ү" : "Ú", "Ф" : "F", "Х" : "H",\
           "Ц" : "Ts", "Ч" : "Ch", "Ш" : "Sh", "Щ" : "Sh", "Ы" : "Y",\
           "І" : "I", "Э" : "E", "Ю" : "Iý", "Я" : "Ia"}

    for i, j in dic.iteritems():
        text = text.replace(i.decode("utf-8"), j.decode("utf-8"))

    return text
```

Figure 3.2: Cyrillic to Latin replacing algorithm

Above code is an algorithm, which converts every letter in a word one by one, performed by Python function. It uses simple built-in function replace() to scan the word and if there is similar letter used in our dictionary, called dic, then automatically matches letters to replace after. But, there raises another problem with

coding, because symbols in new script challenges Unicode characters. In order to avoid this problem (UnicodeDecodeError), there is used an default encoding for Python 2.7.13. So you can simply read and write a Unicode character in a string, literally. Finally, function returns the value as a string to make for new operations ready.

Free from copyright books has been parsed from open-source websites and converted into the new script by algorithm as it is shown in Figure 3.2. Python is the main tool while parsing and creating ePub files. It uses libraries like *urllib2*, *BeautifulSoup* and *requests* for network actions and library *ebooklib* helps to get text from contents and create new book after translation. Where parser libraries works for:

- *urllib2* - is a Python module for fetching URLs (Uniform Resource Locators). It offers a very simple interface, in the form of the *urlopen* function. This is capable of fetching URLs using a variety of different protocols. It also offers a slightly more complex interface for handling common situations - like basic authentication, cookies, proxies and so on. These are provided by objects called handlers and openers [9].

- *BeautifulSoup* - is the parser for syntactic analysis of the HTML/XML files written in the Python programming language which can transform even the wrong marking to a tree of syntactic analysis. It supports easy and natural ways of navigation, search and modification of a tree of syntactic analysis. In most cases he will help the programmer to save hours and days of work [10].

- *requests* - is a Python library which elegantly and just carries out HTTP inquiries and sometimes can replace old and nonfunctional urllib2.

- *ebooklib* - is a Python library for managing EPUB2/EPUB3. It's capable of reading and writing ePub files programmatically [11].

And final step is to collect defined outputs into one file with .epub extension. There are two known ways of creating an e-book. Firstly, programmatically with Python, which uses already described module. Documentation of implementation from group of developers is fully explained in github source. There also exists many available examples. Secondly, elementary way with simple instructions, which consists of easy 3 steps and every personal computer user can pull this off. It is written after the tutorial code.

Following code below is the example of making an ePub file without any problems. Each action is clear for user. Initially, book variable is created and assigned to ePub value in order to give an opportunity for functions of library. Then, step by step book information is added, like a title, an author, a default language and identifier, if it is needed. After basic actions, book needs contents

and every part have to finish with add_item() function. There is no limit to chapters of book. All in all, variable gets values of ToC (Table of Contents) and navigation elements. Lastly, there is added a style of ePub with user's needs:

```
1  book = epub.EpubBook()
2
3  book.set_identifier('id123456')
4  book.set_title('Sample_book')
5  book.set_language('en')
6
7  book.add_author('Author_Author')
8
9  c1 = epub.EpubHtml(title='Intro', file_name='chap_01.xhtml', lang='en')
10 file = open(path + "/content_" + str(i) + ".xhtml", "r")
11 c1.content=res.encode("utf-8")
12 book.add_item(c1)
13
14 book.toc = (epub.Link('chap_01.xhtml', 'Introduction', 'intro'),
15            (epub.Section('Simple_book'), (c1, )))
16
17 book.add_item(epub.EpubNcx())
18 book.add_item(epub.EpubNav())
19
20 style = 'BODY_{color:_white;}'
21 nav_css = epub.EpubItem(uid="style_nav", file_name="style/nav.css", media_type=
        "text/css", content=style)
22
23 book.add_item(nav_css)
24 book.spine = ['nav', c1]
25 epub.write_epub('test.epub', book, {})
```

Second way:
1. Create a new, empty folder with the title of your book as its name, no spaces, all lowercase is best.
2. Zip the empty folder. You now have an empty zip folder.
3. Now you will put the pieces of the ePub into it, in the right order:
3.1. First, drag and drop the mimetype file into the zip folder.
3.2. Next, add the META-INF and the OEBPS folders (you can drag both into the folder at once, their order does not matter, as long as the mimetype file is first) [12].
Finally, change the .zip extension on your compressed folder to .epub.

3.3   Django

Django [13] is free and open-source Python Web framework that is established to build web applications in a short time without needing to reinvent the wheel. Rapid development is a new black of programming. Application uses exactly this tool because of flexibility in designing the API calls. Today, mobile phone application cannot survive without API requests in server-side projects. API is like a waiter in a restaurant, you will get that stuff that you called. Such that, API uses standard to transfer data with GET and POST functions and it is called

JSON. JSON(JavaScript Object Notation) [14] has benefits in delivering serialized large data, so JSON is better format than XML.

3.3.1 Python Web framework

Django uses models to describe data structure and already has customized administration page. Application has main model called Book and model has attributes like:

- **title and title_latin** - multi-script name of a book

- **author and author_latin** - multi-script author/writer of a book

- **image** - multi-script designed unique cover image

- **epub_cyrillic and epub_latin** - two versions of a book

- **description and description_latin** - multi-script summary of a book

- **section_id and section_name** - collections of books

- **updated and timestamp** - automatically generated time values for sorting

```python
1   from django.db import models
2   from django.urls import reverse
3
4   class Book(models.Model):
5       title = models.CharField(max_length=200)
6       title_latin = models.CharField(max_length=200, default="")
7       author = models.CharField(max_length=200)
8       author_latin = models.CharField(max_length=200, default="")
9       image = models.FileField(upload_to="media/", null=True, blank=True)
10      image_latin = models.FileField(upload_to="media/", null=True, blank=True)
11      epub_cyrillic = models.FileField(upload_to="epub/", null=True, blank=True)
12      epub_latin = models.FileField(upload_to="epub/", null=True, blank=True)
13      description = models.TextField()
14      description_latin = models.TextField(default="")
15      section_id = models.IntegerField(default=0)
16      section_name = models.CharField(max_length=200, default="null")
17      updated = models.DateTimeField(auto_now=True, auto_now_add=False)
18      timestamp = models.DateTimeField(auto_now=False, auto_now_add=True)
19
20      def __unicode__(self):
21          return self.title
22
23      def __str__(self):
24          return self.title
25
26      def get_absolute_url(self):
27          return reverse("detail", kwargs={"id": self.id})
28
29      class Meta:
30          ordering = ["-timestamp", "-updated"]
```

3.3.2 RESTful API

REST is a structural style of a software presented for building re-sizable web services, it was firstly taken from the paper of Roy Fielding. And library *rest framework* for Django deals with these operations. Generally, it listens to responses and sends requests to server-side, which has already been realized by Django. In other words, it is like a waiter in restaurant, which gets to your place what you ask. Full documentation of API responses will be available in the end of thesis.

Information is retrieved at background, where terminal user cannot see what is happening inside. It is needed for security and functionality. Web services are constructed in that way in which priority is the speed of information getting. Project uses JSON format to handle data. Each API can be called by particular url assigned in django project. Below code is an example of using an array of books and there is one tutorial element in array with specific id:

```
1  [{
2      "id": 1,
3      "title": "Test book title",
4      "title_latin": "Latin test book title",
5      "author": "Test book author",
6      "author_latin": "Latin test book author",
7      "image": "/media/cover.jpg",
8      "image_latin": "/media/cover_latin.jpg",
9      "epub_cyrillic": "/media/epub_cyrillic/test_original.epub",
10     "epub_latin": "/media/epub_latin/test_latin.epub",
11     "description": "Summary",
12     "description_latin": "Latin Summary",
13     "section_id": 1,
14     "section_name": "Top 10",
15     "updated": "2018-05-01T00:00:24.077726Z",
16     "timestamp": "2018-05-01T01:00:24.077726Z"
17  }]
```

In short, mobile application requests three main API calls. Firstly, to get the list of all books which is stored in database. By default it is sqlite3, because the project does not need huge amount of materials to deal with. Secondly, each book has its own information or details (how project defines). Thirdly, there will be collections of books in further updates for displaying the content more comfortable and exclusive. Finally, small calls that complements other functions.

All operations are available in default django administrator page or project has optimal variant for performing half of the given operations. Second way is implemented for easy access with simple design.

3.4 Android

This section has a brief description of the most installed operating system and has three main subsections. Totally explained why this operating system has been chosen and which features it has among others.

Android is a mobile operating system designed for smart phones and tablets. As of May 2017, it has over two billion monthly active users [15]. Analysis online magazine announced a statistics, where 197 billion mobile apps downloaded in 2017 and this number still increasing. Because, democratic politics of Google gives a chance for developers to come up to popular platform.

On the one hand, price was the one of the main attracted advantages before development. After investment and some time application will reach to Apple devices.

3.4.1   Android Platform

Completed courses in Java programming language gave an opportunity to evaluate possibilities in this area. Android Studio is updated to a new version with faster compilation time and appropriate window size for Windows operating system that offers exclusive features.

An application is created in Android Studio 3.0.1 and needs minSdkVersion 21 (5.0) Lollipop. If we come to the structure of the project, then it consists of bottom navigation bar, which are collected by fragments, not activities. Most down-loadable applications in Play Market demonstrates that navigation at the bottom is easy understandable for user and decreases the time that user spend on application learning. Also, minimalistic design with bold font sizes, which became a trend in a year, gives a pleasure of using and reading without hurting their eyes.

Every single detail had been studied before addition to the project and source is available on github.

3.4.2   Libraries

Set of tools for Android Studio to handle a lot of operations with ready solutions from open-source websites. Libraries in the project are chosen by specific criteria like update patches, contributor numbers, advantages among other similar libraries.

- **ButterKnife** - is a library that cleans and removes unused parts of a code to make perfect for reading

- **RecyclerView** - is a standard of building layouts

- **CardView** - is an element of lists the way are collected

- **Volley** - to handle http requests

- **Gson** - to work with JSON format

- **Glide** - is a download manager for images

- **FolioReader [16]** - is a tool to work with ePub books

3.4.3 Touch Gesture Controls

User experience prompts developers to make predictable apps at first usage. Users want to play with gestures with known finger motions. That's why smart phones had been created. New trends are getting developers to minimize elements on the screen by removing instinctive buttons and re-designing the navigation panels of main page. Adjustable brightness; font size; font style with the finger while reading and saving favourite bookmarks with long click of word; zoom; scroll horizontally and vertically.

Chapter 4

Software

Pre-final chapter includes screens and explanations of chosen methods. In this chapter, the final product is achieved by combining all previous chapters to get one full product. Section "Architecture overview" explains the flow of an mobile application and shows how it responses to the moves.

4.1   Architecture Overview

Let's begin with the database as an initial point of getting information. Project uses PostgreSQL database in production instead of sqlite3, which is fast, to work with JSON serializers and retrieve data to client. However, it will never be called till the one of three buttons in bottom navigation is clicked. LibraryFragment is responsible for displaying all books in ePub format, which are placed in special directory, called Oqyrman. It will be internal storage or external, it depends on user's pick. It means when the application is clicked from desktop it just opens welcome page LibraryFragment without any operations. If user clicks to the second button, DiscoverFragment, it changes the fragment to catalogue of books. In future plans there will be collections of books with creative images of authors. Here, client can explore new books and see details, which commands to display DetailActivity. It is the ending of interactions, but this activity has a download button, where it is optional: Cyrillic or Latin. After downloading, user can return to LibraryFragment to read his book with pleasure. Third button in navigation panel gets a search activity, where you can find books that is currently in database.

Lastly, SettingsFragment, third navigation page, helps user with customization of language and with other adjustments. Here, user can explore FAQs (Frequently Asked Questions), contact with developer and share application with friends.

As it has told, books are available at exact directory and already downloaded books can be opened with third-party applications.
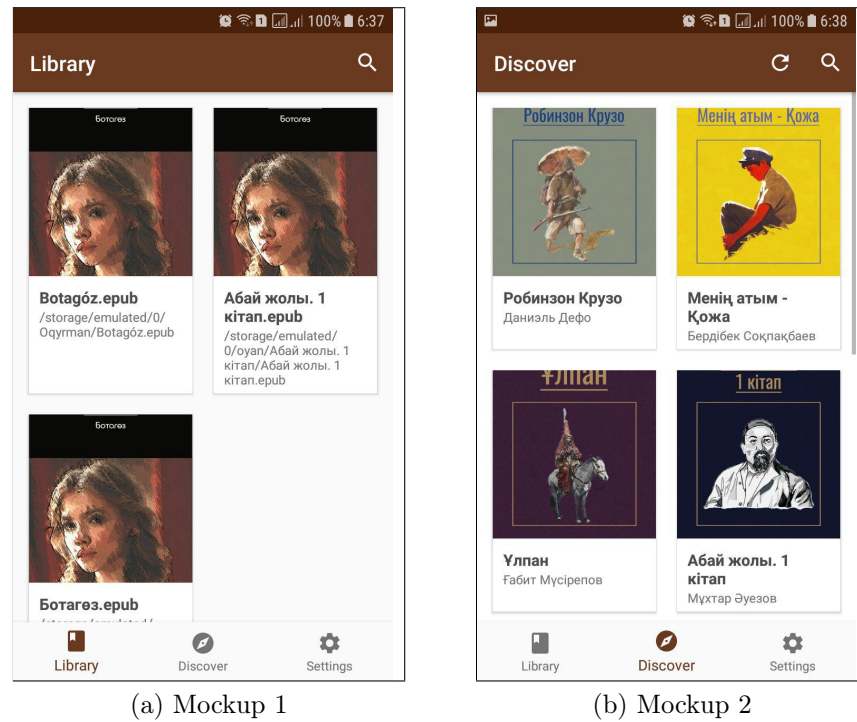
(a) Mockup 1   (b) Mockup 2

Figure 4.1: Template designs

4.2 User Interface and User Experience

Section of UI and UX provides a short review of main steps for practical ideas. User Interface plays a huge role within interaction at unconscious level that could not be seen or fell for ordinary person. In many years of researching application designs the one thing became catchy, it is the influence of coming from iOS to Android. Copying or attempting to make things similar to iOS can be noticed for everyone if you stop for a moment and think. Last year, the new update with bold font sizes became popular by replacing light fonts and it becomes clear that it came from iOS. They completely refreshed the Apple Store and gave new variations of trends. Minimalistic design with single big button and beautiful canvas works blow minds.

Application's design is constructed by Material Design Kit which is created by Google. It helps to run application more quick and perfect. Statistics from focus groups of popular companies established with useful information within combining of architecture and design.

Actually, it is unavoidable fact that design discussed before turns into a new standard, so what is the meaning of reinventing the wheel? New screens will be added in future works.

4.3 Uploading an application

Firstly, Django project would find its place at hosting with given domain after advanced configurations. Then, optimization and refactoring total code is waiting as a further plan. After successful API settings, we can move about to application. Play Store requires a small charge to be registered. With coming investment it would be uploaded to Google Play Console. Finally, it will take two weeks to check the app for them.

In short, the hosting webfactional.com fully supports django project (back-end) with its applications on Ubuntu 16.04. Hosting offers a free support for 2 months and payable continuation. However, application works on local server produced by XAMPP and will be deployed in near future.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

In this thesis, I introduced an application which offers multi-script books in ePub format. As well, an Android e-book reader is developed for the purpose of allowing users to access multi-script e-book content with intuitive gestures.

During this research, implementation of the project is covered step by step, from requirement analysis to executable solution. Furthermore comprehensive explanation of the development progress shows that the electronic books can be built for distinctive purposes in order to give a more comfortable reading experience. I am pretty sure that Oqyrman has an important potential in these hard times and gives an opportunity for readers to be involved to this process. Readers in Kazakh language could benefit from cultural heritage of our ancestors and new authors who are trying to populate book reading in the digital age.

On the other hand, it was a great experience to make an application on my own. It is possible to make all work slowly and surely, but it would be better just if one designer or copyrighter will help to support application. In this research work based on developing e-book reading application, everyone can notice various types of aspects that developers face everyday. By making this software I can say that I have reached almost all my goals that were planned before. Hard work is behind and the left part will not keep users waiting. I think that my application will influence young developers to rethink and get new publishers to provide more content in our language.

5.2 Future Work

In future I am going to continue writing my paper work to achieve the main goal. New features will come with a new update and a production release. If the application gets a popularity in our country and shows good results, iOS users will get a similar application. The list of future plans:

- **Own customized ePub library reader** - because current library does not support advanced functions that I waited for

- **Collections of books** - models are already ready, now I have to implement with new API architecture

  Re-design of navigation panel with the new trends

- **Latin Keyboard** - in future plans, application would support latin keyboard

  Ecosystem of games and challenges are also planned

- **2 resizable screens** - 2 screens would show the difference between scripts

  Users could match two texts in one screen

- **Dictionary** - a dictionary will be added as a new feature with the first update, because it is the most useful function

  Many words in Kazakh language is not understandable even for native speaker

Appendix A

Documentation

This page contains links to the most requested sections of the API documentation.

A.1   Description of the API methods

- *GET* ../v1/api/ - **Book List**. List of all books

- *GET* ../v1/api/<id> - **Book Detail**. Details of a book
  Replace <id> with specific number of book.

- *GET* ../v1/api/section=<id> - **Section List**. List of books with similar section
  Replace <id> with specific number of section.

- *POST* ../v1/book_create/ - Add a new book*

- *POST* ../v1/book_delete/<id> - Delete a specific book*
  Replace <id> with specific number of book.

- *POST* ../v1/book_update/<id> - Edit details of a specific book*
  Replace <id> with specific number of book.

- *POST* ../v1/book_create/ - Add a new book*

* Available only for admin

References

[1] Deutsche Welle. "Kazakhstan to change from Cyrillic to Latin alphabet". In: *Deutsche Welle* (2017). URL: http://p.dw.com/p/2meJg.

[2] Айгюзель Кадир. "Как выглядит новый казахский алфавит на латинице". In: *Sputnik* (2018). URL: https://ru.sputniknews.kz/society/20180220/4668066/kazahstan-alfavit-latinica.html.

[3] Аделя Васильева. "Каким был переход казахского алфавита на латиницу". In: *Today.kz* (2018). URL: http://today.kz/news/kazahstan/2018-02-20/760477-kakim-byil-perehod-kazahskogo-alfavita-na-latinitsu/.

[4] Olzhas Auyezov. "Alphabet soup as Kazakh leader orders switch to Latin letters". In: *Townhall* (2017). URL: https://townhall.com/news/world/2017/10/26/alphabet-soup-as-kazakh-leader-orders-switch-to-latin-letters-n2400623.

[5] Eileen Gardiner and Ronald G. Musto. *The Electronic Book.* The Oxford Companion to the Book. Oxford University Press, 2010.

[6] Wikipedia. *WikiBilim.* URL: https://en.wikipedia.org/wiki/EPUB.

[7] Statista.com - The Statistics Portal. *Number of mobile phone users worldwide from 2013 to 2019 (in billions).* URL: https://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide/.

[8] Wikipedia. *EPUB.* URL: https://en.wikipedia.org/wiki/WikiBilim.

[9] Michael Foord. *HOWTO Fetch Internet Resources Using urllib2.* URL: https://docs.python.org/2/howto/urllib2.html.

[10] Leonard Richardson. *Beautiful Soup.* URL: https://www.crummy.com/software/BeautifulSoup/.

[11] Aleksandar Erkalović. *ebooklib.* URL: https://github.com/aerkalov/ebooklib/wiki.

[12] Joshua Goodwin. *Zip the epub again.* URL: http://electricbookworks.com/kb/creating-epub-from-indesign/after-indesign-export-to-epub/zip-the-epub-again/.

[13]   Django Software Foundation. *Meet Django*. URL: `https://www.djangoproject.com/`.

[14]   Wikipedia. *JSON*. URL: `https://en.wikipedia.org/wiki/JSON`.

[15]   Wikipedia. *Android (operating system)*. URL: `https://en.wikipedia.org/wiki/Android_(operating_system)`.

[16]   Heberti Almeida. *FolioReader*. URL: `https://github.com/FolioReader/FolioReader-Android`.