Bilkent University
Department of Computer Engineering

CS 491- Senior Design Project

Project Name: Papyrus

Project Specification Report

Group Members: Ant Duru, Atay Kaylar, Enver Yiğitler
Supervisor: Fazlı Can
Jury Members: Shervin Arashloo, Hamdi Dibeklioğlu
Innovation Expert: TBD

1. Introduction

In the process of learning a new language, understanding what one reads has an important role. However, reading a novel in a foreign language is not something easy to do for a beginner since there are so many unknown words and phrases. Papyrus is an application to help language learners to overcome this problem. Papyrus is an e-reading platform that provides assistance to language learners by a lot of natural language processing algorithms and various tools. Papyrus is inspired by Kato Lomb, who was one of the first simultaneous interpreters in the world. Kato Lomb presents a method to learn new languages by emphasizing the process on reading. This method tries to extract all the information of a language by reading books or passages. Papyrus adopts this method to help the users for learning languages.

1.1 Description

Papyrus is an e-reading platform for language learners. In Papyrus, three versions of a book will be able to be found: The original book, the translated books, and the audiobook. There are three modes to read or listen in Papyrus. In the first mode, the users will be able to find a book in the language that they want to learn, and in their native language. In this mode, users can view how the original book is translated at any time they want to see the translations of words or sentences. In the second mode, the book and the audiobook are matched, and a narrator is reading the book while the system tracks the narrator and highlights the sentences that are read. This will help readers to comprehend the correct pronunciations of the words. In the third mode, the user can just listen to the audiobook to test their listening skills. Apart from these features, vocabulary is one of the most important

aspects of learning a new language. To help the users to enhance their vocabulary, Papyrus offers a system to create flashcards to learn and practice words using spaced repetition algorithms. When creating a flashcard, the meaning of a word appears according to the respective context. The user's pronunciations of flashcards are checked and corrected. Papyrus offers alignment services for users' private files to be read in the application. So that user is not constrained with the Papyrus Market. Overall, Papyrus will be an essential tool for language learners and readers.

1.2 Constraints

1.2.1 Implementation Constraints

- Git and Github will be used for version control.
- Mobile development will be written in native mobile programming languages. Swift and Kotlin. (Cross-platform frameworks may be used too.)
- The backend framework is not decided yet. Microservices and cloud services are the most probable candidates.

1.2.2 Ethical Constraints

- Books that are in the market will not be pirated.
- User information will be private and be hashed in the server.

1.2.3 Economical Constraints

- The app will be free to use but the books shall be purchased.
- Cloud services may be used.

1.2.4 Sustainability Constraints

- Feedback from the users will be evaluated constantly to improve the app.
- Using aligned data, NLP models can be improved or developed from scratch for additional NLP tools.

1.2.5 Language Constraints

- The app will be available in various languages.
- The books will be available in various languages.

1.3 Professional and Ethical Issues

Since we are providing an environment to read books in various languages, copyrights of the books will be respected. This means that no pirating will occur in our application.

Secondly, users' personal information, their reading preferences, annotations, highlights, and selected words will be kept to improve the recommendation system and overall application usage. However, no information about any user will be shared with third parties.

To work more professionally, every member in the group will make a contribution to the project. All group members will attend all the meetings within the group and the supervisors.

2. Requirements

2.1 Functional Requirements

- Users should sign up for the application to use it.
- The user will be able to buy a book from the market and download it
- The user will be able to create flashcards.
- The user will be able to highlight and annotate parts of books.
- The user will be able to look up the translation of a word instantly.
- The user will be able to see and listen to the aligned text/audio of a part of the book.
- The user will be able to apply NLP algorithms (sentence alignment, forced alignment) on a new text.

2.2 Non-functional Requirements

- The application will have an easy to use interface
- The application will be able to be used on iOS.
- User information will be secured.
- The application will require an internet connection only when the user downloads a book.
- Alignment services for users' own files will be available. After uploading the files, the aligned version will be kept at servers and can be downloaded at a suitable time.

3. References

1.Alfin, K. A. (2016). ***Kató Lomb, Polyglot: How I Learn Languages.*** **UNIVERSITY of WISCONSIN–MADISON.https://african.wisc.edu/kato-lomb-polyglot-how-i-learn-languages-berkeley-tesl-ej-2008/**

2.Feng, F., Yang, Y., Cer, D., Arivazhagan, N., & Wang, W. (2020). Language-agnostic bert sentence embedding. ***arXiv preprint arXiv:2007.01852*****.**