

Intro matt

For our project we've decided to build a text reader with a Maori vocabulary training component. It will allow users to open a variety of text file formats (txt, ePub etc) and it will replace a portion of the words with their Maori translation with the facility to hover over the word to see the original. It will have a library of public domain eBooks for users to read. We see this as an option to replace the users default text reader, something akin to adobe pdf reader but for all text.

The software will come with extended learning features included, and will utilize the diglot weave approach which involves reading text in the users native language then slowly replacing words with the language which they wish to learn. In this case the focus will be on replacing head words with their maori counterparts. It will cycle words in and out to increase the vocabulary retention, and have the ability to mark words as learned. The app will provide the ability to swap out the dictionary mappings to different languages etc english to spanish.

Our team consists of Jet, Jake, Matt & Cody ***Experience of each team member?***

Jake

I am a third year computer science student with interest in low level programming and design. I have some experience in natural language processing. I am well practiced in working with teams from my time as a chef, however this is the first time building an app as a team from scratch.

Jet

I'm Jet. I am almost finished undergrad computer science and am starting honours next year. I have experience with C++ in the context of graphics and computer vision. In addition to this I have designs skills and some artistic ability which I think will be valuable for this project.

What datasets are we using?

For our data we need a high quality English to Maori Dictionary, and repository of public domain books. The latter is easily decided, as Project Gutenberg offers a fantastic database of books to draw from. The dictionary is maybe not quite as straight forward, as although many exist, they are not always up to date and well digitised. We propose to get around this by drawing from as many datasets as we can and using it to inform our final dataset for translating. Some notable sources of this data are Te Ngutu Kura, the Maori spell checker

How will we build it?

Can add diagram of what I think it'll look like. Mention C++, Ultralight etc as possible technologies.

Can mention the challenges we need to overcome to have a working app. Can discuss

Look at roles our team members will take. Talk about Agileish (with that hint of waterfall) which will lead into our schedule.

We have made a rough diagram to show the architecture for the project diagram

As seen in the diagram we are trying to split the application into three main areas:

- Backend API
- Controller
- GUI

Backend API

This is further split into three different sections:

- User Profile
- Translator
- Substitute

We aim to keep each part of the backend isolated, in that it will send and receive data through the controller abstraction layer

Controller

This will handle the data going between different areas of the code. The GUI will send text and any updates to user data to the controller, which will then make the API calls to coordinate the backend. Once the backend has finished its computation, the controller will then send the result to the GUI.

GUI

From a UI design perspective we are aiming to have a low clutter and simplistic design aimed to be easy on the eyes for long reading sessions. We have made some mockups in figma of some of the features that we envision for the app. These mockups will then be transferred to html and css using the ultralight framework as a html renderer. Figma itself allows for the designs to be exported to html and css, though there will also be some clean up on our end. Ultralight

allows us to make an app using a UI framework we understand, html, but be able to easily couple it with C++.

landing_page

reader

As seen we will have a landing page from which you can open recent books or documents. Additionally there is a sidebar which will show a typical file system to navigate to files, or optionally look through the users library.

What is our schedule? cody

- gantt

Gantt charts. Include hard deadlines such as assignments and end of semester + mid sem breaks.

What apps already exist?

We found in our research that this idea has been looked at many times before. The core concept of our app has a basis in research, with the best known promotion of this concept being the Diglot Weave technique as it became to be known, start with R. Burling in his “outlandish” 1968 paper. Other similar techniques have been known as the “Clockwork Orange” technique coming from research in the late 90s that had noticed those who read A Clockwork Orange learned and retained Russian slang words very well.

As for commercial ventures there are some notable examples. Prismatext (prismatext.com) cites the Diglot Weave technique as a core component of its technology. It has public domain books that have been edited by translators to slowly transition into a variety of different languages. Although it itself does not offer the ability to open your own text, and it does not appear to offer Maori as an option.

We also found that the Clockwork Orange Technique was being used, rather amusingly, to write children’s books by onethirdstories (onethirdstories.com). They focus largely on offering physical books and only aimed at younger readers. They also have a fairly limited language selection, Maori not being on the list.

We also found Weeve (<https://shop.weeve.ie/>) which offers both an app and physical books for sale. This one is of particular interest as it is an Irish company that offers texts in Irish, which faces similar challenges to adoption as Maori does. Ultimately, they rely holy on preprepared text to offer their translations, and don’t have the ability to open your own.

Other more promising searches from a feature point of view do appear, though often lead nowhere. We have loomVue(<https://loomvue.medium.com/loomvue->

sneak-peak-320c6e385dba) that seemingly offered an extension to do this but it doesn't appear obvious if this is still around or indeed ever made it to market. They did however take a machine translation approach and dealt with alignment issues.

Through all of this, we find a mixture of the features we like, yet nothing that is: free, open source, and offers Maori as an option.

Is there interest in our app matt

Measuring interest is not exactly straight forward. The fact that there are commercial companies based around a similar proposition suggests that there is interest in the core idea of mixing a second language into a first language's texts. Although it is also notable that none of them do this with machine translation on a commercial scale.

We also believe that we offer something unique, that is the ability to read ones own text, and that the language that is on offer is Maori.

There is also a rather different cultural argument to its use in New Zealand, as New Zealand English (NZE) has a strong propensity to use Maori loanwords in place of their English counterparts. Data from the 2021 General Social Survey (GSS) showed that the ability of New Zealander's above the age of 15 to speak some Maori had increased. Between 2018 and 2021 the proportion of people able to speak more than just a few words or phrases in Maori had increased from 24% to 30%.

The same survey showed that 3 in 5 New Zealander's think that Te Reo should be a core subject in primary school. This suggests that there is at least some demand for Te Reo resources

Summary