Silvia Santamaria Buil

Introduction:

The Problem:

Coding has become an essential skill for everyday life, yet, unless you take an interest in computer science, it is a very difficult skill to learn. In order to make learning to code more accessible, I will create a programming-language learning app that makes learning to code easy by providing engaging and short lessons on coding from the user’s mobile device, so they can learn to code from anywhere.

My client is Mr Tibble, the head of computer science at my school, who would like the app made to ensure that all students have good coding skills regardless of whether they continue to take Computer Science at GCSE or A-Level. The target audience is everyone, as knowing how to code is essential to thrive in modern society. I feel older generations will benefit equally as much as younger generations from my app, as it will help to close the generational technology divide.

The Solution:

My solution to this problem will be an app, based on MFL learning apps, but for programming languages. It will involve gamified exercises that will encourage the user to learn more and spend more time on the app. It will include a streak feature, to encourage players to learn every day, a diamond feature, and a lives feature so that users try harder to get answers correct. They will also be able to compete against other players in weekly leaderboards and get followers/follow people and do ‘friends’ quests’ with them. The UI will be easy to understand so that users of all ages can feel welcome within the app.

The lessons that teach the languages will have many different types of exercises within, as well as a code editor so that the users can test their coding skills within the lessons in the app. I believe that this will help to solve the problem of people not knowing how to code as it will engage users and teach them coding languages.

My project will be limited to only having 1 full course in Python and another smaller HTML and CSS course. Due to memory constraints my app will also be limited to fewer players at first until more memory can be allocated.

Research:

Similar products:

Duolingo:

Duolingo is a mobile application and website that teaches languages (as well as maths and music in recent updates) through the ‘Duolingo teaching method’, which focuses on interactive, personalised and gamified lessons. The backend was originally written in Python, before being rewritten in Scala to improve performance. The frontend was then written in Swift for IOS and Java for Android. The session-generating algorithm uses predictions from their LLM ‘Birdbrain’, to generate exercises that are at an appropriate difficulty for the user, using a system similar to the ‘ELO-rating system’ used in chess. Furthermore, the topics covered in lessons also follow the spaced-repetition teaching algorithm to maximise memorisation.

With 97.6 million monthly users to date, Duolingo attracts users with its free services and bright and colourful UI, as well as its mascot, Duo the owl. After becoming popular in meme culture for the menacing notifications the mascot sends to encourage users to complete their lessons, the app’s social media has become very popular and played into the jokes. Duolingo is the main inspiration for my project, as it has begun to teach subjects other than languages, like maths and music, and so I would like to see if it would be an effective approach for learning programming languages!

Codecademy:

Primarily powered by Ruby and React, Codecademy teaches a huge amount of coding languages to the users of the platform, as well as other important computer science concepts, like cybersecurity and machine learning. Described as an ‘online coding bootcamp’, the website offers many free courses, but the majority are paid, requiring a subscription service to access them. Despite this, the free courses still provide great overviews of different languages, allowing them to be 100% self-studied, letting the user’s decide when and where and how they want to approach their own learning experiences.

Most popular for their web-editors, they provide interactive courses, allowing users to test out the code from their website as soon as they have learnt it. The courses involve a quick lesson page, introducing the key concepts, and then an area to code a task based on the content just learnt. There are also some quizzes when completing topics. For my project, I intend to use an in-app editor for some of the exercises I use, as well as a similar method of teaching the concept before jumping into the questions, since that method seems best for programming languages which can not just be translated into English like MFL languages in Duolingo.

Algorithms needed:

Spaced Repetition Optimization algorithm:

A diagram of a graph

Description automatically generatedSpaced Repetition is a technique used to optimize memorisation. Study of certain concepts is scheduled with increasingly large gaps between revision of the concept to ensure that the concept is committed to memory. This is because, after learning a concept, memory of that concept will decay quickly until it is forgotten. If you revise that concept the next day, it will take longer for that concept to decay; if you revise that concept again a week later, it will take even longer for that concept to decay and so on. In order to help the memorization of different concepts in the programming-language that the user will be learning, I will implement a Spaced Repetition Optimization algorithm.

This will be done by modelling the decay and then computing the optimum time for future revision of the material. The information it will use to model this is:

* The user’s accuracy of this concept during review
* The time taken for the user to recall the concept during review
* How many times the user recalled the concept in reviews correctly in a row
* How long the current interval since the last review is
* How late is the user in reviewing the material vs when it was scheduled to be reviewed

The algorithm will then calculate the user’s ‘strength’ out of 100 for each concept using the above variables. These variables will be stored in a variety of databases storing data about the user used to optimise their app experience. Using this strength, the algorithm will then compute what date this material should be reviewed. This will be implemented through ‘review’ lessons that will pop-up, providing past mistakes or difficult content based on the concepts that need to be reviewed. The results from this lesson will alter the ‘strength’ for each concept, and thus the algorithm will compute the date when the concept needs to be reviewed once more.

Difficulty-ability rating algorithm using stochastic gradient descent:

A graph of a graph

Description automatically generated with medium confidenceThe ELO rating algorithm is a method for calculating the relative skill levels of players in games like chess or Esports. However, it has also been shown to be useful in machine learning algorithms. For example, in my program I will be using a stochastic gradient descent on 2 parameters: the difficulty of the exercise, and the ability of the user, for each exercise completed. This results in a generalisation of the ELO rating system, however, instead of having their ability increase when a player wins a game and the opponent’s decrease, my app will have the user’s ability decrease when they get an exercise wrong and the difficulty of the exercise increase. There can be dramatic changes in these ratings, caused by users with low ability ratings completing exercises with high difficulty ratings, but if a user with a high ability rating completes an exercise with a low difficulty rating, the ratings will not change much.

Stochastic gradient descent is an optimization algorithm that finds the model parameters corresponding to the best fit between the predicted and actual outputs. It is a form of gradient descent that is much faster than traditional gradient descent; traditional gradient descent uses all of the data, whereas stochastic gradient descent uses randomly selected very small samples of data (sometimes even only 1 piece) to calculate the line of best fit. The predicted difficulty values along this line then let the algorithm know which exercises to place next in the lesson for the us er to complete. Stochastic gradient descent is important to use in this situation instead of gradient descent as it will take much less time and therefore is a reasonable algorithm for calculating these predictions after every exercise.

Hash tables + Hashing:

A computer screen with text and numbers

Description automatically generatedHash tables are data structures that create maps between keys and values. They use a hashing algorithm to calculate the index a certain piece of data, or ‘key’, should be placed into in the table. This immensely speeds up the search for any item of data in the table as the ‘key’ can be put into the hashing algorithm to calculate its index, and then the data at that index in the table can be found immediately. Hashing algorithms are a calculation that takes place on the ‘key’ to provide the index of the ‘key’ in the table. Often hashing algorithms use one-way encryption, which increases the security of the table. Unfortunately, sometimes hashing algorithms can lead to collisions where the hashing algorithm produces the same index value for multiple different keys. These collisions can be handled using a variety of methods, such as rehashing, linear probing, and separate chaining. Separate chaining is the best of these methods, inserting all elements that hash into the same slot index into a linked list. This, however, leads to another list that may have to be search through, and some insecurity around the key. A way to reduce the time taken to search through the second list, it could be sorted every time another value is appended to it and then a binary search could be used to find the correct value.

In my project, I intend to use hashing to store the passwords of the users of my application, storing the hashed value in a table with the username. This code is an example of a hashing algorithm I could use, as well as a simple hash table created from a static 1D array, however, as my project progresses, I will use a class to model my hash table instead.

Cross-Table parameterised SQL and Aggregate functions:

Cross-tabulation organises a dataset in table format. The data is organised by specific table headings, and then only shows 1 type of data for these specific table headings. Aggregate functions perform calculations on values, like SUM(), and output a singular value.

In the example shown, I created a small example table called ‘languages’ to show the exp that each user collected. By using a ‘CASE’ method, the table has been organised by UsernameID. This is the specific table heading organising the table. The 1 type of data shown for this heading is the amount of exp earned, grouped by the language they earned this exp in, as well as the total exp in the last column of the table.

In my project, this method will be used to calculate the total exp points for each user: by week for the weekly leaderboard, by language course and in total.