



THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA

UNDERGRADUATE ENGINEERING HONOURS THESIS

**Predicting Responses to Spaced Repetition Flash Cards
with Machine Learning Techniques**

Author:
Jordan WEST

Supervisor:
Dr. Mark SCHULZ

5 November, 2012

Contents

1	Introduction	2
2	Background	2
2.1	Machine Learning	2
2.1.1	Neural Networks	2
2.1.2	Support Vector Machines	2
2.2	Spaced Repetition	2
2.3	Similar Projects	2
3	Goals	2
4	Methods and Materials	3
4.1	Experimental Design	3
4.2	Ethical Clearance	3
4.3	Software Design	3
4.3.1	Requirements	3
4.3.2	Tools	3
4.3.3	Data Entry	6
4.3.4	Screen Mockups	6
4.3.5	Spaced Repetition Algorithm	6
4.3.6	Data storage, formatting and output	6
4.4	Data Analysis and Prediction	6
4.4.1	R Programming Language	6
4.4.2	Usage Data	6
4.4.3	Forgetting Curves	6
4.4.4	Prediction of Recall	6
5	Results	6
5.1	Online Learning Environment (Membit)	6
5.1.1	Overview	6
5.1.2	Bugs and Issues	6
5.2	Usage Statistics	6
5.3	Forgetting Curves	6
5.4	Prediction of Recall	6
6	Discussion	6
6.1	Evaluation	6
6.2	Potential Future Work	6
7	Acknowledgments	6
A	Participant Information Sheet	8
B	Participant Consent Form	9
C	Original Wireframes	10

1 Introduction

test

2 Background

2.1 Machine Learning

2.1.1 Neural Networks

2.1.2 Support Vector Machines

2.2 Spaced Repetition

Supermemo 2 Algorithm (SM2)

$$I(1) := 1 \tag{1}$$

$$I(2) := 6 \tag{2}$$

$$I(n) := I(n - 1) \times EF \tag{3}$$

$$EF := EF + (0.1 - (5 - q) \times (0.08 + (5 - q) * 0.02)) \tag{4}$$

2.3 Similar Projects

Memrise

Memrise is a private company which produces web-based flashcard software.

The Mnemosyne Project

Mnemosyne is open source spaced repetition software collecting anonymised data from its many users in order to evaluate the effectiveness of the implemented spaced repetition algorithm [4]. Mnemosyne uses a modified version of the Supermemo algorithm. The project does not appear to have produced any papers or research publications at this time.

Anki

3 Goals

Develop an online spaced repetition learning environment

Visualise student usage of the learning environment

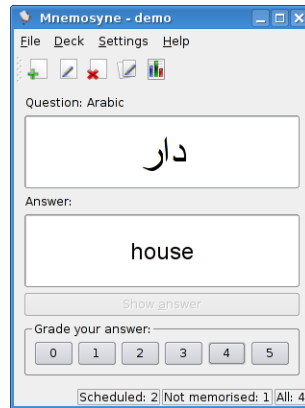


Figure 1: Screenshot of Mnemosyne in use

Predict student responses to flash cards

4 Methods and Materials

4.1 Experimental Design

4.2 Ethical Clearance

As with any project involving humans, the details of the project must be reviewed and approved by the University Human Ethics Committee prior to any student participation.

An application for review was submitted in June and approved with modifications on 25 July 2012 in time for the second week of semester.

The application included details of the methods of data collection, recruitment of participants, and approval by a ‘Gatekeeper’ who provides access to participants - in this case the course coordinator of JAPN1023, Dr Yuriko Nagata.

Data was to be stored anonymously and securely. In order to ensure participation was anonymous, cards containing unique codes were to be handed out randomly to participants to allow them to register online. Student review data was tied to a unique number in the database which could not be traced back to individual students. Email addresses were collected from students in order to allow them to log in and to reset their password if required, however exported review data was stored only against a unique number in the database which could not be traced back to individual students. Participants were also given an information sheet (See Appendix A) and a consent form (See Appendix B) to sign and return before receiving a registration card.

4.3 Software Design

4.3.1 Requirements

4.3.2 Tools

This section outlines the software tools that will be used for the project and reasoning for choosing these tools.

Git and Github (<http://git-scm.com/>), (<http://www.github.com/>)

Git is a distributed version control system (VCS) which tracks changes to source code (often amongst multiple developers) and keeps a complete history of changes. This is invaluable when a change in code occurs that results in a critical bug. Versions can be compared to find the change that introduced the bug, and production code can be reverted if need be [5].

Git repositories can be hosted anywhere, however Github offers free Git repository hosting for open source projects. It also allows users to 'fork' public repositories to create their own version of a project. For this reason it is useful for research projects as the project can be picked up and continued at any time by others.

Git was selected for this project because of its portability (moving repositories between servers is trivial). Github was chosen as it is free, encourages collaboration and is also the tool of choice for the Centre of Educational Innovation in Technology [6] under which this project was completed.

Ruby on Rails (<http://www.rubyonrails.org/>)

Ruby on Rails (aka Rails) is a popular open source framework for developing web applications[1]. Rails was originally extracted from a commercial application (Basecamp by 37Signals) to create a generic application framework [2] written in the Ruby language. Rails is designed for rapid development and provides many guidelines which the developer is recommended to follow in order to speed up development. Additionally, as an open source project Rails has gained many developers who have contributed back to the community by sharing reusable components (known as Ruby Gems) with the community. This means many pieces of functionality can be used in a project without rewriting, speeding up the development process. Gems used in this project include:

Prawn Provides PDF output - used for generating registration code cards

CanCan User authorisation - Allow and deny access to users based on their role (participant, administrator, teacher)

Highcharts-Rails Adds the Highcharts library to the application (See section below)

Heroku (<http://www.heroku.com/>)

Heroku is a private company offering hosting for Ruby on Rails applications with automated deployment. While deploying a Rails application on a server normally requires system administrator knowledge and a significant amount of time to install, Heroku allows deployment via Git and automatically installs dependencies to get an application up and running in less than a minute.

Heroku was chosen over a private server for this project since it was necessary to be able to push updates to the live application quickly in order to respond to bugs and to reduce time spent finding faults in the server.

Backbone.JS (<http://www.backbonejs.org/>) Backbone.js is an open source Javascript framework providing a model oriented structure for web applications. It was selected because...

Highcharts (<http://www.highcharts.com/>)

Highcharts is a commercial Javascript framework which provides graphing capabilities to web sites. Highcharts allows free usage by non-commercial projects. Highcharts was selected for graphing usage statistics on the website because of the features it provides in addition to recommendations on websites such as Stack Overflow [3].

Twitter Bootstrap (<http://twitter.github.com/bootstrap>)

Twitter Bootstrap is a set of default styles for websites and web applications, provided as open-source by Twitter. Using Twitter Bootstrap rapidly speeds up theming of a web application with default looks for navigation, buttons, text and layout.

See figure 2 for a comparison of default styling with and without Twitter Bootstrap

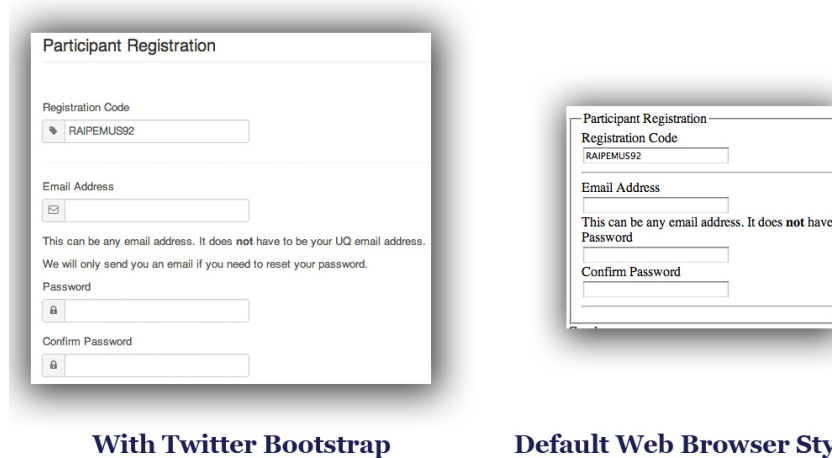


Figure 2: Comparison of a page with no styling and Twitter Bootstrap default styling

More significantly, Twitter Bootstrap offers a ‘responsive’ layout system which provides a reduced screen size (ie. smartphone) layout with little to no extra work on the part of the developer. This means a smartphone version of the web application could be designed at the same time. Twitter Bootstrap was also chosen for this reason.

The R Programming Language R is an open source programming language designed primarily for statistical computing.

- 4.3.3 Data Entry
- 4.3.4 Screen Mockups
- 4.3.5 Spaced Repetition Algorithm
- 4.3.6 Data storage, formatting and output
- 4.4 Data Analysis and Prediction
 - 4.4.1 R Programming Language
 - 4.4.2 Usage Data
 - 4.4.3 Forgetting Curves
 - 4.4.4 Prediction of Recall

5 Results

5.1 Online Learning Environment (Membit)

- 5.1.1 Overview
- 5.1.2 Bugs and Issues

5.2 Usage Statistics

5.3 Forgetting Curves

Generated from Recorded Reviews

Generated from Machine Learning Algorithms

5.4 Prediction of Recall

Support Vector Machines

6 Discussion

- 6.1 Evaluation
- 6.2 Potential Future Work

7 Acknowledgments

This thesis would not have been possible without the support of my supervisor Dr. Mark Schulz whose input and guidance has been invaluable for the project.

I would like to thank Dr Yuriko Nagata for her assistance with this and for allowing me to introduce the software to her students.

Dr Michael Harrington of the School of Languages and Comparative Cultural Studies for his advice with the online learning software, particularly the suggestion to record the time students take to review individual words.

I would also like to thank the UQ CEIT team for their help with this project - particularly Alan Cody and Phil Long.

Last but not least I want to thank my friends for their encouragement and for putting up with me throughout the project, and Mum and Dad for supporting me

References

- [1] M. Bachle and P. Kirchberg. “Ruby on Rails”. In: *Software, IEEE* 24.6 (Dec. 2007), pp. 105–108. ISSN: 0740-7459. DOI: 10.1109/MS.2007.176.
- [2] Cloves Carneiro Jr. and Rida Al Barazi. *Beginning Rails 3*. 1st edition. Apress, 2010.
- [3] Torstein Hnsi. *JavaScript Chart Library*. 2011. URL: <http://stackoverflow.com/a/1826573/80393> (visited on 10/26/2012).
- [4] Peter Bienstman. *Principles | The Mnemosyne Project*. 2012. URL: <http://www.mnemosyne-proj.org/principles.php> (visited on 08/23/2012).
- [5] Scott Chacon. *Pro Git*. 1st edition. Apress, 2009.
- [6] John Zornig. *CEIT Project Methodology*. Jan. 2012. URL: <http://ceit.uq.edu.au/content/ceit-project-methodology> (visited on 10/20/2012).

A Participant Information Sheet



School of Information Technology and Electrical Engineering

HEAD OF SCHOOL

Professor Paul Strooper

Jordan West

Telephone 0438518251

Email

jordan.west@uqconnect.edu.au

CRICOS PROVIDER NUMBER 00025B

Participant Information Sheet

Machine Learning and Spaced Repetition Systems for Predicting Foreign Language Vocabulary Test Scores

Investigator: Jordan West, Undergraduate Engineering Honours Student, Centre for Educational Innovation in Technology

Supervisor: Dr. Mark Schulz, Associate Director, Centre for Educational Innovation in Technology

This study involves monitoring your use of learning software for memorising vocabulary for JAPN1023, from which we build a model of your vocabulary knowledge. By building such a model, we are expecting to predict with some accuracy your future score in a non-assessed vocabulary recognition quiz.

The learning software is web-based and can be accessed from any computer with a modern web browser and internet connection; including from the University library or from your home. You can use the software at your leisure; there are no minimum requirements to participate however the software will be more effective at helping you memorise vocabulary if used for at least a few minutes each day. The study will span the whole of Semester 2, 2012; however you are free to choose how often and for how long you use the software.

In order to gather data on your knowledge of words, a spaced repetition flashcard system is used. Spaced repetition is a method for memorising facts at increasing periods of time and aims to provide the most efficient method of memorisation. The software has been pre-loaded with vocabulary specifically for JAPN1023, however the potential risk of participating is that you might change your study habits to incorporate this software and miss other important content. It is recommended that you use this software as an additional tool to assist your study, and not as a replacement. Participation in this study is expected to help you memorize the vocabulary, however does not constitute a replacement for your normal class study.

Your email address will be collected when you register online, however this will only be used to allow you to login and to send a password reset email if you forget your password.

Your participation in this study is completely voluntary and will not affect your grade in JAPN1023. Participation is anonymous – you will be identified only by a unique code handed out randomly upon your consent. Data on how you use the software will be collected and stored confidentially and securely and in a form such that data cannot be linked with any individual. The teacher will have access to aggregate information on the class as a whole; however will not have access to information about individual students.

You may withdraw from the study at any time, either by logging in to the system using your code and password, or by contacting me on the details listed at the top of this page. Upon withdrawal, your account and all associated data will be deleted.

This study has been cleared by one of the human ethics committees of the University of Queensland in accordance with the National Health and Medical Research Council's guidelines. You are of course, free to discuss your participation in this study with project staff (contactable on 0438518251). If you would like to speak to an officer of the University not involved in the study, you may contact the Ethics Officer on 3365 3924.

If you have any difficulties, questions or concerns about the study, feel free to contact me.

If you would like to learn the outcome of the study in which you are participating, you can contact me at the email above or write your email on the consent form and I will send you an Abstract of the study and findings upon completion.

School of Information Technology
and Electrical Engineering

The University of Queensland
Brisbane QLD 4072 Australia

T (07) 3365 2097

F (07) 3365 4999

Email enquiries@itee.uq.edu.au

Internet www.itee.uq.edu.au



B Participant Consent Form



School of Information Technology and Electrical Engineering

HEAD OF SCHOOL

Professor Paul Strooper

Jordan West
Telephone 0438518251
Email
jordan.west@uqconnect.edu.au
CRICOS PROVIDER NUMBER 000258

Participant Consent Form

Project Title: Machine Learning and Spaced Repetition Systems for Predicting Foreign Language Vocabulary Test Scores

Investigator: Jordan West, Undergraduate Engineering Honours Student, Centre for Educational Innovation in Technology

Supervisor: Dr. Mark Schulz, Associate Director, Centre for Educational Innovation in Technology

Participant Name: _____

- I have read and understand the Participant Information Sheet for this project.
- I understand that my participation is voluntary and that I will not receive any benefit for participating.
- I understand that I may withdraw at any time without penalty.
- I am participating with the knowledge that data will be stored securely, confidentially and anonymously.

Signature of Participant: _____ Date: ____/____/____
(Day) (Month) (Year)

If you would like to be notified of updates and outcomes of this project, please add your email address below to be added to the mailing list:

Email Address (optional): _____

School of Information Technology and Electrical Engineering The University of Queensland
Brisbane QLD 4072 Australia
T (07) 3365 2097 Email enquiries@itee.uq.edu.au
F (07) 3365 4999 Internet www.itee.uq.edu.au



C Original Wireframes

Wireframes/Screen Layouts

Machine Learning and Spaced Repetition Systems for Predicting Foreign Language Vocabulary Test Scores

Register

Thank you for participating in this project. I hope you find this software useful in memorising vocabulary. If you would like to read more about Spaced Repetition, [see this article](#)

You can find the Participant Information Sheet with information about the study and contact details [here](#)

Participation in this project is completely voluntary and you may withdraw at any time.

To create an account and start revising vocabulary, please fill out the following information

Account Code (This is the code written on the card you were handed after giving consent to participate)

Password

Password again

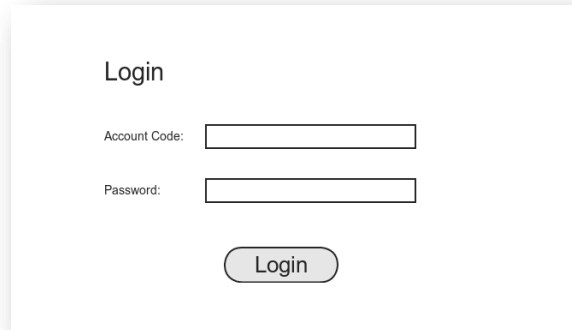
Gender: ☐ Male ☒ Female

Is English your first language? ☒ Yes ☐ No

Note:
This is the only questionnaire information that will be asked of participants.

Register

Figure 1 - Registration Page. After participants have signed a consent form, they are randomly handed a card with a web address and a unique code written on it. When the participant accesses the web address, this page will be displayed requesting information from the participant.

A login form with a title 'Login', two input fields for 'Account Code' and 'Password', and a 'Login' button.

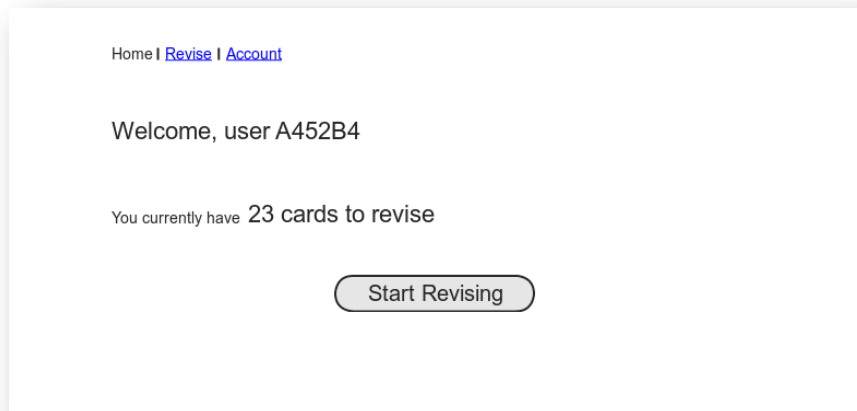
Login

Account Code:

Password:

Login

Figure 2 - Login Screen

A home page layout with a navigation bar, a welcome message, a card count, and a 'Start Revising' button.

Home | [Revise](#) | [Account](#)

Welcome, user A452B4

You currently have 23 cards to revise

Start Revising

Figure 3 - The Home Page. Displays the number of cards currently due for revision.

[Home](#) | [Revise](#) | [Account](#)

天気予報

Type the meaning of this word

Figure 4 - Revision Screen 1

[Home](#) | [Revise](#) | [Account](#)

天気予報

Correct Meaning: Weather Forecast

Your Answer: Weather Fore-cast

Was your answer correct? How easily did you recall the answer?

Incorrect	Correct - Easy	Correct - Difficult
-----------	----------------	---------------------

Figure 5 - Revision Screen 2

Congratulations

Congratulations, you have finished your cards for today.
Tomorrow there will be 16 cards due for revision.

[Return Home](#)

[Log out](#)

Figure 6 - Revision Complete Screen

Fill in the blanks

[Help](#)

	Japanese	English
Q1	天気予報	<input type="text"/>
Q2	季節	<input type="text"/>
Q3	<input type="text"/>	Temperature
		...
Q28	強い	<input type="text"/>
Q29	変	<input type="text"/>
Q30	<input type="text"/>	Young

Submit

Figure 7 - Quiz Screen

[Home](#) | [Revise](#) | [Account](#)

User A452B4 - Account Settings

Participation in this study is completely voluntary. Should you wish to withdraw from the study you may do so without penalty at any time by clicking the "Withdraw Participation" button below.

Note that if you do so, all data associated with your account will be immediately deleted and will not be included in the final study. You will no longer be able to log in to the system.

If you have any questions regarding withdrawal, or any difficulties using this function, please contact Jordan West at jordwest@gmail.com

Withdraw Participation

Return to Home Page

Figure 8 - Withdrawal of Participation Screen

You are about to withdraw participation from the study

All data associated with your account will be deleted and you will no longer be able to use the system.

Are you sure you want to do this?

Yes, withdraw participation

No, go back

Figure 9 - Withdrawal of Participation confirmation