



School of Computing final year project

Matthew Anthony James Hawkins

INDIVIDUAL PROJECT (ENGINEERING) (PJE40)

Project Initiation Document

A flashcard web application to help efficiently learn Japanese

Project Initiation Document

1. Basic details

Student name:	Matthew Anthony James Hawkins
Draft project title:	A flashcard web application to help efficiently learn Japanese
Course:	Software Engineering
Client organisation:	BritVSJapan.com

Client contact name:	Matthew Anthony James Hawkins
Project supervisor:	Petronella Beukman

2. Outline of the project environment and problem to be solved

Who is the client? What do they do?

The client is my small side business, BritVSJapan.com which gets around 15,000 visits a month.

BritVSJapan.com is a website which originally started off as a blog for me to discuss my experience with learning Japanese. It now provides advice for new comers to the language and aims to provide advice on language learning based on my experience and research into language acquisition, that hopes to make a significant cut in the time it takes for one to learn a Japanese or any other language.

What is the problem? Why does it need to be solved?

With the recent boom in web and mobile technologies in the past 10 or so years we have seen a massive rise in the amount of learning services available online. Some of these services are directly aimed at language learning, while others are aimed at learning almost anything, however out of the hundreds of services that are currently available, only a small portion of these services are built upon research on how we learn, and in the terms of languages, how we acquire foreign languages.

Consequently, we now have an abundance of applications that just aren't that efficient for people to learn foreign languages with.

One of the main pieces of advice that you will find when looking up "how to learn a language" online is to use a flashcard software that implements "Spaced Repetition" as these systems are known to be effective for meaningful retention (Ausubel et al., 1965, p. 147).

However, after using one of these systems, Anki, for 3+ years I, and many others, have noticed a few issues with the software that make studying languages slightly harder than it should be.

Most of the popular flashcard applications that implement spaced repetition algorithms on the market are not directly aimed at language learners (they tend to allow you to study for just about anything), meaning that they are not necessarily setup in the best way for an easy and efficient language learning experience.

If these applications fixed these issues and based development of their applications on research on language acquisition, then language learners would be able to produce better results in a quicker time period.

As a language learner myself I know that it takes a very long time to learn a 2nd language to a substantial degree of fluency (where one could work and live in the target country), so anything that can help to make the process easier or that allows one to progress faster is incredibly valuable.

3. Project aim and objectives

What is the overall aim of the project?

The overall aim of the project is to allow language learners of Japanese to study more efficiently via simple flashcards. The application will provide some pre-made cards but will allow and encourage students to create their own cards too.

To be able to do this some research into language learning, spaced repetition and other learning techniques will be necessary.

To meet this aim one of the objectives will be to adjust, build upon or even fix issues with current spaced repetition algorithms to develop a better algorithm and my own flashcard software that's specific for language learning and that is more efficient than other apps on the market.

Any research that suggests and certain learning techniques are great for language learning will be noted and if possible, these will be turned into requirements to implement in the system.

The application will achieve these aims by simulating physical flashcards using a question-answer pair.

4. Project deliverables

As there is a high chance of the functions and features of the application changing throughout the process, a spiral model style development process will need to be implemented. Due to this the artefacts used throughout the process may have to be changed and updated during the process.

The first deliverable will be a literature review on the subject of software aided language acquisition. The findings here will have an impact on the requirements of the overall system.

As I will be implementing the spiral model approach to this project I will have to consider 4 key stages in each spiral. These phases are planning, risk analysis, engineering and evaluation. In each of these phases there will be several documentations which will include:

Various prototypes will also be produced throughout the project. All these will be delivered including the final program.

The next deliverable will be the report where the system and how it was produced will be discussed in more detail. In the report, a variety of documentation will be included to help aid this. At this current point in time this includes:

- System requirements specification (user requirements and functional/non-functional system requirements)
- Design documentation (system architecture, use case, sequence diagrams, GUI mock ups etc.)
- Testing documentation (unit testing, stress testing and usability testing)

5. Project constraints

Time is a large constraint especially due to the nature of this type of application where in theory it could be constantly maintained and updated to increase "learning efficiency". There is only so much that can be done in the timeframe available.

Lack of research in the domain of language acquisition may be a constraint. There are lots of conflicting research and studies.

Sometime will be spent learning about the technologies that are required to use to produce the application.

I am not familiar with JavaScript, so some research into popular frameworks for building web apps and how to use these frameworks will take time away from development.

6. Project approach

Using the starting points for research that I have gathered in this PID I will begin researching for the literature review. The process for this will start off with the resources that I have already gathered but I will try to only reference papers or books that are reputable (in a reputable journal, published by a reputable university etc.) This means I will have to investigate what publication sources I should be researching. I may also discuss other publications that are cited by a lot of other publications as these tend to be the ones that are more well-known among the language learning community and could give some insight into some problems and their solutions.

The research for the literature review will be on language acquisition and software aided language learning. This research will affect how certain functionality and features will be implemented and will cause the original requirements to change and develop over time. These features will need to be planned and designed and then a spiral model development process will need to be used to get a prototype released and tested with an actual user base as soon as possible. After testing the first prototype with users, information such as how many flashcards, words, pieces of grammar etc., as well as how well they have retained this information will be recorded and analysed. Any changes to the system can be written up into requirements and a new iteration can begin. When the prototype of the next iteration is created then the data gathered from previous tests can be compared to see how effective, or not, the system has become in reaching its overall aim.

I will be using the spiral model throughout the development process to produce multiple prototypes so that each version can be compared with each other in terms of whether they reach the requirements. I've also decided on the spiral model as the requirements are unclear and complex at this stage. As time goes on and as research is done, these requirements will become clearer which means that I would be unable to take a waterfall approach to this system.

The application will be written in HTML, CSS, JavaScript, PHP and will connect with an SQL database to store user data, progress and flashcard data ready for scaling to mobile and desktop applications as well in the future.

HTML, CSS, JavaScript, PHP and SQL knowledge are required to develop this web app. While I have knowledge of all of these, I am not particularly strong with JavaScript and haven't yet used any of the commonly used frameworks that are out there for building web applications. I will have to work out first which framework will be best suited to this application and learn how to use the framework before and during the development process.

7. Facilities and resources

User testing maybe done online but it could also be done in person in which case some form of public space with some computers would be required. This can be done at the university at certain times when certain classrooms aren't in use. Ideally testing will be completed over the internet.

Web hosting is required for hosting the finished site/software and database. This has already been acquired.

A computer and the relevant software, such as a programming IDE is required. This has already been acquired.

8. Log of risks

<https://docs.google.com/spreadsheets/d/1yk0W34RsJVDnsudRCgsK2rwdacpCFQISBTxrG8ABdHw/edit?usp=sharing>

UP769535							
A flashcard web application to help efficiently learn Japanese							
Status	Risk Impact	Risk Probability	Risk Description and Type	Risk Impact to Project	Indicators and Causes	Mitigation	Solution if Risk Occurs
Open	High	Low	Loss of all project data	Project has to be started from scratch or from earlier start point. Completely wastes time and will add huge pressure and stress causing a bad project.	Hardware/software failures. Failure to create regular backups or pushes to Github.	Regular saving, backups and use of Github to prevent complete data loss.	Resort to most recent backup of work. Note down as much information that can be remembered as possible and attempt to add back into project. Re-do work that has to be done again.
Open	Medium	Low	Loss of web hosting	No live version means no testing can be done.	Run out of funding. Infringement of web host's ToS. DDOS or similar attacks.	Keep track of funding. Read web host's ToS before hosting project. Consider security risks.	Attempt to talk with hosting service to re-instate hosting. If this fails then go with a different provider. If funding runs dry then get free hosting, or resort to a local environment for testing as well.
Open	Low	Low	Web hosting downtime	No live version means no testing can be done.	Contact from web hosting company about downtime.	Have a flexible schedule.	Adjust schedule.
Open	Medium	Medium	Risk of not meeting the schedule	Work isn't completed on time and affects the amount of work that can get done within the given timeframe. This can also reduce the overall quality.	External factors such as possible issues with family/friends/health or other coursework/projects. Literally any external factor that isn't related directly to the project but that could take time away from it.	Create and stick to a tight schedule with deadlines.	Take time away from external factor and put extra time into project in order to meet schedule deadlines.
Open	High	Medium	Development of "spaghetti code" i.e. producing bad code that could cause issues further into the project.	Depending on what and where this is done, it could cause minor issues or it could prevent the entire system from functioning properly, resulting in huge changes	Taking shortcuts/getting lazy when programming. Not sticking to design documentation.	Be mindful of best practices when coding. Stick to documentation.	Fix code immediately.

9. Starting point for research.

When it comes to research I will start off by using Google Scholar and the University Library to find papers related to the topic areas of interest found below:

Main subject areas to research

Language learning/acquisition

- Learning techniques
 - o Spaced repetition/spacing effect/testing effect/forgetting curve/Leitner system/incremental reading
 - o Memory palaces
 - o Mnemonic elaboration
- Input vs output

Computer assisted language learning – CALL

- Mobile assisted language learning – MALL
- Mobile learning - ML

I will be looking at a wide range of sources and will try to focus on reputable papers from the following Journals:

Journals

- The Encyclopedia of Applied Linguistics
- Multimedia-Assisted Language Learning
- Research Foundation for English Language
- Language Learning & Technology

For language learning specific papers, I won't limit papers by the date they were published as language learning is something that has been around for hundreds of years and by limiting the research I could miss something crucial. When it comes to papers related to technology, and software, for learning I will only include papers from the past 10 years (2008 onwards) to make sure the information is up-to-date and relevant.

Papers

Mobile-Assisted Language Learning

<https://onlinelibrary.wiley.com/doi/full/10.1002/9781405198431.wbeal0768>

Emerging Technologies for Language Learning

<https://onlinelibrary.wiley.com/doi/full/10.1002/9781405198431.wbeal0365>

Exploring Smartphone Applications for Effective Mobile-Assisted Language Learning

<http://kmjournal.bada.cc/wp-content/uploads/2013/05/15-1-2Kim.pdf>

Some Emerging Principles for Mobile-assisted Language Learning

http://www.tirfonline.org/wp-content/uploads/2013/11/TIRF_MALL_Papers_StockwellHubbard.pdf

Using Gamification to Enhance Second Language Learning

<https://dialnet.unirioja.es/servlet/articulo?codigo=5495904>

The Impact of Mobile Dictionary Use on Language Learning

<https://core.ac.uk/download/pdf/82156394.pdf>

A Trainable Spaced Repetition Model for Language Learning

<http://www.aclweb.org/anthology/P16-1174>

EMERGING TECHNOLOGIES FROM MEMORY PALACES TO SPACING ALGORITHMS:
APPROACHES TO SECOND-LANGUAGE VOCABULARY LEARNING

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.472.6575&rep=rep1&type=pdf>

Spaced Learning Enhances Subsequent Recognition Memory by Reducing Neural Repetition
Suppression <https://www.mitpressjournals.org/doi/full/10.1162/jocn.2010.21532>

An Investigation into The Effect of Targeted Vocabulary Learning Using a Spaced Repetition
Flashcard System on TOEIC Scores <https://www.agulin.aoyama.ac.jp/opac/repository/1000/12507/>

The effectiveness of computer-based spaced repetition in foreign language vocabulary instruction: a double-blind study https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=1120&context=engl_pubs

Computer Assisted Language Learning

<https://www.tandfonline.com/doi/pdf/10.1080/09588221.2010.520675?needAccess=true>

Principles and Practice in Second Language Acquisition by Stephen D Krashen of University of Southern California

http://www.sdkrashen.com/content/books/principles_and_practice.pdf

Ebbinghaus, H. (1885/1913). Memory: A contribution to experimental psychology. New York: Columbia University.

Second Language Acquisition The Effect of Age and Motivation by Einar Garibaldi Stefánsson

<https://skemman.is/bitstream/1946/15018/1/BA%20EinarG.pdf>

Nakata, T. (2008). English vocabulary learning with word lists, word cards, and computers: Implications from cognitive psychology research for optimal spaced learning. *ReCALL Journal*, 20(1), 3–20. Retrieved from <http://dx.doi.org/10.1017/S0958344008000219>

Nakata, T. (2011). Computer-assisted second language vocabulary learning in a pairedassociate paradigm: A critical investigation of flashcard software. *Computer Assisted Language Learning*, 24(1), 17–38. Retrieved from <http://dx.doi.org/10.1080/09588221.2010.520675>

Nation, P. (2001). Learning vocabulary in another language. Cambridge: Cambridge University Press. Retrieved from <http://dx.doi.org/10.1017/CBO9781139524759>

Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2012). How and when do students use flashcards?

Existing Systems and Algorithms

Some existing systems and technologies that will be useful to look at and compare are:

Anki

SuperMemo

SM2 Algorithm

Memrise

Leitner's learning box/Leitner system

Pimsleur method – Paul Pimsleur

Duolingo

Rikaichan

Brainscape

Cerego

Course Hero

Lingvist

Mnemosyne

Pleco Software

Quizlet

Skritter

Synap

WaniKani

Online resources

Some issues with Anki - <https://vladsperspective.wordpress.com/2017/03/14/optimize-your-anki-youre-overtesting-yourself-on-too-few-cards-make-huge-gains/>

10. Breakdown of tasks

The approach will start with setting out the user requirements while beginning secondary research into the necessary topic areas.

Based on the user requirements some designs will be mocked up and artefacts created to get a better understanding of the system and how everything will interconnect. During this process anything useful discovered from research will also be noted ready for the next spiral.

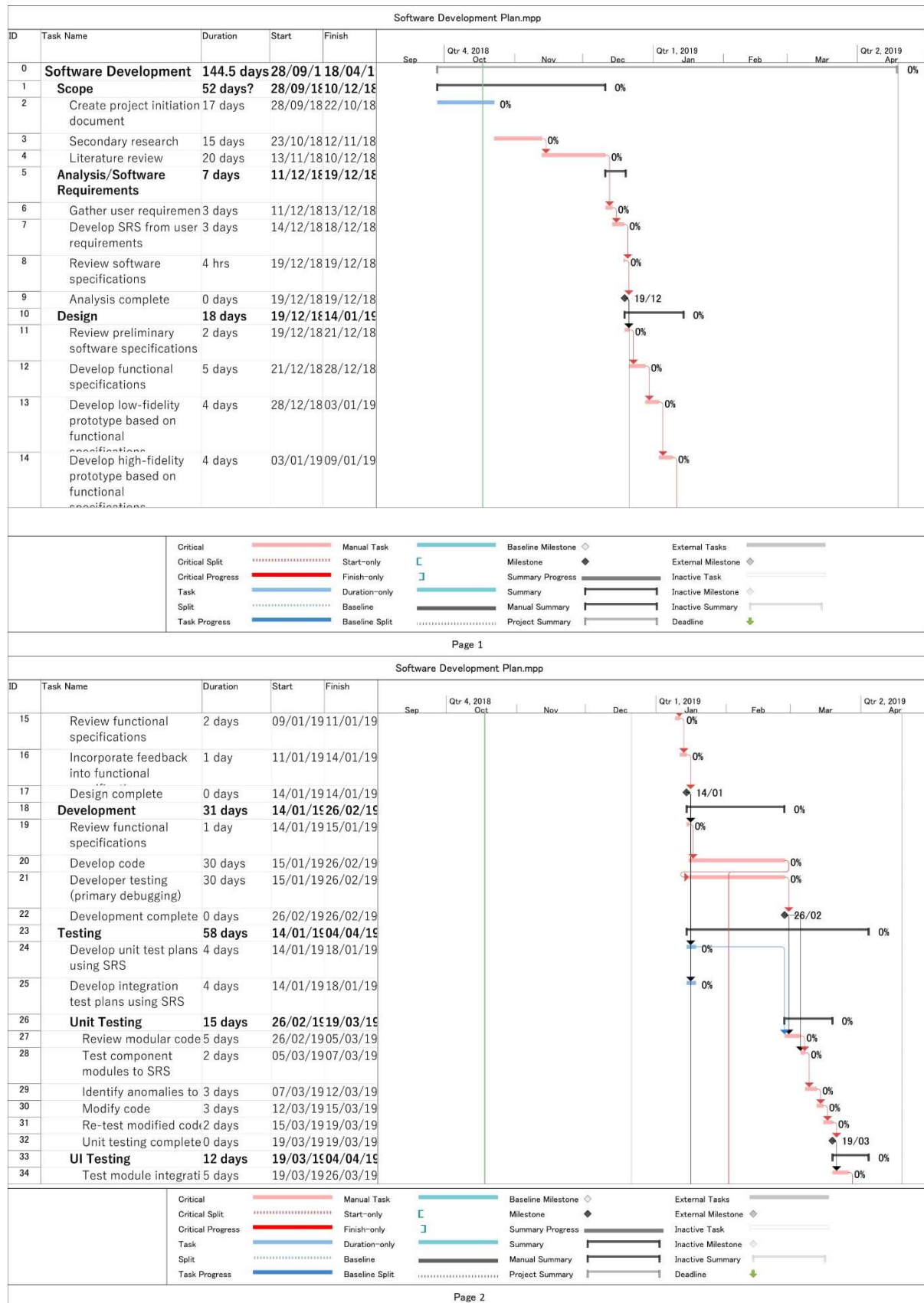
A prototype of the system will then be developed based on the current design documentation and user requirements.

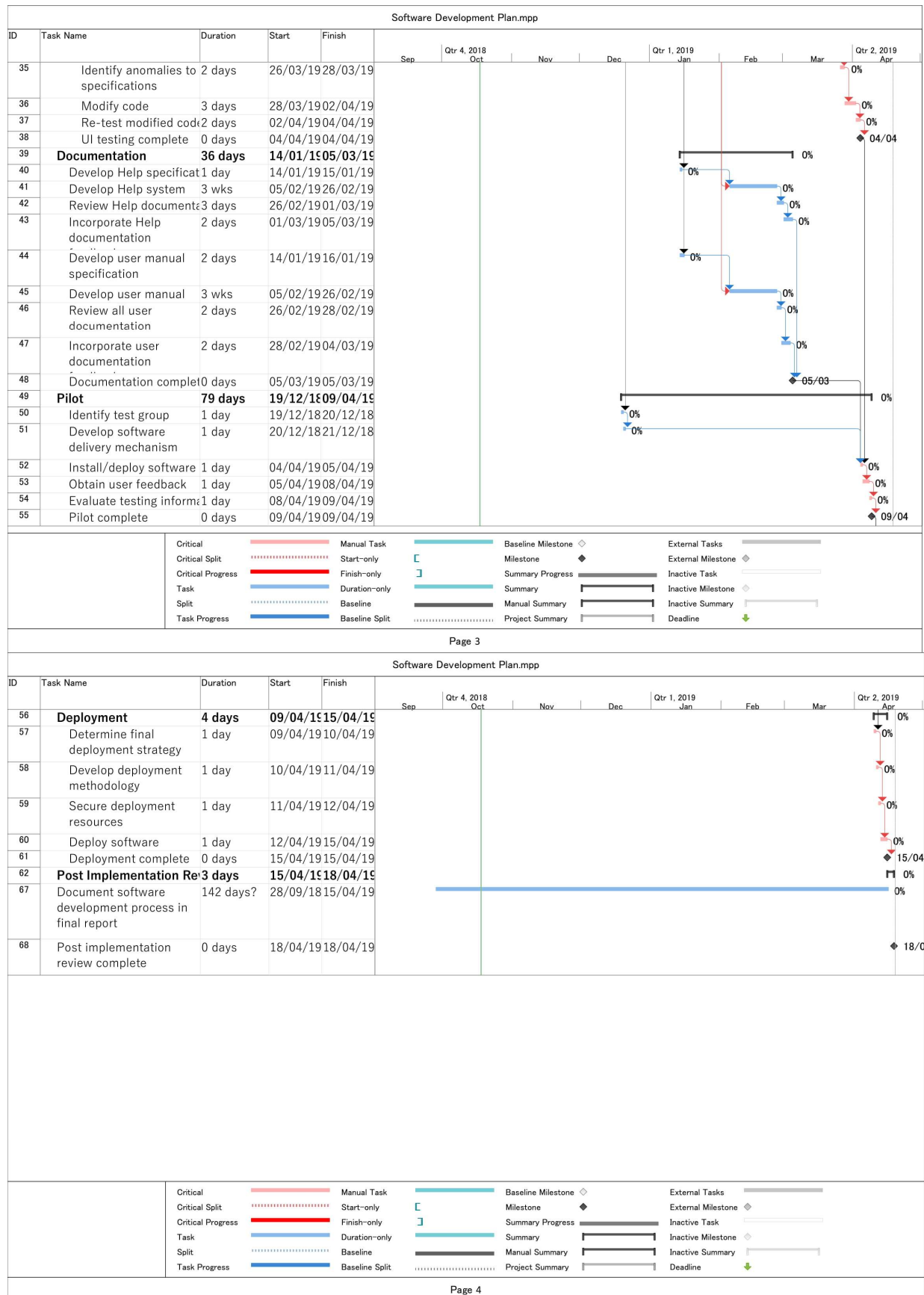
Tests will be done on this system to see how effective it is. Once data has been collected, any research or changes needed to be made to the system will be turned into requirements and the next spiral iteration will take place.

Tasks:

- Research topic area
- Gather user requirements based on research and my experiences
- Write up system requirements from user requirements
- Create design documentation (use cases, state diagrams etc.)
- Develop basic prototype based on design documentation
- Test prototype using unit testing and GUI testing.
- Gather end users to try the prototype and record the prototypes “efficiency”.
- Use the gathered data from the prototypes and from the continuous research to make changes to the user requirements.
- Repeat the steps above until a finalised application has been developed.

11. Project plan





12. Legal, ethical, professional, social issues

Multiple user tests are required to properly be able to tell whether the application has improved in terms of “efficiency”. In terms of legal issues, I will need to make sure that I handle these user’s personal data correctly as well as making sure that the system is secure to prevent any possible attacks to get users information from the system. This means preventing cross site scripting, SQL injections etc.

There won’t be any ethical issues. Any test users will be 18+ years of age and will sign a consent form telling them what data will be stored and how the data will be used for the project.

A basic ethics form has been completed.

No professional issues.

For social issues I will have to be culturally sensitive when choosing people for my test groups as there may be some people out there that may not be comfortable with learning Japanese during the group test. To prevent this, I will make it clear to anyone who maybe considering doing these tests that they will be learning a bit of Japanese on the consent form.

Appendix

Ausubel, D. P., & Youssef, M. (1965). The effect of spaced repetition on meaningful retention.

Journal of General Psychology; Provincetown, Mass., Etc., 73, 147–150.

Retrieved from <https://search.proquest.com/docview/1290528359?accountid=13268>