CSE5 ITP Application Development
Project Plan

Project Title: Mediterranean Archaeology

Team XD

19187484-Biswas Nandamuri

19171661-Hema Priya Movva

19148739-Krishna Mohan Chiluveru

Table of Contents

1.OVERVIEW

La Trobe University has researchers working in 45 specific disciplines where Excellence in Research for Australia ranked at or above world standard.

A.D. Trendall research centre for Ancient Mediterranean Studies is one among the best research centre working on the art and archaeology of South Italy and Sicily with centre's unique resources.

A.D. (Dale) Trendall was a legendary figure and one of the foremost historians of Greek art of the 20th century. He was the principal authority on the red-figure vases produced in the Greek colonies and native areas of South Italy and Sicily during the 5th and 4th centuries BCE.

Trendall Research Centre has the following objectives:

- To promote research in the general area of Ancient Mediterranean studies, particularly in the archaeology of South Italy and Sicily during the Classical period.
- To disseminate within the general community in Australia the results of the latest research in Greek and Roman art and archaeology through the sponsorship of conferences, lectures and seminars.
- To make available, at the Director's discretion, the resources of the Centre to all scholars and graduate students, whether from Australia or overseas, who wish to use the library and archive.
- To maintain and extend both the Library and the Archive (as unique research resources in Australia) through the acquisition of books and periodicals relating to Greek and Roman culture, and images of South Italian red-figure vases.

The project aim is to develop a web application for presenting data about Mediterranean archaeology to make it available online which is helpful for students, archaeologists, and many others for their reference, study or to make a research. During this project a database is created with the ancient Greek pottery images and made available along with its description and details about the pottery.

It is always exciting for students to know about the history and gain knowledge which will help them in their studies and also research work. For archaeologists, this website will act as a reference for their research where they can find all the requirements at one place. There are many people who are keen to know about the ancient products and their history, and this website will make their interest possible by giving an access and to learn more.

During this, complex data will be converted into the simple format by recognising each element and structure of the complex data. Basically, we get the data for specific fields for the image which we refine, restructure and put them together to release them online in a website and provide access to the users to study, refer and gain knowledge with the content.

The documentation, website are the main project deliverables that will be handed over to the client. It is important for the team to take care of these two deliverables and release them with all the requirements that are needed to the client. Though we have few risks, we can manage them and convert them into considerable challenges. This gives an opportunity to face and implement our ideas successfully.

To process a project and complete successfully, every team must follow a methodology to plan and proceed accordingly. In this project, as a team we have chosen agile methodology for the application development.

As an agile development process, we have a chance to provide guidance to the stakeholders on designing and understanding the website, user requirements and expectations. But as a huge task we must consider the timeframe for the project delivery and divide the tasks. For this we must train the staff properly in all the required areas to cross communicate with the other team members.

1.1 Project Objective

As stated in the overview of the project, the main objective of this project is to make sure we successfully deliver the product deliverables to client. The project deliverables are as follows: A website (with the required data), Documentation (which includes the entire work by the team).

As it is an application development process, it needs to be planned and managed accordingly with the software team and allot the tasks to the team. Follow-up is mandatory for the team lead to be on track with the deadline. And we must prepare the documentation simultaneously with the software development as it is also a project deliverable. There should be proper set of standards and guidelines that the team must follow and proceed for successful completion of the project.

1.2 Project Constraints

There are few constraints which we need to focus during the project development.

Developing a web application involves quality of the product, i.e., how user-friendly is the web application? Is it accessible to everyone? How easy is it to access? Are there any privacy concerns with the authentication details? Are there any unwanted pop-ups that ruin our privacy?
So, we need to consider each constraint as a challenge as they will be accessed by the users.

As far as the project is related to coding; development and implementation of the data and database will be quite understandable and easy to implement.

Adjustment of time for each sprint is essential, and every sprint has its responsibilities and allotment of the tasks to each team member. In our case, analysing and assigning work is done a week before the start of new sprint for all the team members. Multiple applications (Trello, bitbucket, Slack) need to be used for allotting and implementing tasks which help in the completion of the project. These applications will highly reduce the risk of complexity in the project.

We make use of Trello, where the team can manage the entire project with a visual overview of the work that is being done for the project. We as a team, should make sure to maintain the trello board as updated as possible to have less complications regarding the project. This is the place where we update the tasks, propose and schedule meetings and also set agenda for the meetings.

Bitbucket is a web-based repository which will help us in source code and development projects. We choose this to be the code repository as it has many features including pull requests for code review, merge checks, code search, issue tracking, smart mirroring.
These features are very much useful for us during the code development.

Slack is a communicating tool that creates a workspace for the project. We can upload/download files, share between the team members or to a group. This made our work easy to share important files related to the documentation, to provide research links.

As all the software applications used in this are open source files and libraries, it does not cost much. The probability of investing in web templates is high which are useful for web development which will eventually be in the last sprint. As it is the main product deliverable, this needs to be considered with highest priority. So, investment in web template is worth it and selection of the layout should be made so that it is user attractable and user friendly.

By the end of the final sprint a complete product can be expected with planned features and user-friendly application which consists of all the requirements by the client.

The webpages with history of ancient relatable will have a huge impact for few categories of people. Students, Archaeologists will get involved with these contents for their knowledge and research. If required, we can provide additional data if requested by the user, depending on type of data and also aid in understanding the contents.

1.3 Project Risks

Every possible aspect that might be the cause of a break in the development of the project is considered as a risk factor. One such issue that as a team we faced would be decoding the structure of the data.

Event Risk	Probability	Impact	Time Scale (Weeks)	Mitigation	Contingency
Matching Respective fields in the record	Likely	Significant	2 weeks (Sprint 3)	Finding different logics based on which fields can be extracted and mapped.	Understanding what each field depicts, and then mapping the rest of the fields based on the similarity of each record.
Finding Suitable OCR libraries	Likely	Minor	1 week (Sprint 1)	Browse for all the available libraries prior to the requirement.	Implementing Approach 2 i.e. using XML file.
Extracting Specific records from Textbook	Almost certain	Significant	3 weeks (sprint 3-4)	Writing code logically to separate individual records from textbook pdf.	Search for unique patterns in all the records and use them to extract required information.
Extracting individual fields from record	Almost certain	Significant	3 weeks (sprint 4)	Manually Search for similar patterns to match fields.	Search for unique patterns in all the records and use them to extract required field information.
Delivering the final product on time	Rare	Minor	-	Completion of extraction and database creation prior to the given deadline.	Request for extension so that project can be delivered completely.

2.PROPOSED SOLUTION

Solving the Data Accuracy & Font understanding issue are the primary concerns as they are considered as the major part of the project, without them being available there is no possibility to proceed further.

For these problems to get solved, as a team we decided to have two different approaches which could be beneficial. These two approaches will be developed simultaneously. The one which has a better and easy way to finish the project will be finalized for use.

The first approach makes use of a Python library, i.e., poppler which helps to convert a pdf file/document into an image. Poppler will be forming sublayers, and it helps in converting the pages to images and provides the page count. However, the issue with this is time

insufficiency for the execution of that sublayer. So, the code which helps in running this library is considered and modifications are made to it.

By doing the modifications, both the page count and the file to image processing are implemented. Though the accuracy is not fully achieved, a maximum of nearly 90% is reached.

The second approach which we choose to take forward the project is to convert pdf file into a XML file. During this, a meta data is obtained which contains font, font style, height, size in pixels.

This data helped us to segregate the individual records from the pdf book along with the individual fields. This mainly involves the brute force technique to identify the font size of the required data.

2.1 Business Requirements

Final product is the business requirement. In our case, client needs a website with pottery images and related data of the pottery and to have access to it as a final product.

Hence a software product is a final project deliverable for the client's business requirement, which contains all the required pottery with the related data so that the user can have access to it for their need.
For this purpose, Oxford university database is used, which contains fields that needs to be replicated in our website. So, we must extract individual fields from the record as per the oxford database and develop the website.

Evaluate existing processes

A.D. Trendall had written a book that contains the entire records and the pottery images along with the history of ancient archaeology. This is the existing way that the researchers are using for any references and research purpose. As an improvement to it, this book is converted into a pdf format which is used for extraction of the data and make it available online using a website.

Interviewing users regarding the issues that are being evolved timely, face-to-face discussion of what the client is willing to change, i.e., improvements, and understanding project in their terms to learn the issues and modifications that need to take place. Querying about what was the previous issue and the need to be addressed.

There is no financial data that is related to the project; all the latest software versions are used. So, there is no scope for legacy applications, requirement of pdf for data extraction is primary. If the pdf is not provided, then it can be considered as a regulatory issue.

New business rules and workflow

As a basis for application development a business analyst should consider new rules and strategies.

Once the product is released, we take feedback from the users and the stakeholders. This will need improvements and updates from the project team. So, a business analyst needs to

document the rules. This work is divided between the team members from different departments where there is a need for improvement.

The work flow between departments will be divided by the project lead after the analysis report from the business analyst.

Specific User Interface (UI) requirements

User interface requirements should be clear, and all the details must be specified in the documentation which gives the client a better view at the interface.

1.Search: This is the first requirement that an average user looks for when he/she visits the website. So, this must be very easy to identify and access.

2.User friendly: All the website developers aim for this requirement as the most important one. This will bring a lot of users to the website. So, we need to try and try until it is the best possible website which is easy to use.

3.Guide: Providing a tour at the website, contents and access ways for the user will benefit them and will make it easy to find what they are looking for. So, we need to provide this for them and ask them to leave a feedback on how better we can serve them.

4.Quick response: Customer service should be provided within the website with frequently asked questions and responses. If they are looking for something different, they can contact us via phone or email.

The response for the queries should be quickly done by the management team.

5.Privacy: Customer privacy is what matters for any owner. So, we need to be clean enough and take care of all the privacy concerns and prepare the product. We also need to provide statement of the privacy for the users regarding the website usage and their personal details (if given any).

Specific technology requirements

Pytesseract: This is a python wrapper library, which is used around tesseract-OCR for optical character recognition from images.

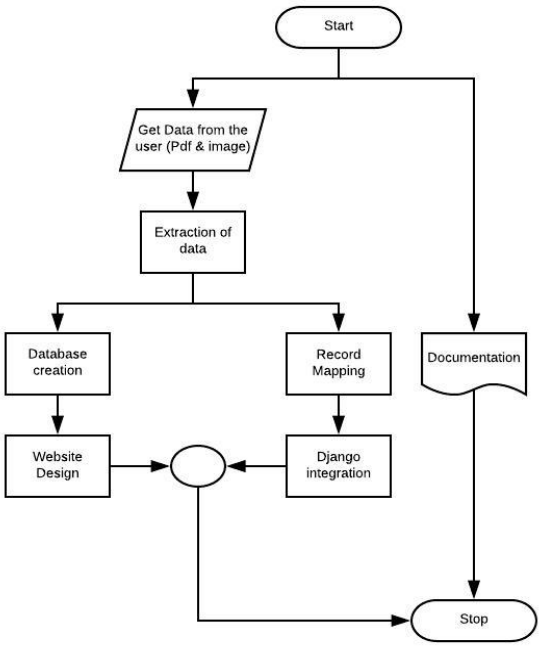
Poppler: This is used to convert each individual page in the book provided into individual images containing the data regarding the potteries.

PDFMinor.six: This is used for extraction of information from a pdf document. With this we can obtain exact location of the text in the pages.

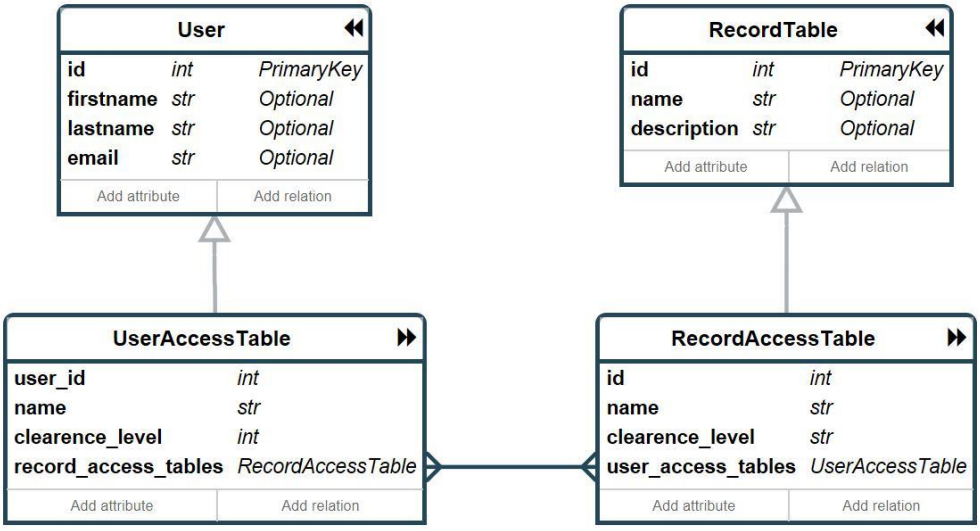
Django: Its an open-source framework, used to reduce complexity of creating websites faster and easier.

Bitbucket: We used this application to maintain the code in a secure way which is privately maintained by the team members.

2.2 Architecture



Class Diagram:



Sample Database

Functional Specifications

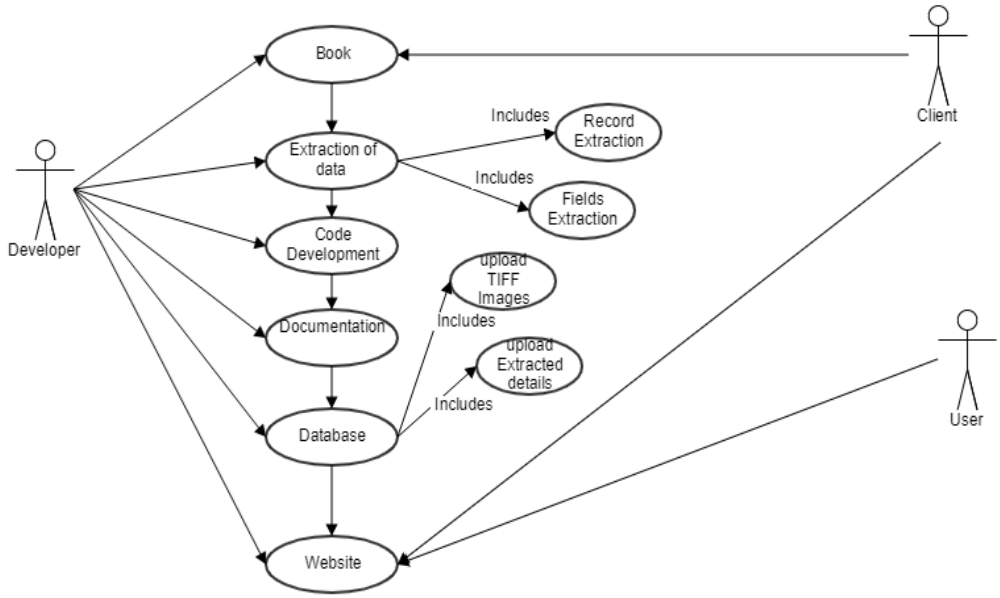
The functional specifications in software development specifies the functions that a system must perform. The purpose of this is to define the requirements that need to be implemented by the developed software.

User story, use cases, system requirements specifications come under functional specifications.

User story: This is a description of features with respect to end-user perspective. This also describes the type of user and their requirements. In our case a user expects a user-friendly website which contains the Ancient Greek pottery images with the respective information which will be one of the user stories.

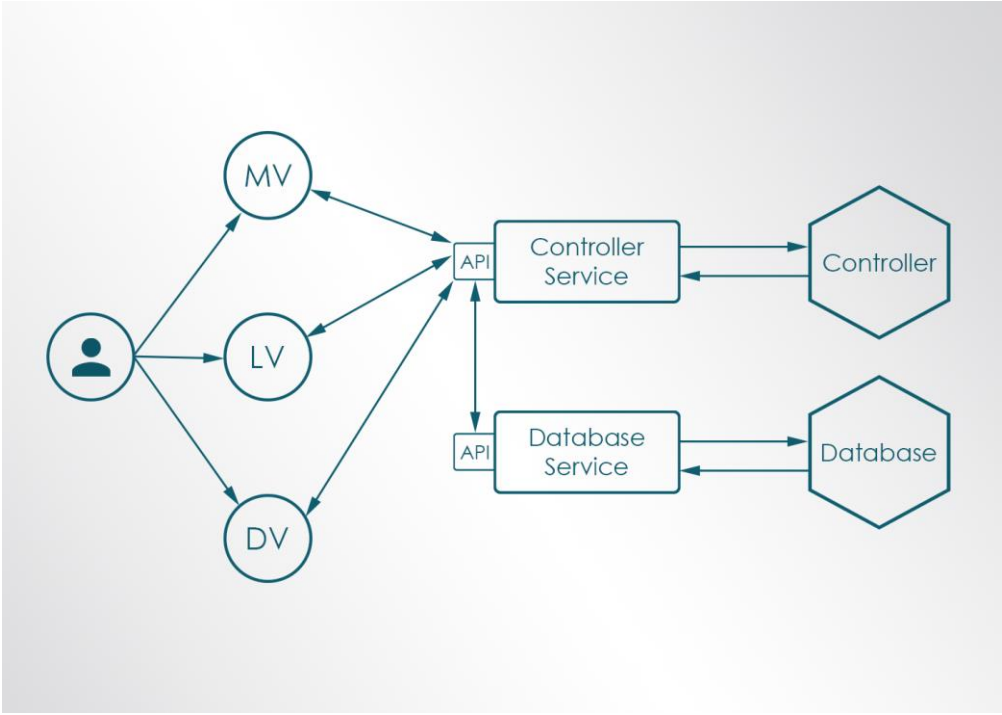
Use cases: These are basically the actions that a person or a role or an external hardware system will perform with the main system. There are several stake holders involved in our project. A client expects the product from the project team and also follow up the team. Client will also analyse how the present market is going and take steps accordingly.

Similarly, a developer will deal with the coding, testing and development process. In such way, all the stakeholders will be connected to their respective use cases which can be better explained in a use case diagram.

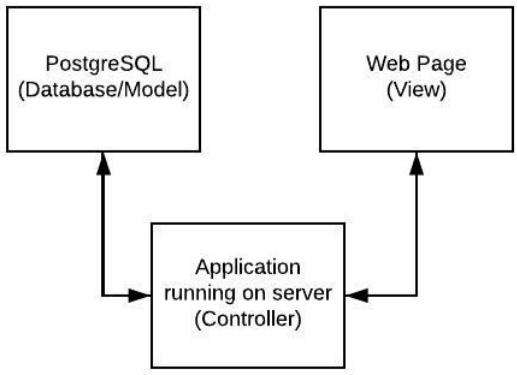


Technical Specifications

Network Diagram:



Network/Architecture Diagram



MVC Architecture

Development Languages:

Language	Python (Back-end) HTML, CSS, JS (Front-end)
Frameworks	Bootstrap, Django, Jinja v2(version 2)
Dev-Server	Python http. Server (testing), Gunicorn
Database	PostgreSQL

Peripheral Specification:

Input: Input in the form of pdf data, Respective images, File system access.

Output: Web-browser.

Platform Specification: File System Access.

Security Specifications:

This application must be a highly secured product as there is an important data of the ancient history Greek pottery. This should never be mis-used, so need to be carefully maintained.

All the data must be securely maintained. Also, the privacy for the users should be considered as a factor of security. They should be given a detail statement of how we maintain their data and provide proper response when needed.

2.3 Development

Features that are included for each release and development process of each phase are:

Development of Web applications for Static file Location.

Development of Web Application for dynamic file location.

Development of application with windows as the main server Operating System.

Development of application with both windows and Linux as main server Operating System.

2.4 Testing

Software Development is a repetitive process, and so it does continuously test the data or code that is created. In this project, testing is conducted by a specific set of people, i.e., Automated Software Testers (Continuous Integration using Pipelines in bitbucket). As Agile methodology is being used, an iterative process is followed. It can be said that whenever a commit is made, it will run a test. Involving customers at the time of testing is not encouraged in this project, so both Bug Reporting and Formal Feedback takes place. Bug Reporting process happens in a short interval of time for every commit so that if any errors occur it can be handled at the soonest possible.

2.5 Deployment

Major milestones for the project involve all the tasks that need to be accomplished for each sprint. Each sprint is scheduled for two weeks with proper agendas for each sprint. We consider each sprint to be a milestone.

SPRINT 1: Finalizing the libraries for the software development was the agenda which will be a major selection for the entire project.

SPRINT 2: Data extraction from pdf file, Copying TIFF images to database. Milestones will depend on the respective sprint. They are scheduled and implemented in an organised format.
As considered, we developed work related to second approach which involved brute force method. This really helped us to find a way to extract the records with the meta data.

SPRINT 3: Segregation of the data for respective fields and prepare an excel file which contains the entire data of the pottery that we separated from the records.

SPRINT 4: Extraction of the individual record from the pdf and data of the fields from each record with two approaches.

We finally decided to go with the second approach to extract the records which will convert the Pdf file into XML file, where the XML file holds the meta data such as font style, height, width, font size in pixels.

With the help of brute force technique, we manually compared the font size of the pdf file containing necessary and unnecessary data. This helped us to distinguish the data that we needed to form the database.

3.PROJECT RESOURCES

3.1 Roles and Responsibilities

Responsibilities	Team Members (Roles)		
	Biswas Nandamuri	Krishna Mohan	Hema Priya
Developing a project plan		Project lead	
Coordinate & Supervise		Operational lead	
Best Practices	Operational Lead		
Communication	Operational Lead		
Assigning tasks	Functional Lead		
Strategic Input		Operational Lead	Operational Lead
Gather Requirements		Process Lead	Business Lead
Process Analyst	Process Lead		Process Lead
Test solutions	Business Lead		
Establish a project schedule		Project lead	Operational lead
Development	Developer		
Documenting the Process		Business Lead	Business Lead

Project Lead – Delivering the project is one of the core responsibilities of being a project lead for the respective project. And in support to the team lead, every team member should cooperate and work in a rightful way.

Structuring proper paperwork of how the project is going to be evolved, i.e., step by step procedure of things happening and for all this to happen; it is very crucial to have a timeline for every task and time-to-time meeting to discuss the progress and upcoming tasks. A proper review of how the tasks have been handled need to be considered by the project lead.

Functional Lead – Being in a specific area and handling it is the primary task of functional leads. They can be considered as bridges who help in communication with the project lead.

Dividing and Assigning tasks is quite essential. An unorganized organization can lead to many conflicts. As to avoid the situations, it is always better to have a leader who will assign all the tasks that are required to be completed in their respective sprints. Handling problems related to team members and cross department communications should be taken care by the functional lead.

Operational Lead - Managing various team members, the productivity of the work that a team is working on, and the quality of work that is being delivered will be carried out by operational lead.

This team will know what the team members are capable of, and so they send the employee history to the functional team as they can assign the tasks. Facilitate coordination and

communication between support functions. They also help in suggesting strategic plans in development.

Process Lead – Managing day-to-day activities. This role ensures that all the process, activities are performed, and that team members are assigned with enough work. Review of the work completed takes place by the process lead.

Developer – Installation of the required software's, developing the code, accessing multiple libraries for generating the code, testing of the code using a sample test data to check the accuracy of the code, maintenance of the software and system are few activities that will be carried out by the developer.

Business Lead – Checking if adequate resources are available for the development. It is the responsibility of the business lead to oversee and supervise the team activities and also have an idea of the present market to deal with.

3.2 Issue Escalation

Accuracy rate during the extraction of the data is a major problem that need to be solved quickly. This is an important part of the result or the project outcome. The team need to focus on this problem to solve it and manage the project towards the intended outcome.

3.3 Project Staffing Plan

In an application development project, we mainly need to focus on:

Project management and planning: Carrying out an application development project needs a proper team who can manage the situations, who have an ability for decision making, plan the tasks and organize the team by proper allocation. So, we need a team who have these abilities.

Programming: Programming is a core part of an application development. So, a developer with an idea on how to take forward the project is mandatory for this.

Testing: Testing is the stage where the code developed is tested, and the application is developed. It needs to be tested with a sample data which will give us an idea on what exactly the code is doing and what improvements must be done to achieve the intended result.

Documentation: This is a part in which everything that needs to be done, which is in process and also the tasks that are done will be included. This is basically a paper work with every proper detail on how the project is going on. We can make use of this to improve few aspects i.e. progress of the project, planning, development, customization etc. so that the team can overview and focus accordingly to improve them.

Training: Training is important for the team. Every team member needs to know, how far the project is implemented, they need to be trained in the required fields which will help them during the team meetings, cross department communication.

3.4 Project Materials

Hardware/ Software	Infrastructure	Peripherals	Co-location space	Licensing
Laptops	Server machine	USB	Latrobe university	Apache license
Mobiles, iPad	Database	Hard disk	-	-
Applications: Python, visual studio code, GitHub, Bitbucket, Trello, Slack.	Client machine	NIC	-	-

Licensing: Copyrights of the work and allowing the licensee to use those rights are considerable factors for licensing.

Permission	Conditions	Limitations
Commercial use	Licence and copyright use	Liability
Distribution	State change	Trademark Use
Patent Use	-	Warranty
Private Use	-	-
Modification	-	-

4.PROJECT APPROACH

4.1 Development Model

In our development model, we developed the project with two different approaches simultaneously to achieve the result. For these two methods we made sure, we followed agile methodology. As this project needs a simultaneous development and iterative process, we decided to go with it.

Agile methodology: This uses a continuous iteration process for development and testing of the software product.

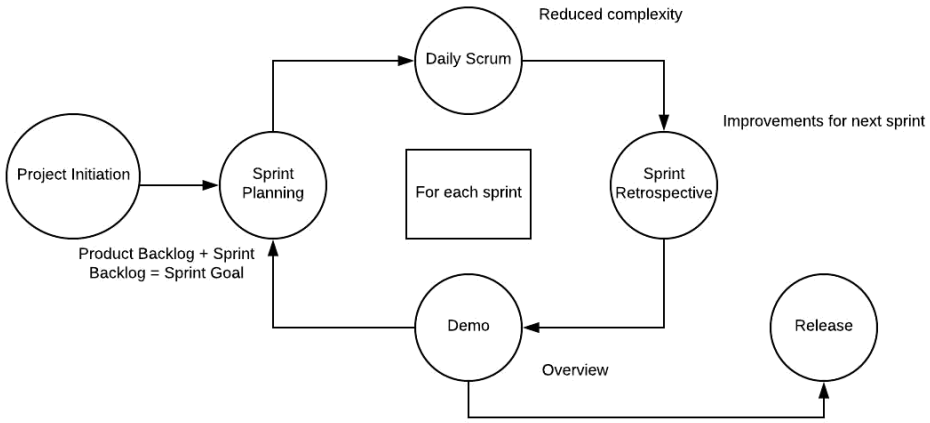
We made use of the agile methodology, which can simultaneously work on software development, webpage designing, extraction of the data, testing the developed product. This basically contains sprint planning in which the sprint backlog is chosen from the product backlog. This is iterated for every new sprint.

Daily scrum: This includes daily scrum meetings during which we discuss the progress of the sprint, this helps the team to reduce the complexity and update the progress of the work. Then we set an agenda for the next 24 hours.

Sprint retrospective: During the retrospective session, we discuss the improvements that need to be done and make sure that we don’t repeat them in our next sprints (if any).

Demo: This basically involves the overview of the sprint, where we discuss the entire sprint and the project outcome.

After the demo, if the product has gone through all the sprints and ready to be released, it will be tested to cross verify with the final intended product. If all the requirements are satisfied, then the product will be released.



Agile Methodology

4.2 Configuration Management:

Management of few items are necessary during a project. Software project includes the code which might need to be updated depending on the testing stage to get a proper result.

We will have various releases for an application development project because of the requirements from the client. Depending on the requirements on the type of software, hardware and peripherals we need to change our platforms and do the process in order to have less complications with the output.

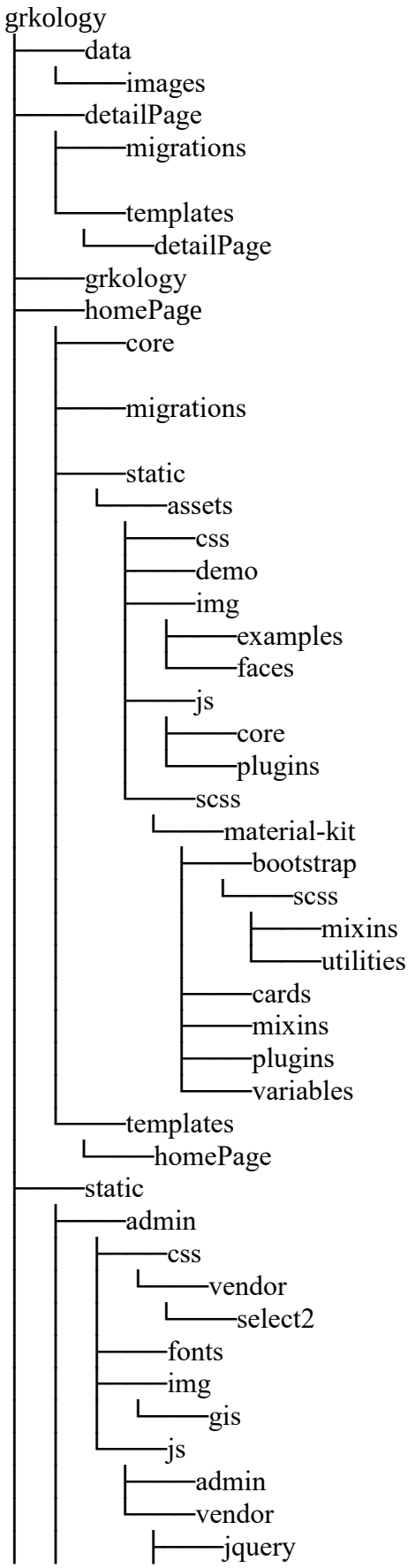
Analysis of the documentation and also approval for the submissions from the team are necessary and all the team members must know the respective changes regarding the project and the flow. Documentation of the project needs to be updated simultaneously along with the other aspects.

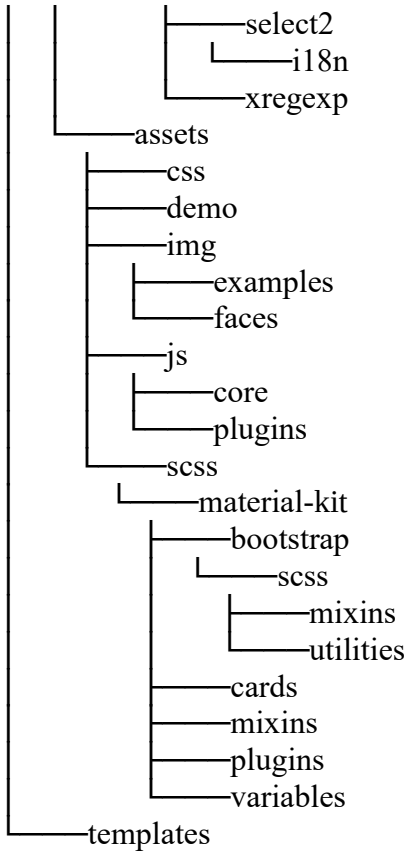
Product baseline must be built prior to the sprint and need to be tracked with the help of burndown chart. This will give us an idea of tracking the project according to the plan.

Best way to track the changes that happen in any code/project baseline could be noticed using Version control system. There are few VCS's which can be used such as CVS, GIT, SVN, Bazaar. We have chosen to go with GIT as it is a distributed VCS, free open-source software, fast and small, secure and flexible to use as users can select any workflow they prefer.

Implement code library system: Managing a large part of data is difficult as well as complex. To overcome this situation, libraries are introduced, these libraries are known as modules.

Library Structure:





Tools

In developing a code library system, we use a technique to have modules for every task. This will help us to maintain and perform the tasks at ease. As a part of agile technique, dividing the tasks is a major benefit for project processing. So, we consider the technique of modules for the code library system. Documentation analysis and approval by the team is assumed to be a technique to reduce complexity between team members and cross department functioning.

Reporting

Reporting during the project is important part of the documentation where we will be notifying the team regarding any changes that matters. Project baseline analysis report need to be discussed and also notify to implement changes (if any).

For every release after the sprint, there should be a detail report of the project status. This will give us an idea on any improvements, changes, mistakes which need to be taken care for the next sprints.

Inspect and paper the data by each team member which can be used for reporting. Audit data report should be made and reported. This gives us the information on the type of data and also the amount of data that is being used for the entire project.

Archiving

The images from which the data is extracted need to be archived until the extraction process is completed. So, we copy all the images and archive them until the code is developed. Then we

start the extraction of the content from these images. Once all the images are successfully extracted, we delete these images.

The documentation of the project is also archived as we simultaneously prepare it along with the software development. Whenever we need this to be updated, we recall the last documentation.

4.3 Communication Management

Communication between the team members to share the information regarding the project is mandatory. We can have team meetings with proper agenda and also have a review for every meeting. Attendance of the team members depends on the agenda, but reporting, changes and approvals must be made only after considering the team members approval.

We need to keep track of the status of distribution of information among the team members, it is important to have this both for the team and also for the project.

We need to schedule meetings only with proper agenda for the session and make sure to discuss all the points and also propose meetings to carry discussions in the next sessions.

4.4 Change Management

There are few change initiators in our software development which were discussed during meetings and approvals were done after the entire team is satisfied with the changes.

CHANGE INITIATOR	NATURE OF CHANGE	REJECTION/APPROVAL
1.Web page designing	Expected	Approved (by team)
2.Library selection	Unexpected	Approved (by team)
3.SQL to PostgreSQL	Expected	Approved (by team)
4.Logo designing	Expected	Approved (by team)
5. Approach 2 (Pdf to XML file)	Expected	Approved (by team)

These changes are mandatory for the project to have the intended output and all the team members should know the changes whenever made.

4.5 Testing

Testing the product will be done before each release and will have a review of the product output. We make sure that we are taking the product in correct direction. The developers analyse the product by proper testing which will minimize the issues.

In our case, the product obtained after each sprint will be tested and analysed. These testing are scheduled by the software developer and are considered to be milestones.

All the bugs that are found during the testing will be considered to have changes and update them accordingly. We prepare the feedback document after the testing is done, so that it will be easy for the developer to make appropriate changes and proceed accordingly.

4.6 Documentation

Preparing the documentation with all the details and delivering them when needed is a part of the team work. Clients expect the paper work from the project team and request to deliver them when they need it. It is useful to prepare it simultaneously rather than to do it when needed.

We need to prepare a soft copy and deliver the document in required format to the client.

It is one of the project deliverables, so review of the documentation before being delivered to the client is mandatory.

Summary of the extraction

Extraction Process

Extraction the records and fields are the main criteria of this project and for this to happen entire pdf file needs to be converted into a text format. As multiple fonts styles and inscriptions are included it is difficult to extract them using normal process. So, we have chosen to follow two approaches to find the better solution

Approach 1: Using Pytesseract OCR library to extract the pdf file into a text file. While using this library we came across many problems one such issue was accuracy. We used page Segmentation technique to overcome this by doing multiple attempts to solve the problem.

Approach 2: Converting the pdf file into XML text and extracting the entire data with that. The Same problem repeated even with this approach though it was not that intense as Approach 1. Just by making some small changes we achieved nearly 95% accuracy in this. When compared we felt Approach 2 was much more efficient and reliable to work. So, we converted the pdf to XML and the next process is carry forwarded using this approach.

Extraction of Record

Converting the entire pdf gave us a bulk of data from which we were supposed to extract records which were related to the pottery for which we are creating the database, website.

When converted into XML file we got the metadata of that text i.e. Font style, height, dimensions, distance from the start, font size and so on. We selected Font size as the element because the entire pdf was 10px but the records which we needed were of 9px of size. Using this difference, we were successful in extracting all the records.

Selecting the individual record is the tricky part as there is no pattern to extract them. We used regular expressions to find similarities between the records and we came across a few of them. But the best option was the Asterisk symbol (*) that was present before the vase/record numbers. One problem we faced was the symbol was not present for all the records.

((?=*1).*(?=*2\s)) – for the extraction of individual records

Extraction of Fields

We selected the fields and elements from the pottery database of the University of Oxford website. Fields such as vase no, fabric name, Technique, provenance, Collection name, Author name, Publication record, Date range and so on were the elements that we were supposed to extract. Fabric name and technique have some constant values for this textbook. So, we extracted vase no, collection name and provenance fields from the records using regular expressions.

Provenance: In all the records that are extracted this field is present in only a few among them. ***From*** is the common term that will be present before every provenance. When using the regular expression, we gave a command such that the next word that is present after **from** need to be considered as provenance.

\.*fro[mn](.*)Ht

Explaination:
4 Paestum IV/452, **from the area of the Heraion, Loc. IV. Ht**. (to top of vase) 14; the handles have been restored. PLATE 16 e PP-s, NF 6; PAdd, no. A 33. (a) Nude woman seated on pillar, holding phiale in 1. hand, [b) Eros running
/
gm
\.* matches the character . literally (case sensitive)
* Quantifier — Matches between zero and unlimited times, as many times as possible, giving back as needed (greedy)
fro matches the characters fro literally (case sensitive)
Match a single character present in the list below [mn]
mn matches a single character in the list mn (case sensitive)
1st Capturing Group (.*)
.* matches any character (except for line terminators)
* Quantifier — Matches between zero and unlimited times, as many times as possible, giving back as needed (greedy)
Ht matches the characters Ht literally (case sensitive)
Global pattern flags
g modifier: global. All matches (don't return after first match)
m modifier: multi line. Causes ^ and $ to match the begin/end of each line (not only begin/end of string)

Collection Name:
Extraction of the collection name is possible by considering the starting number of the record and alphabetical recognition is given as the end i.e. H, fro, G these were given to stop the search.

\s(.*),\sfro

Explanation:
4 **Paestum IV/452**, from the area of the Heraion, Loc. IV. Ht. (to top of vase) 14; the handles have been restored. PLATE 16 e PP-s, NF 6; PAdd, no. A 33. (a) Nude woman seated on pillar, holding phiale in 1. hand, [b) Eros running
/
gm
\s matches any whitespace character (equal to [\r\n\t\f\v])
1st Capturing Group (.*)
.* matches any character (except for line terminators)
* Quantifier — Matches between zero and unlimited times, as many times as possible, giving back as needed (greedy)
, matches the character , literally (case sensitive)
\s matches any whitespace character (equal to [\r\n\t\f\v])
fro matches the characters fro literally (case sensitive)
Global pattern flags

g modifier: global. All matches (don't return after first match)
m modifier: multi line. Causes ^ and $ to match the begin/end of each line (not only begin/end of string)

5. ESTIMATE

Estimation of the budget is quite low according to the development strategies that we are planning. As it is an application development project, this mainly depends on the coding, website development, testing or deployment.

With respect to time estimate, it depends on the number of sprints. We are considering 4-5 sprints with two weeks span for every sprint. So, this will take an estimate of 10 weeks to complete the project.

Software Developer	• Extraction on records and individual fields • Database creation • Website development	150 hours * $55.00/hr = $8250/ea.
Software Analyst	• Understanding Requirements • Gathering Software requirement specifications	100 hours * $43.50 = $4350/team $4350 / 3 = $1450/ea.
Business Analyst	• Documentation • Assessing business models	50 hours * $75.00 = $3750/team $3750 / 3 = $1250/ea.

Estimation of entire project is done based on the sprints assessed and the work done in each sprint by each team member, based on the roles and their responsibilities everyone is assessed and the cost is provided. Based on the calculations, the estimated cost is $10950/ea.

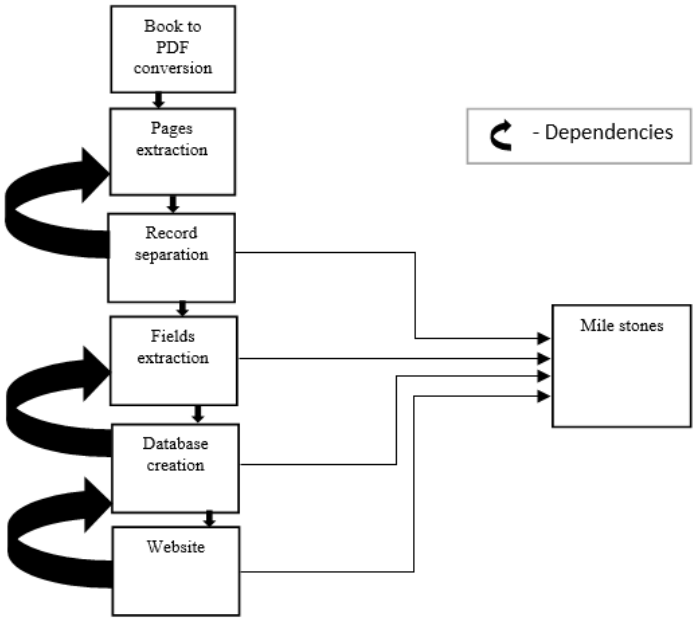
As the entire project is about building a website, hosting can be made after considering with multiple third-party platforms like GoDaddy, SiteGround, InMotion etc. Website can be launched under third party platforms where we are supposed to invest up to $10-50/month. Launching a website without third-party can cost nearly $250/month.

6.SCHEDULE

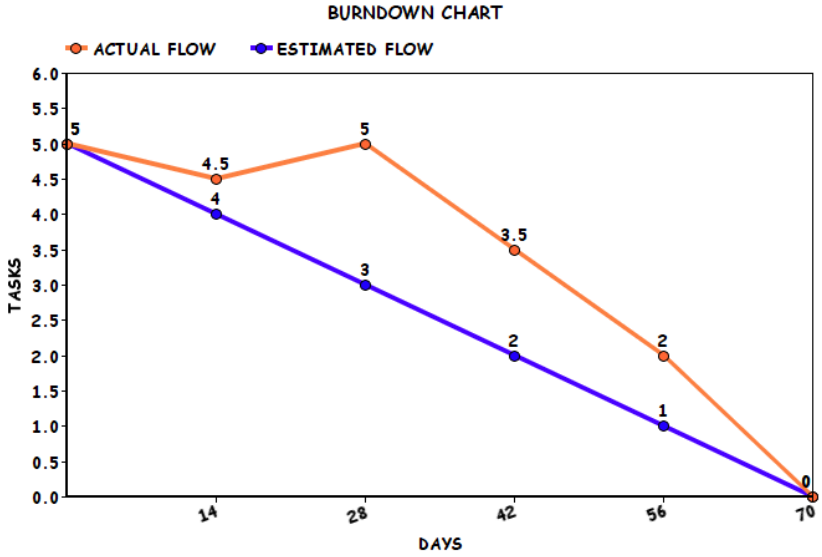
The application development takes 4-5 sprints which will be completed in 10 weeks.

Though the schedule of this development process varies depending on the sprints, but it is important for the client and the team to have an average estimate of the time to complete the project and deliver the product successfully.

Prioritizing the project tasks is one of the better options by which we can select the sprint backlog from the product backlog and develop the product accordingly so that we can finish the tasks in time and submit the product delivery successfully.



Burn-Down Chart



REFERENCES:

1. 1.https://www.latrobe.edu.au/trendall (Overview)
2. 2.https://www.onlinecharttool.com/graph (Burndown chart)
3. https://www.payscale.com/research/AU/Job=Software_Engineer/Salary
4. https://work.chron.com/much-computer-engineer-paid-per-hour-21954.html
5. https://borisfrolov.wordpress.com/2013/02/03/project-risk-management-in-tfs/
6. https://towardsdatascience.com/10-common-software-architectural-patterns-in-a-nutshell-a0b47a1e9013
7. http://www.nakov.com/blog/2011/06/29/software-architectures-client-server-multi-tier-mvc-mvp-mvvm-ioc-di-soa-cloud-computing/
8. https://pythex.org/
9. https://docs.python.org/2/library/re.html
10. http://www.cbs.dtu.dk/courses/27610/regular-expressions-cheat-sheet-v2.pdf