Carnegie Mellon University Qatar

Team Argo

(Qur'anic Botanic Garden)

Web presence and news aggregation system

Project Proposal and Requirements

Version 3.0

Prepared for: 67-373 (Software Development Project), Summer 2017

Prepared by: Abbas Ali, Abdul Wasay, Amna Al-Ansari, Umair Qazi

Professor: Selma Limam Mansar

May 18th, 2017

TABLE OF CONTENTS

REVISION HISTORY

Date	Description	Author	Comments
May 11th 2017	Version 1.0 Initial Document setup	Umair	Setup of design and structure of the document with headings
May 16th 2017	Version 2.0 Completed Sections	Team Argo	Filled in content for different sections
May 18th 2017	Version 3.0 Compilation and proofread	Umair & Wasay	Compiled different sections in one file plus proofreading

1. Executive Summary

Organizations want to get a handle on their web presence to analyze their influence and impact on the community at large. Qur'anic Botanic Garden aims to better understand and catalogue their web presence using our system, Argo.

Argo is a web presence and news aggregation system that will enable the organization to download all news posts, tweets, Instagram posts, YouTube posts and articles on the web which are related to Qur'anic Botanic Garden. Argo will also include the feature of cataloguing these posts and articles in a local database for easy search and query. This system will help Qur'anic Botanic Garden to measure their influence and impact on the web. This report presents the problem, solution, feasibility, limitations and risks of carrying out this project.

2. Community Partner Background

The Qur'anic Botanic Garden (QBG) is one-of-a-kind garden which aims to assimilate all the plant species mentioned in the Holy Quran and the Hadith. The project was introduced by UNESCO under the auspices of Her Highness Sheikha Moza and inaugurated on the 17th day of Ramadan (September 17th, 2007.)

The mission of QBG is "to promote knowledge of the plants, botanical terms, and conservation principles mentioned in the Holy Qur'an and Sunnah" (Qur'anic Botanic Garden Vision and Mission, 2015).

Since their inception, they have been working to gather all the plants and have been successful in cataloging and housing 60 plants, of which 26 are mentioned in the Quran, and rest are mentioned in the Hadith. QBG also aims to have a physical garden soon. The design of the physical garden has been approved and will be built in between Oxygen Park and QFIS in Education city. QBG also has a laboratory that houses plants not only from the Middle East but other regions as well. The organization has also done extensive community outreach, has conducted various sustainability and recycling events to promote awareness about nature and its impact on the community. It is located near Education City Community Housing and has a total of five employees.

Qur'anic Botanic Garden wants to get a sense of their web presence. It has conducted many events, whose information is relayed to QF Press, who then send it out to other news media outlets. Currently they have been gathering data about themselves on the web manually. Sally El-Sayed, Secretary at the QBG, has been doing the laborious task of searching for different terms on google hoping that results that are relevant to QBG and its events will pop up. The organization needs to produce a report of its web

presence to present to sponsors and stakeholders to sustain itself in the long run. Malik Baloch, who is the business support coordinator, has experience in networks and has also been providing IT support to the organization. Team Argo aims to assist QBG in their needs and will implement an application that would allow QBG to gather and search news and articles about them which are available on the internet.

3. Technology Infrastructure & Management

People within the organization are all using Windows machines that are connected to the internet within the Qatar Foundation network. The organization does not have an IT team and Sally and Malik are managing its technology needs. The technology infrastructure is not of paramount importance to QBG since it is a small organization with a focus to collect plants mentioned in the Qur'an. The client is willing to provide a computer or a laptop for the project and the will be responsible for any costs incurred after the project is complete and the system has been deployed.

4. Problem Analysis

QBG is currently downloading news, articles and posts manually. One of the employees, Sally, goes online to find news articles and other data about QBG and their events. She then manually copies URLs and saves them into a list. The manually downloaded articles are then catalogued in Excel or Word spreadsheets. This entire process is time consuming, difficult and laborious. In addition to this, the links that are saved sometimes expire or become inactive, thus having an offline copy becomes a necessity.

A solution that was considered by the client previously was: to ask QF press to send back

an email with a report of all the news outlets or sites that have published QBG's articles. However, this did not work out because QF press was not consistent with their feedback or gave delayed responses in sending back these reports.

There are numerous reasons why a solution to this problem would be beneficial to QBG. One of the main reasons is being able to produce a report easily and quickly, either for reporting to QF of their progress or for producing an annual report. Integral parts of this report include a section on statistics of the number of events for specific categories, and sources of news. The solution also needs to simplify the client's efforts in documentation and compilation of an archive for all their activities, presence and mentions on the internet. Furthermore, the solution to this problem would help QBG progress further as a company because they will able to carry out further analysis on the data to observe market penetration and then have a better understanding of current activities and where they lack. The solution will also enable them to target audiences for a further spread in awareness about their programs.

5. Solution Alternatives

The purpose of the project is to capture Qur'anic Botanic Garden's web presence. For this purpose, five possible solutions were discussed. The alternatives that were not chosen and the reasons why are below:

Java Application

There is a lack of official third party API support. The language is complicated, and can be hard to debug when something goes wrong. The team members do not have much experience of coding in Java.

Outsourcing work to a Marketing firms

Qur'anic Botanic Garden is a project within QF which does not have a huge budget. For now, QBG's web presence is not wide enough to warrant the outsourcing of a task that can be done for free and manually.

Open source solutions or applications

The requirements of the projects are very specific and there is no open source solution which covers the parameters and the scope required by QBG. The open source systems that are available are complicated and too broad in scope to be used by the organization.

Web Application

QBG has a small team and they require a simple localized solution. Stakeholders of the organization also do not allow hosting of web application for security reasons.

6. Requirements and Features

A solution and a framework is required that can be implemented within the given time-frame. The components of that solution must be familiar to most of the team members. For these reasons and many that follow, we chose to go with a Python application. Python is a manageable and usable programming language, with the widest support for APIs (YouTube, google, twitter, PDF conversion) required for the project. The python library is also extensive enough to encompass the scope of the project in terms of fetching, saving and viewing the posts of QBG. A system with Python also maximizes the utilization of the skillsets that each team member possesses allowing for an efficient development process. This system is also the most economically viable as it does not require the involvement of an external agency to catalog the posts that QBG requires.

Python also allows for a hybrid user interface that enables the system to display results in the best possible formats.

a) Project Requirements

i. Functional Requirements

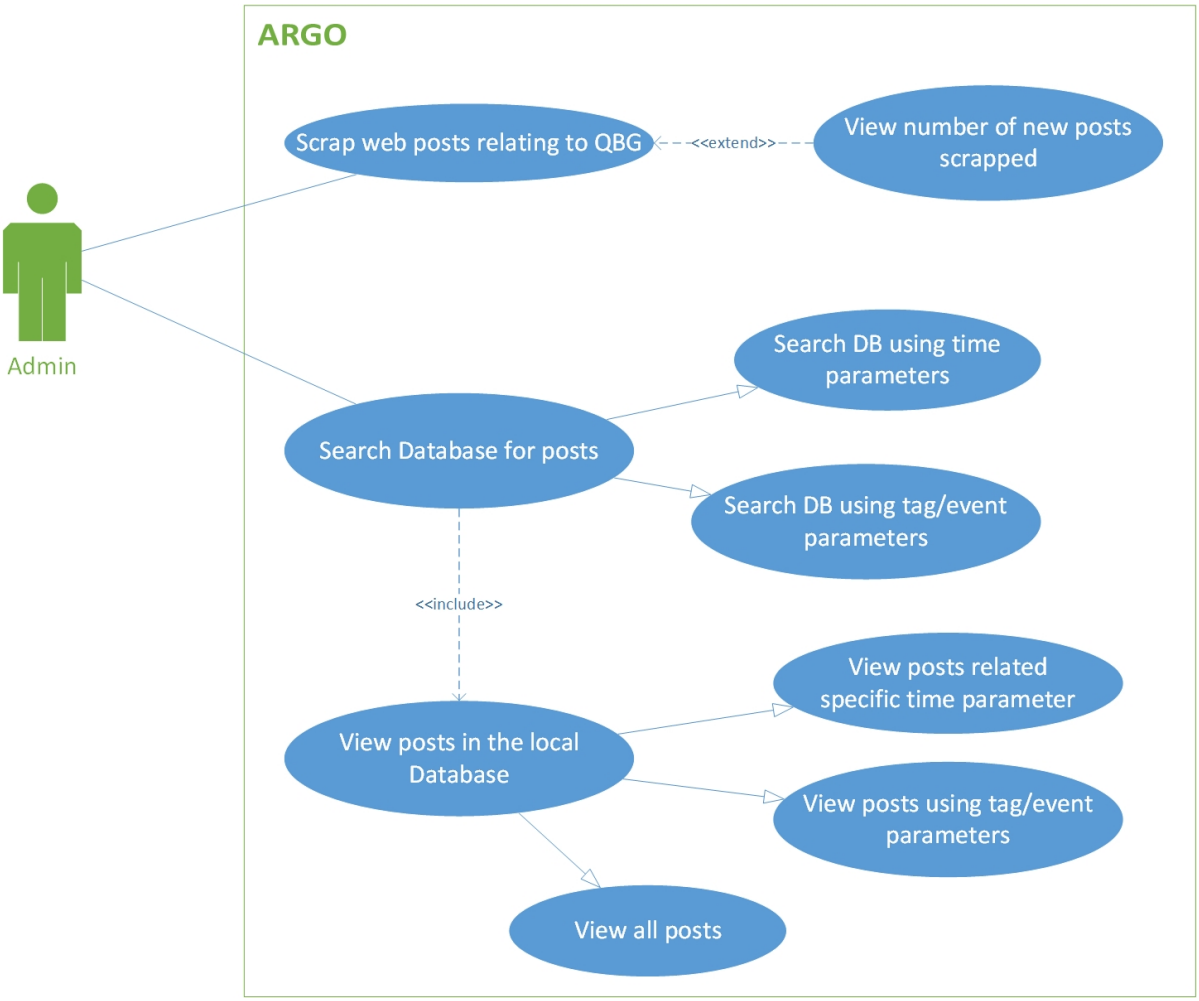
- The system shall allow the user to search the whole web for posts related to Qur'anic Botanic Garden.
- The system shall show how many new posts (compared to the database) have been found.
- The system shall allow for the snapshots/PDFs of the found posts to be saved.
- The system shall allow for tag and time search to be conducted on the local database of saved posts.
- The system shall allow all the saved posts in the local database to be viewed.
- The system shall allow the user to check the number of posts saved in the local database pertaining to an event or tag.

Feature Number	Feature	Description	Prioritization (M, S, C, W)
F1	Local Database to catalog posts of QBG	A local database that includes the posts related to QBG that are found on the web and stored locally	**Must**
F2	Active Links of Posts in the Database	Database saves the active link of the source for the post related to QBG	**Must**

F3	PDFs or snapshots of posts in Database	Database saves the PDF or snapshot of the post related to QBG	**Must**
F4	Searchable Database based on time	Local Database of QBG posts can be searched through with time parameters.	**Must**
F5	Full web scrape to get all search results to be stored in the local Database	Full web scrape using all available APIs adding records to the local database	**Must**
F6	Categorized Database	The Database assigns categories for the posts (e.g. for a specific QBG event)	Should
F7	Searchable Database based on event tag	The Database can return all the records that pertain to a particular event or tag	Should
F8	PDFs/Snapshots in the same format as online version	The snapshots and PDFs in the Database are of the same format and visual outlook as the post on the web	Should
F9	Unique entries (no duplicates) of found posts in local Database	The Database does not allow for duplicate posts after a scrape	Should
F10	Installation Guide	A guide to teach the user how to install the system on a new computer and what libraries and files are needed for its smooth operation	Should
F11	User Manual	A guide that shows the overview and flow of the system, to understand its features and how to use it	Should
F12	Login Credentials for system use	Secure login and password to limit the number of users that have access to the system	Could

F13	Manual tag addition to Database	Posts found in the Database can have event tags added to them to aid search and categorization	Could
F14	Numbers and statistics for each event category stored in Database	The number of posts in the Database for each kind of event tagged, displayed individually	Could
F15	Specific searches saved to a separate file/folder	Posts retrieved after specific searches on the web saved in a separate folder for use by QBG	Could
F16	Flash/YouTube video content downloading	Downloading of video (flash and youtube) to be saved alongside of the PDFs and snapshots of the posts found about QBG	Wont
F17	Graphs of detailed statistics of each event	Compiled statistics of number of people attending, number of views of each post and graphs showing all these as a summary for QBG to analyze their web presence	Wont
F18	Accessibility of the application on an intranet	An application running on a shared local network to be accessible by multiple machines	Wont
F19	TV/Video/Radio content downloading	The download and cataloging of radio/TV content alongside of the PDFs and snapshots for QBG posts	Wont
F20	Content from Facebook, Snapchat, private social media accounts	Posts relating to QBG on Facebook, Snapchat and private social media accounts	Wont

ii. Use Case Diagram



iii. User Stories

- As an administrator, I want to scrape the web for all QBG posts and view the number of new records found.
- As an administrator, I want to search the local database for QBG posts within a certain time frame.
- As an administrator, I want to search the local database for QBG posts with a certain tag.
- As an administrator, I want to view all the QBG posts in the local database.

- As an administrator, I want to view all the QBG posts in the local database within a certain time frame.
- As an administrator, I want to view all the QBG posts in the local database with a certain tag.

iv. User Classes and Characteristics

There will be only one user class for the system. This user (Administrator) class will have access to all the functionality available. And since the system will run on a dedicated computer, it limits the number of people (to be administrators) by limiting access to the dedicated computer.

v. Operating Environment

- The system will operate on a dedicated Windows machine with a stable internet connection.
- The python program console and all the required libraries will be available and installed on the dedicated computer for the smooth operation of the system.
- The search results will be displayed in an application interface (native to python) or a web interface using a regular web browser installed on the computer.
- The PDFs or snapshots of the posts will be available to be viewed on the same system.

vi. Assumptions and Dependencies

- It is **assumed** that the provided machine will not change significantly. The client will provide a dedicated Windows OS running machine with a capable internet connection that is needed to fetch the posts related to the organization from the internet. If the OS for the dedicated computer changes, significant changes need to be done to the system for all the libraries to work properly. If a dedicated computer is not provided, hosting the program on a private network will require significant changes to the way the system is accessed and stores data in the Database.
- It is **assumed** that the APIs will not change significantly. The OAuth2 protocols to access social media APIs is recently updated from OAuth1 and will be active, hence not requiring an update in protocols for API access. A change in authentication protocols will require a new license to be generated to allow for continued API access.
- It is **assumed** that the organization's web presence will not increase dramatically in a short time period. The search queries performed using the free to use APIs yield enough results to fit the requirements of the client and will not be needed to be upgraded to the paid options. If the client's web presence increases significantly in a short time, then the APIs will not yield the full results that are available on the internet.
- It is **assumed** that the client will use the system on only one machine. The client understands that the system will have to be configured on a single system, with dedicated space for the Database. If the client wants to use and configure the system on multiple devices, then the risk of duplication

and inconsistencies with the DB may occur.

- It is **assumed** that the client has the necessary software on the dedicated machine. Access to a PDF/JPEG viewer and a web browser will be available on the dedicated computer to allow the viewing of posts derived from the system. Unavailability of any such software will not allow the results (from the system) to be viewed properly.

b) External Interface Requirements

i. User Interfaces

The program will be implemented using python. It will include a searchable database that will have the capacity to view posts in an application/ browser. The client requires that the system be accessible on a dedicated computer/laptop so a user interface accessible to a single user will be offered to them for the purpose of this project.

ii. Hardware Interfaces

For the hardware: a screen, a mouse, a keyboard and an internet connection is needed. The system can also be deployed on a laptop, if such a need arises.

iii. Software Interfaces

The software and packages required for the project include a Python interpreter with all supported libraries/files properly installed on the Windows machine. The user will be able to access the system on the dedicated computer/laptop which will run the program on a regular basis (or possibly at the click of a button).

c) Non-Functional Requirements

i. Performance

- The system shall be accessible by one user at a time, and there will only be one dedicated computer for it to prevent having different database content for different users/computers.
- The first scrape from the web should not take a very long time to produce results, approximately no more than 10 minutes. Later scrapes will not include results that have already been catalogued, hence it should scrape much faster.
- Displaying and querying from the stored database shall not take longer than 30 seconds to return results.

ii. Security

- The program could have a login screen with username and password before allowing any access to the data.
- Safety checks in the program would display to the user when no new results are found from the latest web scrape and so nothing has changed in the existing database, preventing duplicates.
- The program and the database will not be saved on an open network or internet to keep the client's content private and keep up with QF's sharing rights.

iii. Software Quality Attributes

- Program/database should be available at all times to utilize and produce reports.

- Software should require as little maintenance as possible and if maintenance is required, it will be explained in the user manual.
- Software must be robust, should not crash in any scenario but show diagnosable errors for the user to fix using the user manual.
- Software should have a simple and user friendly design.

iv. User Documentation

- User Manual shall be created for the client, this will include how to: install the python program plus supporting libraries, and maintain it over the years, as well as a guide on how to use the software to produce desired results.

7. Feasibility Study

Economic: As per the client's request, the system will be deployed on a dedicated computer with an internet connection. The cost of that computer will be covered by QBG. It is economically feasible for QGB to cover the cost of the computer and an internet connection both in the short and long-term.

Operational: The size of the organization is relatively small, so there would not be any red-tape within the organization. QGB has a flat organizational structure. The system would be operated by Sally and Malik. Operationally, they would be responsible to maintain the computer, on which the application will be deployed and hosted.

Technical: QBG possesses all the necessary hardware and software features to implement the system. See the section on Technology Infrastructure & Management for more details.

Schedule: The system will be operational a week before the end of course (roughly June 7th, 2017). The organization will have sufficient time to test the system and achieve the anticipated benefits. The system will not radically change the running and working of the company but will assist QBG in collecting regular data that is on the internet.

Legal and contractual: This proposal is an informal contract between QBG and students. The system will be owned by QBG after it is formally deployed on their machine.

8. Risk Analysis

Risk	Description	Likelihood of occurrence	Magnitude of potential loss	Mitigation strategy
Retrieval of duplicate results (not unique records in database)	Each time a web scrape is conducted (harvest the internet for data) the results might include data that has already been recorded in the database	High	Low	Ensure checks are implemented within the code that remove duplicate and maintain the integrity of the database.
Not enough time to complete the full scope of the project	With the time constraint for the summer, there is a high risk of the requirements not being met within the given time frame.	High	High	Implement the strategy: "promise low and deliver high". This helps reflect the client's expectations according to the time allotted. Careful project planning and Gantt charts help to stay on track and deliver within the project's time limit
APIs get updated/become incompatible after project deployment	The privacy and security settings might affect how the APIs work and might cause some of the APIs to fail and cause the program to crash	Medium	Medium	The user manual will include steps to take in the case that an API fails and what the user should do to replace it with the latest API. Safety checks will be implemented to allow the program to keep running even when an API fails.

Transferring database to another computer causing data redundancy and disparity	Unless the program and the database are stored on a shared network, each computer that has the program installed will have a separate database causing data redundancy and disparity.	Medium	Low	Include instructions in the user manual to handle setting up of the program and the database on another computer.
Displaying database records collectively will not work with Tkinter alone	To view records from the database at a high level of functionality requires a well sorted and user friendly interface. Tkinter (python's native interface) may not be the best interface solution.	Medium	Low	Have a backup plan to use a web browser and ruby on rails or HTML only as a front end of the application/program with python only running the back end

9. Team Structure and Capabilities

Ali Abbas

Skills: Specialty in sockets/networking (python and Java), mobile application development with Visual Studio and C# (iOS and Android), some experience with Microsoft Access for databases.

Abdul Wasay

Skills: Worked with social media APIs and well versed in Python. Design skills in Adobe suite of software for user requirements. Good presentation and writing skills.

Amna Al-Ansari

Skills: Relevant background in programming languages such as Python, Java and Ruby on Rails. Capable and able to present, write and discuss documents efficiently.

Umair Qazi

Skills: Well-versed in Python, Data Structures and algorithms. Experienced in functional programming and systems programming.

10. Conclusion

The problems faced by the client is that of not being able to efficiently and effectively find news about the organization on the internet. This occurs because news about QBG (client) is sent out by a third party that does not communicate back where the news has been shared. Hence news and articles must be searched for manually. This takes time and effort and runs the risk of links expiring and the news disappearing forever. The proposed system that is solving this problem uses multiple APIs to search the web for QBG news/articles and catalogs them in a local database for the organization's personal use as well as their annual reports. Therefore, team Argo will implement a software that supports many APIs, downloads news about QBG and enables the organization to search efficiently and accurately.

References

Google API Library. (2017). Retrieved May 18, 2017, from https://console.developers.google.com/apis/library

Qur'anic Botanic Garden Vision and Mission. (2015). Retrieved May 18, 2017, from http://qbg.org.qa/

REST APIs — Twitter Developers. (2017). Retrieved May 18, 2017, from https://dev.twitter.com/rest/public

Appendices

Agenda	Client Meeting	Team Meeting	Date/Time	Done
Meeting to discuss possible questions for client.		x	May 9th, 10:30 to 11:30	✓
Introduction and discuss basic requirements.	x		May 10th, 8:30 to 10:30	✓
Discussion of first client meeting.		x	May 11th, 10:30 to 11:00	✓
Discussion of solutions.		x	May 13th, 12:00 to 2:00	✓
Suggestion and discussion of possible solutions.	x		May 14th, 1:25 to 2:10	✓
Division of the work amongst the team members for proposal.		x	May 15th, 9:00 to 9:40	✓