



University
of Glasgow | School of
Computing Science

Darwin: A Genetic Programming Framework for Cloning Web Services

George Kouzmov

School of Computing Science
Sir Alwyn Williams Building
University of Glasgow
G12 8QQ

Level 4 Project — 31 January 2014

Contents

1	Introduction	iv
1.1	Genetic Programming	iv
1.2	Web Services	v
1.3	Problem	vi
2	Context	vii
2.1	Background	vii
2.2	Organization	viii
2.2.1	Meetings and Minutes	viii
2.2.2	Requirements	viii
2.2.3	Agile Development	ix
2.2.4	User Stories and Sprints	ix
3	Design	xii
3.1	Multiple Technologies	xii
3.1.1	RESTful Web Services	xii
3.1.2	Genetic Programming Framework	xiii
3.1.3	Code Inspection and Generation	xiii
3.1.4	XML Files	xiii
3.1.5	Graphical User Interface	xiv
3.2	Architecture	xiv
3.2.1	Overall Design	xiv
3.2.2	Client	xv

3.2.3	Evaluator	xvi
3.3	GUI Design	xvi
3.3.1	Design Principles	xvi
3.3.2	Purpose of the GUI	xvi
3.3.3	Iterations and Final Design	xvi
4	Implementation	xvii
4.1	Technologies	xvii
4.1.1	DEAP	xvii
4.1.2	CherryPy	xvii
4.1.3	Inspect	xvii
4.1.4	Requests	xvii
4.1.5	wxPython	xvii
4.2	Implementation	xvii
4.2.1	DEAP Set Up	xvii
4.2.2	Generation of Individuals	xvii
4.2.3	Evaluation Over a Network	xvii
4.2.4	Code Generation	xvii
4.2.5	GUI	xvii
5	Evaluation	xviii
5.1	Unit Testing	xviii
5.2	System Testing	xix
5.3	UI and Usability Testing	xix
6	Problems Encountered and Future Development	xxi
6.1	Problems	xxi
6.2	Future Development	xxii
6.2.1	Improvements	xxii
6.2.2	New Functionality	xxii

7	Conclusion	xxiii
7.1	Contribution	xxiii
8	References	xxiv

Chapter 1

Introduction

In recent years web services have become a main method of communicating between applications and accessing needed information, creating similar to a distributive environment for many developers. Because of the standardized communication used by the web services they are expected to be used even more in the future providing even more commodities for both developers and users. The nature of web services is a black box system which means we are unfamiliar with its contents and we only know how to access it and what it returns.

That's where the idea came from, for copying a web services using genetic programming, which became a proof of concept project that reached very far. The genetic programming involves generating programs based on set of functions and then evaluating the quality (correctness) of the program via a fitness function. To prove that cloning of a web service can be achieved using genetic programming we used a symbolic regression problem where the genetic program needs to find a mathematical expression that yields the same result as a given one from a fitness function. But instead of using our own fitness function we are using a real web service hence generating a program that has the same functionality as the web service. Proving this meant that theoretically by increasing the complexity of our function set we can achieve copying of an even more complex web service.

1.1 Genetic Programming

Genetic programming is a branch in artificial intelligence inspired by biological evolution. It is a machine learning technique that optimizes a population of computer programs until they start performing the user specified functionality. Like the more widespread genetic algorithms (GA), genetic programming uses evolutionary algorithm based methodology to iterate through solutions until a correct one is found. In genetic programming a predefined size **population** is iterated through a user defined number of **generations**. The population consists of **individuals** which are programs represented as a tree structure. This tree structure is constructed from a predefined **primitive set** of functions and **terminals**. The terminals represent the bottom nodes of the tree, they can be both arguments and constants.[insert graph of an individual] A genetic program needs a **fitness function** it's a function through which each individual is evaluated and compared. This way the correctness of the program can be calculated. Usually there are two type of fitness function minimum and maximum where minimum is looking for individuals with fitness values closer to 0 (where 0 usually means that a solution is found) and maximum for the individuals with highest fitness result.

The way a genetic program works is it randomly generates individuals for the first generation evaluates them through the fitness function and after that two types of functions are applied - **crossover** and **mutation**. Crossover

is used to combine the genetic information of two individuals which means switching the nodes of a tree of one individual with another. With tree-based representation replacing a node means replacing the whole branch. When mutation is applied to an individual a single node is replaced with a new randomly generated one again changing a single node can mean changing the whole branch. After crossover and mutation are applied to the population a new generation that consists of the children of the old population, individuals from the old population and mutated individuals is created hence the iterations continue until the generations are completed. A result may not be found in the end of the generations.

One of the problems with genetic programming is searching the correct parameters to solve a specific problem. For each problem a specified number of populations, generations needs to be provided also the percentage of the population to which crossover and mutation are going to be applied. In more advanced cases the types of mutation, crossover and fitness function algorithms used need to be specified. Another complication is choosing the correct fitness function to evaluate the correctness of the individuals. If an incorrect fitness function is defined an individual with seemingly good fitness value may be far from correct. The primitive and terminal sets need to be chosen correctly as well. If too many primitives are chosen the generations might not be enough to find a solution to the problem because we are expanding the search boundaries. However if small sets are chosen we are limiting the search boundaries or the individuals won't have the big enough primitive set to satisfy the required functionality.

So using this methodology multiple problems can be solved. A common and a simple one is symbolic regression. In the case of symbolic regression an individual represents a mathematical expression and is looking to get the same results as another expression specified in the fitness function using the primitive set of operators available. In this project we are using this technique as a proof of concept for the capabilities of the framework.

1.2 Web Services

Web services are one of the main methods to provide utility computing. A definition for a web service is

A software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-processable format.

They represent an easy way for developers and clients to access and compute information. Usually through an API users are being able to communicate with a web service although sometimes they need to be accessed with raw URL or HTTP request. Unlike traditional client/server models, such as a Web server/Web page system, Web services do not provide the user with a GUI. Web services instead share business logic, data and processes through a programmatic interface across a network. The applications interface, not the users. Developers can then add the Web service to a GUI (such as a Web page or an executable program) to offer specific functionality to users.

Web services hold many benefits for developers. They represent components on the web that can be plugged in any project or software so it's functionality can be used. It improves communication between companies and users because information is distributed much faster. A good example of a web service is of Google who provide an API to their GPS maps which when given longitude and latitude return a map with the location of the user. This service has improved the user experience for many devices and because of it's good implementation has made the implementation of other similar services almost obsolete.

A problem with web services are that if they go offline or they have a bug, they affect the whole network and every user using them. Every non-open sourced framework is essentially a black box and the way it uses it's parameters is unknown. However these problems are minor to the overall benefit of the web services. Because of the way the internet works the amount of web services has essentially created a distributed system with multiple components. Developers can integrate complex systems easier than before and can easily plug in additional functionality as long as the right web service is there

1.3 Problem

The purpose of the project is to develop a framework that can clone a web service. Theoretically it is possible to complete this through genetic programming. Initially the project was separated into three steps - evaluate individuals over a network, clone an existing web service and improve a web service. The project aims to clone an external web services and provide a GUI to generate the configuration files needed for the framework to work.

Theoretically when using an external web service we know the parameters we are giving to it, we have an idea of the way it works and we know the results. Combining this knowledge meant it was possible to use an external web service as a fitness function to our genetic program and a way to specify the primitive set used. After part two of the project was completed it was proved that it is possible to clone the functionality of a web service. Because it is a proof of concept project a simple web service was used in order to verify the expected result. However the project proves that with the right choice of primitive set it is possible to clone the functionality of even more complex web services.

Chapter 2

Context

A key part of the project was organization and project management. With a complex project like this there were many factors and issues that needed to be considered before development could start. Research had to be done to identify other similar projects like the *Gen-o-fix: an embeddable framework for dynamic adaptive genetic improvement programming*. Technologies had to be considered since multiple approaches to the implementation were available. Frequent meetings and iterations over requirements were key so the goals can be identified. Software development methods like agile development had to be used to help structure the project, focus on the key goals and important features so sprints can be set up and development can start.

2.1 Background

Before finalizing the idea of the project, research had to be done so previous related work can be found and compared. The project has a very specific scope and there aren't many projects working towards a similar goal using genetic programming. The two of the projects were *Gen-o-fix: an embeddable framework for dynamic adaptive genetic improvement programming*[reference] and *A Genetic Programming Approach to Automated Software Repair*[reference].

Gen-o-fix explores the concept of genetic improvement programming (GIP) which goal is to automate software maintenance. *Gen-o-fix* is a GIP framework which allows a software system hosted on the Java Virtual Machine to be continually improved (e.g. make better predictions; pass more regression tests; reduce power consumption). Initially *Darwin* had the same goal but because of the complexity of the problem and the short amount of time it was decided to be left for future development. Similarly to *Gen-o-fix*, *Darwin* is also focused on creating a user-centric tool rather a research-centric one.

A Genetic Programming Approach to Automated Software Repair is a project which idea is to repair bugs in off-the-shelf legacy C programs. To achieve that the framework does three main modifications to the standard GP technique: (1) genetic operations are localized to the nodes along the execution path of the negative test case; (2) high-level statements are represented as single nodes in the program tree; (3) genetic operators use existing code in other parts of the program, so new code does not need to be invented. Initially *Darwin* was supposed to include similar functionality as well but it was again later excluded and kept for future development.

Both of the frameworks were focused on post-development processes like bug fixing, code improvement and code maintenance. These ideas are different from *Darwin*'s one of cloning web services however the concept for creating and modifying existing code through genetic programming is observed in all three of the project. This made the papers a good source of information and approaches used for the development of *Darwin*.

2.2 Organization

Frequent team meetings, proper software development method and good project management were key for the success of the project. Because of it's complexity there were many unknowns in the beginning and as an inexperienced developer into the field of genetic programming there were many basic that needed to be covered. Through this basic it was later possible to set the requirements, sprints and organize the project around them.

2.2.1 Meetings and Minutes

The core purpose of the meetings was to iterate through the development process with the client. Initially it was only explanation of the project, genetic programming, possible technologies and identify what can be developed for the time frame set for the project. Every week since week two there has been a project meeting. They were helpful for me and for the client so we can clear the ideas that we had, form them into one and work towards the same goal. The client's expertise in genetic programming and frequent meeting helped me understand how it works in little time so it was possible to begin development.

Since day one minutes were introduced to follow my development. This included reading papers, learning concepts, testing technologies. Minutes became the main way of following the project progress and a way for setting small goals for the week. Minutes also shaped the meetings by being able to discuss step by step the goals set from previous time. Every meeting had the same structure, it was discussed - what was done, what were the problems, new issues were addressed and possible solutions. The format of the minutes is shown in figure [insert minutes example]

2.2.2 Requirements

For requirements engineering an iterative approach was chosen. Through the first several meetings different ideas and goals were considered until requirements were finalized. The MoSCoW method was used to reach a common understanding with the client about the importance of each requirement and how relevant it was to the project. MoSCoW method involves grading each requirement for how important it is for the project using the following technique: (M - must, C - could, S - should, W - won't). This way main goals and issues can stand and development can focus on them. Set of requirements were chosen for each of the three sprints as seen in figure[insert table of requirements].

However later on because of the project complexity and the small time frame the third sprint was changed from *improving a web service* to creating graphical user interface for easier configuration of the framework. Even though requirements were set small changes were done through out the development process to suit the current situation and improve the software product[insert iterated version]

Before the requirements for each sprint were built there were basic requirements that needed to be identified. [show basic requirements] Framework for genetic programming and a REST framework had to be chosen. Weeks of research were dedicated to find the most suitable frameworks. The only initial requirement was that the project

should be done in python and later on after the client agreed with the technologies used they were added to the basic requirements. Knowing and understanding the technologies that would be the building blocks of the project helped guide the writing of the sprint requirements later on.

2.2.3 Agile Development

As a development method agile development was chosen for the project. But in order to maintain a good project management a tool for agile development was needed. Git hub[ref] was the version control used and Pivotal-Tracker[ref] was the agile tool of choice.

PivotalTracker is a web based tool where all the development process was stored, place to put all the user stories, functional and non-functional requirements, organize them and modify them easily. In the system it was easy to see when were the development bursts for the sprints and when was the post development time, when bugs were fixed, code was re-factored as seen in figure 2.1. This way the client can monitor how the actual development is going when was I working and on what feature. Pivotal tracker was used to monitor the progress of the project as seen in figure 2.2.

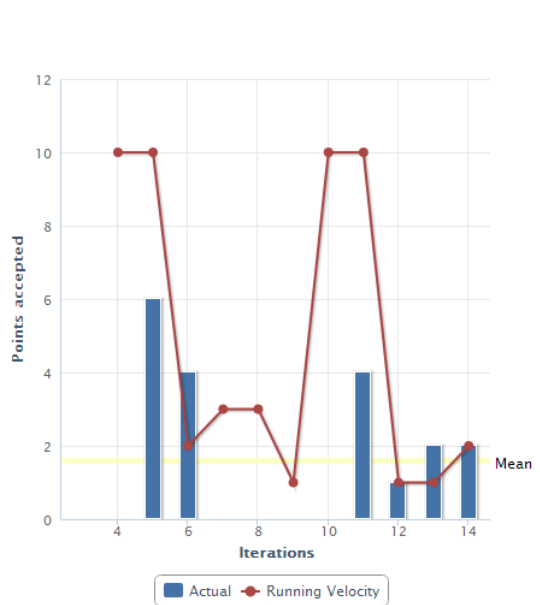


Figure 2.1: Velocity graph showing the sprints and their duration

Git hub is the version control tool used for the project. Being a public code repository gave the opportunity to present the code easily to the client and others as well. Frequent commits were done during the sprints. This way when the code broke during a sprint it could be reversed back to a working version.

2.2.4 User Stories and Sprints

As mentioned the project is focusing on a user-centric tool. Which meant user perspective on how the project should work and look was important. A list of user stories was created for the client explaining the way he and any potential user was going to interact with the framework. The list of stories was later iterated until there were set of user stories defining each of the sprints as seen in figure [ref]

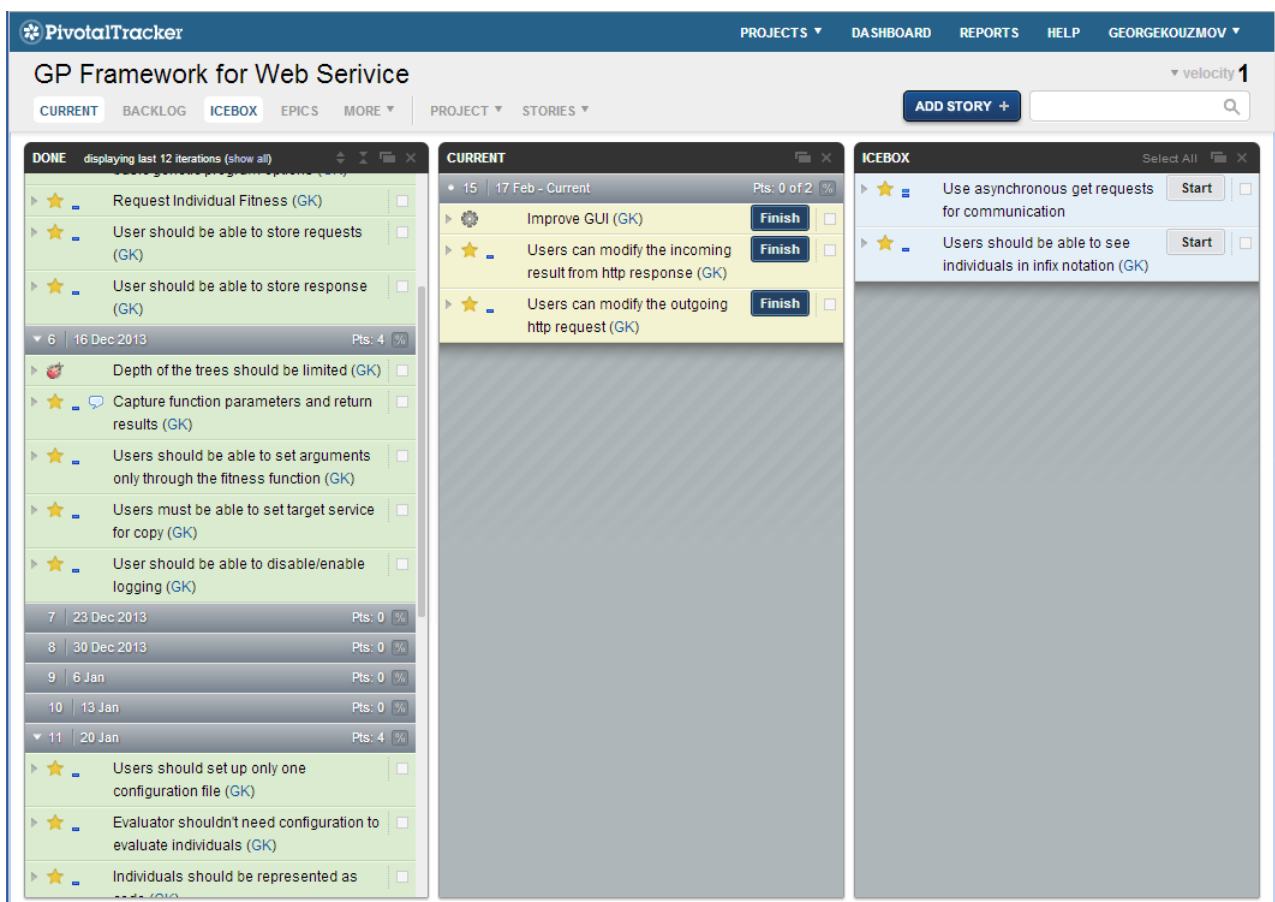


Figure 2.2: PivotalTracker and the way it is used by the project

Sprint one - was focussed on creating an evaluating web service separated from the client. That was done with the idea that with later development in the project this will be used for improving a web service using genetic programming. However this idea was left behind but the functional decision gave the potential ability to evaluate multiple individuals at the same time using multiple evaluating services so it was a good base for the project. First sprint included configuration of basic genetic programming parameters and using a pre-defined primitive set or a user defined one. However later on both were implemented. Generating python code out of the best solution was another feature for sprint one which was complete in order to create the base for a framework. Since DEAP was already the choice for the genetic programming framework integration with it was a requirement as well.

Sprint two - involved using a URL address to set up an existing web service as an "*Oracle*" which meant using them as a fitness function to the currently evolving web service. Because the user story was the user interacting with the system only through a single URL address that meant a lot of automated configuration needed to be done to create a simple easy to use framework. Unit tests had to be completed to make sure development was on the right track.

Sprint two - aimed for creating a more interactive and easy way to configure the system. Initially that meant only creating an XML file parsing all the configuration parameters in the framework. However a GUI for generating the XML file was added to ease the process. The purpose of the GUI was to ease the configuration without removing the previously existing methods hence keeping the advanced functionality of the framework.

Chapter 3

Design

When designing the framework the idea behind it was to provide a simple way to control the key parameters in a genetic program. Since the beginning of the project we were aware of the scope of the project and the complexity it can reach so another key design feature was modularity. Component based software engineering practices were used throughout the project, helping to organize the multiple technologies used in the project.

3.1 Multiple Technologies

Among the key issues in the development was combining all the technologies used in the project. A genetic programming framework, REST based web service, an XML parser, a GUI, code inspection and generation libraries; each was needed in order to create a framework that can clone web services. For each component that was needed there were multiple options available so the best ones needed to be chosen which meant trying out pack of frameworks and libraries.

3.1.1 RESTful Web Services

RESTful web services were needed for creating a service that can evaluate individuals over a network. Simplicity was important for the framework we are choosing. Three of the frameworks encountered were WebPy and CherryPy, Django all of them lightweight and commonly used. Other options for a REST framework were Flask and Bottle. However CherryPy and Django are the only frameworks that handle POST and GET request correctly, without any problems. In order to fix this issues with WebPy, Flask or Bottle additional libraries would of been needed. It was desired to avoid more dependencies because they would just make the project more complex. That meant the choice was between two. The more simple choice was CherryPy but that wasn't the only reason why the project was chosen. Another of David White's projects was using a cloud of RaspberryPis which meant this potentially can be used as a test for the Darwin framework. Because CherryPy provided support for RaspberryPy's it was the obvious choice.

3.1.2 Genetic Programming Framework

Genetic programming frameworks didn't have as many choices as RESTful web services however there was the option of implementing one. Pyevolve, pystep, pygene and DEAP were my only choices for genetic programming frameworks in python. None of them was minimalistic and simple enough, however implementing my own framework seemed as a much more complicated task that can slow down the project and it's final result. So a decision was made, that one of the frameworks was going to be used for the project which meant spending time learning more about the features they provide and how to use them correctly. The most common and supported one was the DEAP framework. It was a newer one but it had really good documentation, tutorials and examples making it the correct choice. However as a future development option for the project remains creating a genetic programming framework related to the project.

3.1.3 Code Inspection and Generation

Code inspection and generation was needed in order to achieve simplicity for the user. The idea was with fewer parameters to do a lot which meant a lot of code generation and code inspection. For inspecting code the best solution was *inspect* library which provided a way to extract source from function calls. It is a library that comes with python package has great documentation, examples and tutorials which proved only positive for the speed of development since not a lot of time was spend learning the inspect library.

Among one of the problems with using DEAP was it didn't support conversion between individuals and python source code or AST (Abstract Syntax Tree). Representing an individual as raw code or an AST meant that the evaluating web service can be simplified even more. This was an important part of the project however there are a lot of choices but the most correct one was the AST library which provides code generation while representing code as a syntax tree. This feature was very compatible with the way individuals are represented in genetic programming and it was more robust and safe way of doing code generation rather than using raw string. This started the idea of an open source contribution to the DEAP framework for generating python source and AST for a DEAP individuals.

3.1.4 XML Files

XML files are commonly used for configuring systems. Since the system is working with many parameters starting the code and defining each and every one took a lot of time and made the code look messy. XML file was a simple solution to a simple problem. There are many libraries in python used for parsing XML files - expat,minidom,BeautifulSoup,lxml ect. In the case of choosing an XML parser the chosen one was the most simple to understand and implement and lxml gave me that impression considering the good documentation it had.

3.1.5 Graphical User Interface

Graphical user interface was the next step into project simplification. It's role is to generate XML files for the user with the ease of controlling all the parameters needed. For creating GUI in python there are several good choices Tkinter,PyQT,wxPython. In my past I've had experience with Tkinter and I wasn't excited about working with it again. It looked much more complex compared to the other two. In sites such as StackOverflow the community was supporting wxPython. An additional benefit to using wxPython was the fact that there were a lot of good tools for generating layout. However my previous experience with GUI frameworks like Swing and it's similarities with wxPython, gave me the advantage of understanding wxPython quickly and writing a decent GUI in no time. In the end I didn't use any of the layout generating tools like wxGlade, wxDesigner or DialogBlocks because learning to use them was going to cost almost the same time as learning wxPython.

3.2 Architecture

A top down approach was applied for identifying the major components in the system. Client's requirements were the starting point for identifying those components. They were clear and simple enough to create a high level view of the framework , with which to explain the framework's basic behaviour and how it's going to work see figure 3.1.

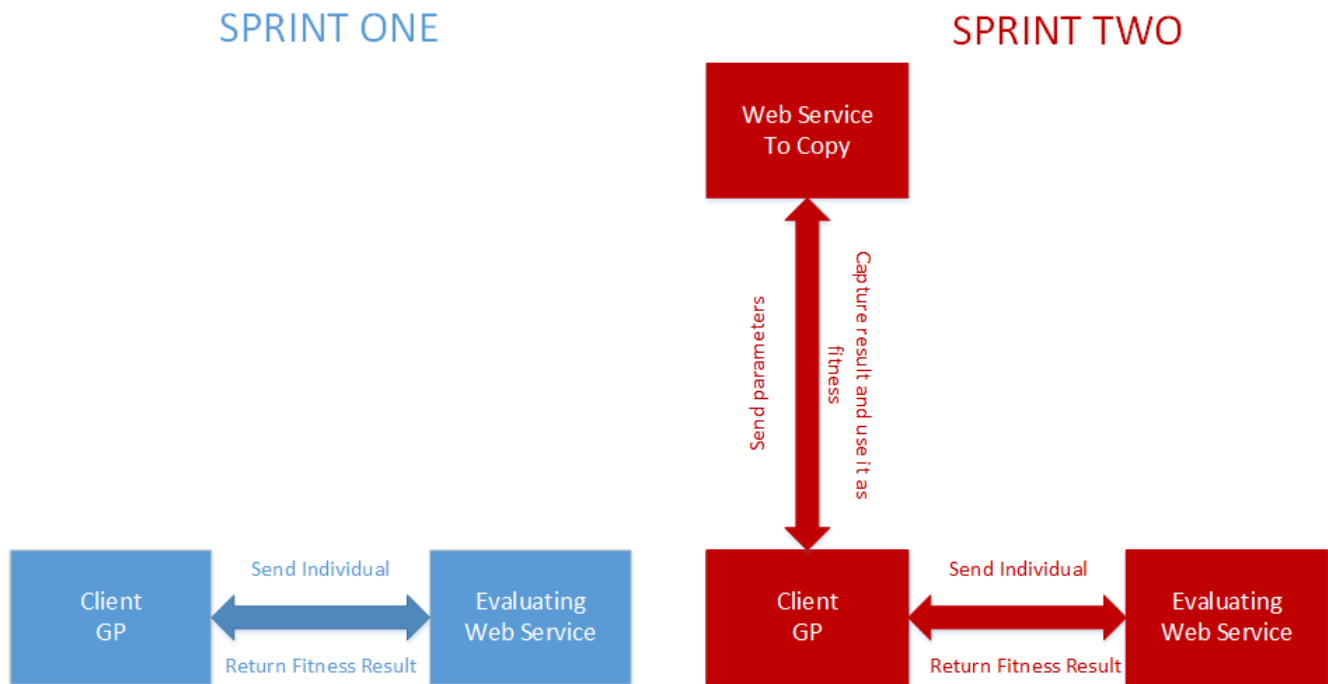


Figure 3.1: Top level view of sprint one and two

By displaying the way the high level components communicate it was later possible to identify lower level ones.

3.2.1 Overall Design

After the high level components were designed, the user stories helped create a more detailed view of them. At that point of time the project focused on sprint one which meant focusing on evaluating individuals over a

network. User stories helped identify how user was going to approach the framework hence it was able to create components with known input and return parameters. The result was a high level view of first sprint with all of it's components identified see figure 3.2

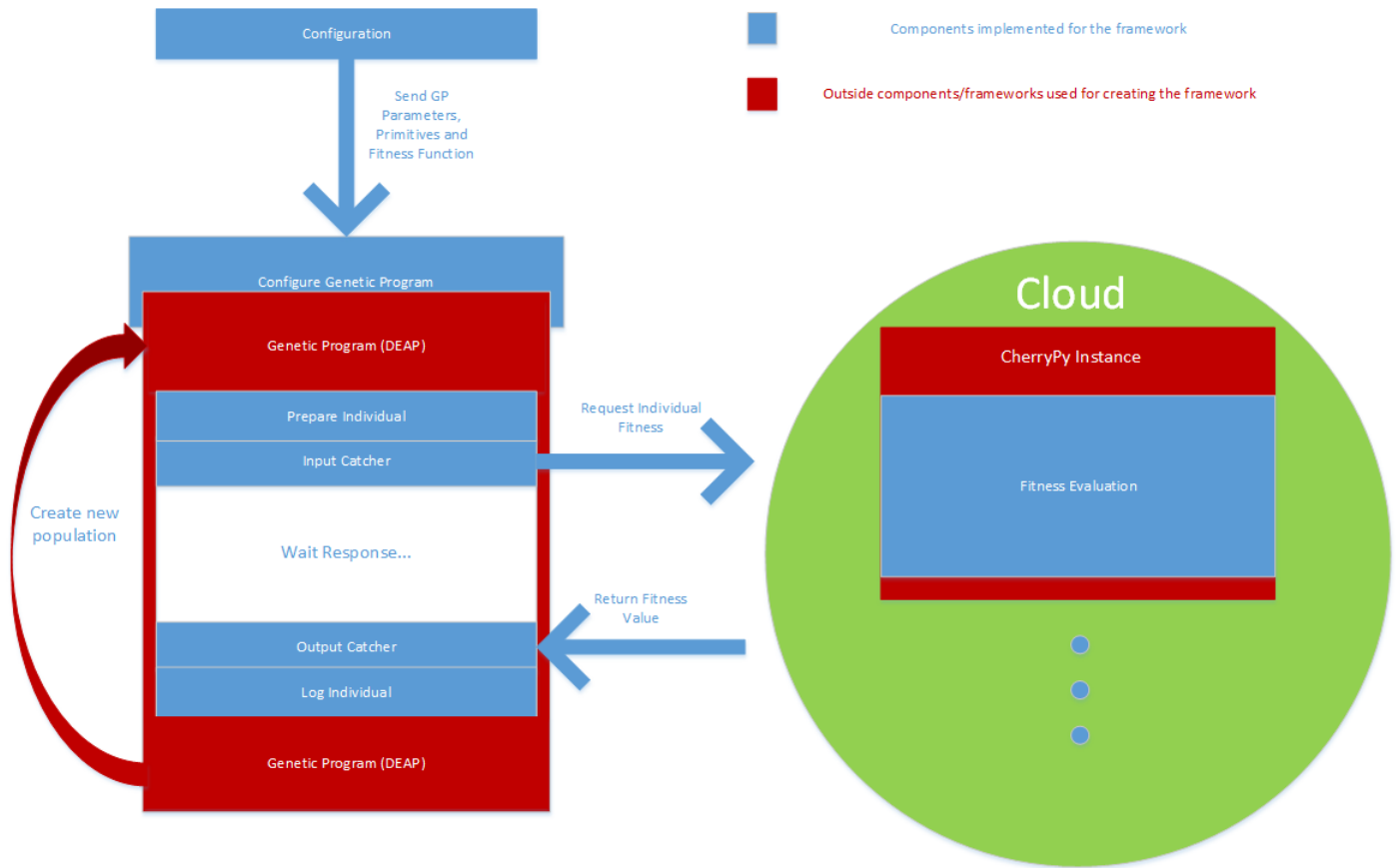


Figure 3.2: High level view of sprint one and the identified components used

Even though python2.7 unlike python3.3 is a semi object language for better use of software engineering practices and better ordering of the project, objects were used to represent the different components. The diverse capabilities of python and large set of libraries helped design a complex framework with very simple architecture. Later on more specifics for the client design will be explained. As you can see in figure 3.2 Configuration, Configure Genetic Program and Genetic Program represent the three main components of the client and CherryPy with an evaluation component represents the evaluating web service. This design helped us insert the technologies we are using at the correct places. For example Input Catcher and Output Catcher were later on substituted by a single component that using the requests library handles the requests and logs the traffic to the evaluator. These steps from top to down helped us clearly identified the classes that are going to be in the components.

3.2.2 Client

Before explaining how the client works, DEAP framework needs to be explained. It is configured through set of method calls to a *toolbox* this *toolbox* is prior configured by a *primitive set* where all the primitives and terminals that the genetic program is going to use are defined. The toolbox give easy API for accessing and controlling the individuals, generations and population as well as methods that invoke mutation, crossover and evaluation over the population.

Knowing how the DEAP framework works, it was later obvious that the same parameters that were needed for DEAP's configuration were going to be needed for Darwin's as well. Through the *Configuration* class shown in figure 3.2 we are able to set all the primitive functions and the fitness function needed for the program. Later on the client decided that there should be a preset number of primitive sets and the user will choose through the GUI which to enable. However in the case of an advance user if additional primitives need to be added the class can simply be extended and add them as normal methods. No other configurations regarding the primitives need to be done since Darwin is handling everything else. [insert class diagram]

In figure 3.2 you can see that for the configuration of the genetic program we need other parameters which are not specified in the *Configuration* class. When setting up the framework they can be set manually or configured by an XML file that carries all the information needed. There are only two parameters that are mandatory and the framework won't work without them - the address of the evaluating web service and the target service that needs to be cloned. Any other parameter is automatically set up with a standard value for solving the symbolic regression problem. The class that configures the framework acts like a data model since all the parameters needed through any other step of the framework is instantiated there. This data model is later used by the *Populator*.

The *Populator* class is basically a wrapper around the DEAP framework. What the class do is it creates the first population then sends each individual over the network to the evaluating web service. After each individual's fitness value is updated it repeats the process until the number of generations are reached. Besides executing the actual genetic program the (Populator) logs the traffic between the services and handles the output of the genetic program - either displaying the best individual or generating a web service working with the best individual.

3.2.3 Evaluator

The final big component is the Evaluator [reference to class diagram]. It is a CherryPy service that can be ran on any machine running python and has only one method - evaluate. It takes an individual, test arguments in the form of a dictionary and fitness results evaluates the individual compares with the fitness function and return the result. Initially the individual was in a DEAP represented form however that meant that in order to evaluate it the same DEAP configuration needed to be set up on the evaluating side as on the client one. This was impractical, slower, it was introducing many dependencies and much harder to set up. The Evaluator needs to be easy to set up, easy to modify because it is expected to be ran on multiple machines including RaspberryPis. Because of that the individuals was later represented as source code which meant it was able to run on any machine that runs python. This way simplicity was achieved for this component by moving most of the overhead work to the client.

3.3 GUI Design

-why use gui

3.3.1 Design Principles

3.3.2 Purpose of the GUI

3.3.3 Iterations and Final Design

Chapter 4

Implementation

4.1 Technologies

4.1.1 DEAP

4.1.2 CherryPy

4.1.3 Inspect

4.1.4 Requests

4.1.5 wxPython

4.2 Implementation

4.2.1 DEAP Set Up

4.2.2 Generation of Individuals

4.2.3 Evaluation Over a Network

4.2.4 Code Generation

4.2.5 GUI

Chapter 5

Evaluation

Evaluation is an important part of system development. *Darwin* being a complex framework extensive tests needed to be ran over each of it's component. Unit tests were used to verify the functionality of the core components, system testing was used to evaluate the overall functionality and stability of the system. User evaluation was used to test the usability of the framework and the graphical user interface.

5.1 Unit Testing

By separating the framework into several base components it was easier to split each of the components into it's core functions and apply unit tests to verify their correctness. However due to the complexity and dependencies in the framework test writing was complicated. Many of the functions depended either on the DEAP toolbox or on the primitive set which was responsible for generating DEAP individuals. They couldn't be reproduced manually for the tests, so multiple components had to be ran first in order to verify the unit tests. Since DEAP was an external framework with a stable release it was assumed that it is functioning correctly so no unit tests were ran on it specifically however components using it were extensively tested. Each of the components had testing aims however not all functions were covered since their complexity or dependency was too high, thats why they were left for system testing later on. The strategy for unit testing looked like this:

1. Unit tests on each of PrimitiveConfig's methods
 - (a) Test pre-set primitives
 - (b) Test code generation
 - (c) Test utility functions
2. Test all of Populator's methods without the configuration methods or the more complicated ones which would be part of the system testing.
 - (a) Test methods related to individual manipulation
 - (b) Test code generation
 - (c) Test fitness gathering functions
3. Handle all the functions in the helper module excluding Proxy class that is going to be aprt of the system testing.
 - (a) Test utility functions

(b) Test code generation

Through the unit tests many errors in the code generation functions were found so they had to be re factored. The imports used for the custom primitive set weren't working properly so they had to be fixed. After the unit tests were created initially they were used for verification of code correctness after a change in the code. If the unit test didn't work that meant there was an error in the code.

5.2 System Testing

There were many dependencies in the framework, which made testing a very difficult task. So instead of spending a lot of time trying to write complicated unit tests for each function/method in the code it was decided it's easier to do system testing. In the case of methods like *evaluate* where individuals were send to the evaluating web service unit tests weren't an option because in order to test them individuals had to be generated manually. Generating individuals was easy but generating the same ones every time so the test data can be robust was a problem. That's why after other framework functionalities were verified system testing was done to prove the correctness of the more complex components.

For system testing the goal was to use a generated web service for cloning and a real world one. The first step of the system testing was easy and simple enough. Testing proved successful and useful results were pulled out of it. For example a bug was detected in which DEAP individuals were generated with depth above 90. This meant that the python stack was breaking and couldn't evaluate the individual properly. The bug was fixed by modifying the way DEAP's decorator methods were used in the individual generation. This helped clear a major issue and second stage of system testing could being. In the second stage Wolfram Alpha[ref] was used as an external web service targeted for cloning. The first issues was the way Wolfram Alpha was receiving parameters and the responses in XML format. Since the framework didn't have functionality to handle the issue system testing was stopped and development over request and response handlers was started. After completion of the tasks system testing was initiated again. Framework passed through first and second stage of system testing. Second stage was tested with Wolfram Alpha. The framework succeeded to generate a web service that had the same functionality as the expression used as input for Wolfram Alpha. The mathematical formula used was measuring the face of a circle. It was a formula that was simple enough so cloning time can be short but complex enough to show the frameworks functionality.

5.3 UI and Usability Testing

Important feature for a framework is usability. Participants were used to evaluate the quality of the framework and check if the goals to create an easy to use and simple framework were achieved. The participants were presented with tutorial for the framework and it's documentation each of them had to have basic knowledge in programming. The choice of first to fourth year students presented a good study group since it presented people from different programming experience. The participants had to asses the quality of the tutorial and documentation and answer the following questions:

1. Was it helpful to understand the basics of the framework?
2. How complex it was to create the first examples?
3. Were you given enough information to continue working with the framework on your own?

4. Do you feel confident that you understand the framework basics?

In another part of the evaluation they had to assess the graphical user interface. They had to answer the following questions:

1. Was it intuitive enough?
2. Was it simple?
3. Was it complex?
4. Did it present enough options?
5. Did it present too many options?
6. Would you think tips would be helpful above each field?
7. Can you give additional feedback that can help improve the GUI

In addition through the evaluation the participants were observed and notes on their actions were taken to compare with their feedback. This way more accurate conclusions could be drawn so the correct features could be improved and re-worked.

The studies helped spot problems in...[wait for results]

Chapter 6

Problems Encountered and Future Development

It has been a great journey developing this project and one that will hopefully continue further. Since the beginning there have been many difficulties with the project mostly related with the meta-programming field it's focused on. Finding sources and doing research has been problematic because development in the field of genetic programming isn't on the same scale as other more popular fields in programming (e.g Machine Learning, Artificial Intelligence). However the result from the project has motivated me to continue explore genetic programming even more and expand the framework.

6.1 Problems

Multiple Technologies - among the main issues with the project is the amount of technologies used. This meant each of them had to be understood so quality development in the area can be done. The framework was related to web services, genetic programming, code generation, GUI which meant extensive research in each area so the correct libraries or frameworks can be chosen. The learning curve was steep for the project however each of the fields became easier to work and understand after the basics were clear.

Proof of Concept - before the project reached any results it had a *proof of concept* label. Exploring new fields and expanding your knowledge in specific fields can be fun but it's hard when not a lot of information is presented. Three times during the project I had to reach out to *stackoverflow* for help related to issues I couldn't find anywhere on the web resolved. The first time it was related to the way CherryPy processed its requests and responses. The second one was focusing on the problem of converting DEAP individuals into executable code or AST. Later on I had to tackle the problem on my own since such a functionality wasn't introduced to DEAP. However the developers from DEAP were kind enough to provide support for further questions and the code generation has spawned the possibility for an open source contribution to the DEAP framework with an extension that converts DEAP individuals to AST's. The third time was related to code execution in python and the way arguments were passed to a sub process.

6.2 Future Development

Since the beginning of the project the development has been very ambitious and all the goals that were set were accomplished. Which gave the opportunity to think a lot about improvements for the framework - anything from re-factoring, improving features to adding additional functionality.

6.2.1 Improvements

Expand symbolic regression - throughout the project symbolic regression was used as a proof of concept so the framework functionality can be showed. However non mathematical web services are harder to implement using the framework and this can be improved so cloning can be more automated. This can be done by adding a better method for overriding at the evaluating web service and tweaking the basic primitives sets by adding primitives for control flows.

AST for code generation - currently all the code that is generated is in string form and through unit tests it was proved that it's working well however using string manipulation rather than manipulating ASTs is not as robust. A good re-factoring plan is to do all the code generation with ASTs and keep generation from strings as little as possible. This would also help for creating the base for an open source contribution to the DEAP framework that adds the functionality for individuals to be converted to ASTs and/or executable python code.

Asynchronous requests - among the initial requirements of the project was using a cloud of evaluating web services to evaluate individuals over the web hence improving the speed especially for more complex individuals. However that idea is a small project on it's own and it was going to bring a lot of complexity to the project. Since the core of *Darwin* is completed it's a solid place to start from however for dealing with threads and asynchronous requests python is not the best language to handle the issues, implementing it in C would be much better

6.2.2 New Functionality

Improvement of web services - improving web service performance using the framework was an initial requirement in the project but due to complexity and time issues it was opt-out from the project. The concept can still be implemented and used for improving current web services that the developer has created so improved web service can be compared by performance speed, memory usage and stability to the target one. After certain threshold of improvement is reached there can be automatic deployment of the generated improved web service.

Using unit tests as fitness - a way of measuring fitness can be a set of unit tests the individual needs to pass. This way individual can be tested if certain functionality exists and hence force the genetic program to evolve towards an individual with the required functionality. From that point a dynamic addition of unit tests can be included so functionality can be introduced through the evolution of a program. Comparing it to biological evolution its like adding a new threat for an organism in its environment hence the organism needs to adapt.

Chapter 7

Conclusion

7.1 Contribution

Chapter 8

References