



# **INFORMATICS INSTITUTE OF TECHNOLOGY**

**In Collaboration with**

## **UNIVERSITY OF WESTMINSTER (UOW)**

**BEng/BEng.(Hons) in Software Engineering**

**Final year Project 2014/2015**

### **Terms of Reference**

**For**

**Project Title: Identify Inherited Diseases based on DNA (IIDDNA)**

**By**

**Iddamalgodage Don Lahiru Manohara - 2010070**

**Supervised By: Mr. Achala Chathuranga Aponso**

.....

Signature of Supervisor

.....

Signature of Student

## Table of Contents

1. Project Background .....	1
2. Aims.....	2
3. Scope.....	2
4. Objectives.....	3
5. Features of Prototype .....	4
6. Project Deliverables .....	5
7. Resource Requirement .....	5
8. References .....	6
9. Activity Schedules .....	7

## 1. Project Background

“What are the chances that we will one day discover that DNA has absolutely nothing to do with inheritance? They are effectively zero.” (Sam Harris, 2014). DNA is established at connection, and does not change throughout human’s life time. Human receive one-half of DNA form their mother and on-half from their father. Genetic mutation occur when DNA changes, altering genetic instructions. This may result in inherited diseases. Analyzing DNA, and identify inherited diseases will show future traits of human life. Parents can be decided their unborn children’s has future risk of having inherited disease.

Extract of meaningful information from large experimental data set is a key element of bioinformatics. DNA pattern analysis and extract meaningful information is a technique to identify inherited diseases. Different algorithms are in data mining techniques involve finding diseases, and visualize sample data set. There are many researches are going into DNA analysis and genetic disease. GWAS (genome-wide association study) is a one of major institute support scientist to identify gene involved in human disease. “Identifying patterns of copy number variants in case control studies of human genetic disorders” research conduct to bring missing DNA data in cell, apart from them it can identify genetic disorders (Abdullah K. Alqallaf, 2009). The different of DAN sequence are called single nucleotide polymorphisms (SNPs). Researchers have found SNPs that may help predict an individual’s response to certain drugs, susceptibility to environmental factors such as toxins, and risk of developing particular diseases. Research like “A SNP and KEGG Based Approach to Mine Risk Pathways Associated with Bipolar Disorder” proved the SNPs are help to identify disorders (Liangcai Zhang, 2008).

Even though outcomes of research were very successful, DNA analysis and find inherited disease based researches are rare. Most of the researchers are explain the approaches to determine genetic disease. But there is no clear path explain of inherit DNA analysis and identify diseases. “The inference of genes that are associated with human inherited diseases (diseases gene) has been a task of grate challenging in biological and medical studies” (HE Peng, 2012, p. 1). Making these research approaches are apply to identify inherited disease based on DNA analysis.

The purpose of “IIDDAN” project is identify inhered disease based on DNA analysis data set which applying different data mining techniques and bioinformatics algorithms to predict probability of having inherited diseases.

## 2. Aims

The aim of the project is to research, design, and develop a prototype outcome of the research, identify inherited diseases based on DNA analysis data. Also demonstrate the algorithms and methods which found from the research.

## 3. Scope

First part of the project is search, and finds approaches which researchers were taken in their researches. Based on these approaches try to find out solution to identify inherited disease using visualizes, and running different algorithms on data set in different statistical analysis tools.

When successful approach found from the testing and running different algorithms on data sets, develop the prototype application with limited futures. User input sample test data with given format, the application runs the algorithm and display probability of having particular disease.

Due to the time constrains the prototype support very basic functionalities but yet sufficient to demonstrate outcome of the research. The feature like visualize data set is not supported in prototype.

## 4. Objectives

1. Writing Terms of Reference report to identify the problem domain of the project and define scope, objectives and deliverables.

❖ **Output Artifact: Terms of Reference Document**

2. Conducting a literature survey on the following topics to gain depth knowledge and analyze past works done in those areas and estimate the future work needed to be done.

- Different approaches for analysis DNA patterns.
- Existing application and prototype for DNA analysis, and identify diseases.
- Data mining techniques related to DNA analysis and classify diseases.
- Predictions algorithms for probability of having classified disease.

❖ **Output Artifact: Literature Review Document**

3. Identify and analyze requirements by conduction online surveys and interviews for a prototype evaluation. This phase helps to improve prototype features.

❖ **Output Artifact: Software Requirement Specification (SRS)**

4. Design the application incorporating the requirements prioritized through analysis and using appropriate design methodology to guide the implementation.
5. Select appropriate development technologies and tools which are very ideal for the development of the application. Identifying the appropriate technology is significant to develop an efficient prototype.

❖ **Output Artifact: Design Specification**

6. Develop a prototype for demonstrating proposed features.  
Developing the following components is necessary to fulfill the goals of the project.

- Development prototype to demonstrate a solution found from the research.

❖ **Output Artifact: Application to demonstrate a solution found from the research.**

7. Test the prototype by unit testing and also by a black box testing approach to ensure all the requirements gathered in the requirement gathering phase are incorporated and functional.
8. Evaluate the prototype by allowing potential users to use the prototype and obtain feedback to ensure the outcome is good to be usable and user-friendly.
  - 8.1. Enhance prototype according to user feedbacks.
9. Submit all the project deliverables with in the dead line.

## 5. Features of Prototype

- ❖ Plugin data source or data set  
The application is able to plug in different data sources with pre-defined data format to analysis, and produce the appropriate result.
- ❖ Implemented algorithms found from the research  
The best approach which found from the research, implement the application with appropriate libraries and custom logic to develop the business logic of the application.
- ❖ User Interface for display result in understanding manner  
Eliminate complexity in the process of analyzing data, change threshold in algorithm and easily understand the test result etc. Design simple user interface to understand the process of the analysis and plugin data sources. User will be able to understand process of the application without much hassle.

## 6. Project Deliverables

- ❖ Terms of Reference
- ❖ Literature Review Document
- ❖ Requirement Specification Document
- ❖ Design Specification
- ❖ Prototype
- ❖ Prototype Report
- ❖ Test Report
- ❖ Interim Report
- ❖ Final Report

## 7. Resource Requirement

### Software Requirements

- ❖ MATLAB
- ❖ Eclipse IDE
- ❖ R software
- ❖ Java

### Hardware Requirement

- ❖ Dual Core 2.0 GHz CPU
- ❖ 2GB DD2 Ram
- ❖ 5GB Hard Disk Free Space
- ❖ Intel HD Graphics Card

Note – Above requirements are subject to change

## 8. References

1. Sam Harris, *The End of Faith: Religion, Terror, and the Future of Reason* [online]  
Available from: <https://www.goodreads.com/quotes/261279-what-are-the-chances-that-we-will-one-day-discover>  
[Accessed 9<sup>th</sup> September 2014]
2. *How genetic conditions are inherited*, 2014. Available from:  
<http://www.nhs.uk/Conditions/Genetics/Pages/Facts.aspx>  
[Accessed 8<sup>th</sup> July 2014]
3. *What are single nucleotide polymorphisms (SNPs)?*. Available from:  
<http://ghr.nlm.nih.gov/handbook/genomicresearch/snp>  
[Accessed 8<sup>th</sup> September 2014]
4. Alqallaf A.K., Tewfik A.H., Krakowiak P., Tassone F., Davis R., Hansen R. , Hertz-Picciotto I., Pessah I., Gregg J., Selleck S.B., '*Identifying patterns of copy number variants in case-control studies of human genetic disorders*', pp. 1-4, Available from: IEEE Xplore digital library. [21 May 2009].
5. Liangcai Zhang, Lina Chen, Yan Zhao, Liangde Xu, '*A SNP and KEGG Based Approach to Mine Risk Pathways Associated with Bipolar Disorder*', pp. 34 – 38, Available from: IEEE Xplore digital library. [20 Oct. 2008].
6. He Peng , Jiang Rui, '*Integrating multiple gene semantic similarity profiles to infer disease genes*', pp. 7420 – 4725, Available from: IEEE Xplore digital library. [27 July 2012].



## 9. Activity Schedules

ID	Task Name	Duration	Start	Finish
1	<b>Term Of Reference</b>	<b>42 days</b>	<b>Fri 8/8/14</b>	<b>Mon 10/6/14</b>
2	Identify Project Scope	19 days	Fri 8/8/14	Wed 9/3/14
3	Identify Aims	11 days	Mon 9/1/14	Mon 9/15/14
4	Identify Objectives	11 days	Mon 9/1/14	Mon 9/15/14
5	Submission draft Terms Of Reference	0 days	Tue 9/16/14	Tue 9/16/14
6	Evaluation Of draft Terms of Reference	10 days	Tue 9/16/14	Mon 9/29/14
7	Submission of Final Terms of Reference	0 days	Mon 10/6/14	Mon 10/6/14
8	<b>Literature Review</b>	<b>104 days</b>	<b>Fri 8/8/14</b>	<b>Wed 12/31/14</b>
9	Conducting a literature search	104 days	Fri 8/8/14	Wed 12/31/14
10	Critical analysis the project with existing projects	22 days	Mon 9/1/14	Tue 9/30/14
11	Write Literature Review report	13 days	Wed 10/1/14	Fri 10/17/14
12	Submission of Literature Review	0 days	Mon 10/20/14	Mon 10/20/14
13	<b>Requirement Analysis</b>	<b>29 days</b>	<b>Wed 10/15/14</b>	<b>Sat 11/22/14</b>
14	Conduct online surveys	13 days	Wed 10/15/14	Fri 10/31/14
15	Identify main requirements	12 days	Sat 11/1/14	Sat 11/15/14
16	Writing requirement specification document	6 days	Sat 11/15/14	Fri 11/21/14
17	Submission of requirement specification	0 days	Wed 10/22/14	Wed 10/22/14
18	<b>Designing the prototype</b>	<b>71 days</b>	<b>Tue 10/28/14</b>	<b>Tue 2/3/15</b>
19	Design the system architecture	55 days	Tue 10/28/14	Sat 1/10/15
20	Submission on interim report	0 days	Mon 2/2/15	Mon 2/2/15
21	<b>Implementing the prototype</b>	<b>66 days</b>	<b>Tue 11/25/14</b>	<b>Tue 2/24/15</b>
22	Find suitable technology to implementation	11 days	Tue 11/25/14	Tue 12/9/14
23	Setting up development environment	3 days	Tue 12/9/14	Thu 12/11/14
24	Development	55 days	Tue 12/9/14	Sat 2/21/15
25	Prototype report and demonstration	0 days	Mon 2/23/15	Mon 2/23/15
26	<b>Testing</b>	<b>24 days</b>	<b>Mon 12/15/14</b>	<b>Thu 1/15/15</b>
27	White Box Testing(Unit Testing)	9 days	Mon 12/15/14	Thu 12/25/14
28	Black Box Testing	16 days	Thu 12/25/14	Thu 1/15/15
29	<b>Evaluation</b>	<b>28 days</b>	<b>Thu 1/15/15</b>	<b>Sat 2/21/15</b>
30	Showcasing prototype for potential users	3 days	Thu 1/15/15	Sun 1/18/15
31	Analysing the feedback	3 days	Thu 1/15/15	Sun 1/18/15
32	Enhance Prototype based in the feedback	27 days	Thu 1/15/15	Fri 2/20/15
33	<b>Final Report</b>	<b>187 days</b>	<b>Fri 8/8/14</b>	<b>Mon 4/27/15</b>
34	Preparing final report	171 days	Fri 8/8/14	Fri 4/3/15
35	Submission of draft project report	0 days	Mon 4/6/15	Mon 4/6/15
36	Submission of bound copies of final project report	1 day	Mon 4/27/15	Mon 4/27/15