**COM570: Project**

**Dissertation**

**School of Computing & Information Engineering**

**Jason McKee (B00553970)**

**Data Cleansing System**

**Supervisor: Gerry Parr**

**Second Marker: David McSherry**

**Date: 2nd May 2014**

****

Plagiarism Statement

I declare that this is my own work and that any material I have referred to has been accurately and consistently referenced. I have read the University’s policy on plagiarism and understand the definition of plagiarism as given in the Project Handbook. If it is shown that material has been plagiarised, or I have otherwise attempted to obtain an unfair advantage for myself or others, I understand that I may face sanctions in accordance with the policies and procedures of the University. A mark of zero may be awarded and the reason for that mark will be recorded on my file.

Signed: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Acknowledgements

Thanks to anyone who contributed directly or indirectly to the project, in any way.

Contents

[Abstract 1](#_Toc363725067)

[Introduction 1](#_Toc363725068)

[Analysis 1](#_Toc363725069)

[Design 2](#_Toc363725070)

[Implementation, Testing and Evaluation 3](#_Toc363725071)

[Conclusions 4](#_Toc363725072)

[References 4](#_Toc363725073)

[Appendices 4](#_Toc363725074)

[A1. Analysis Models 4](#_Toc363725075)

[A2. Design Models 4](#_Toc363725076)

[A3. Code 4](#_Toc363725077)

[A4. Test Suite 4](#_Toc363725078)

[A5. Questionnaire Results 5](#_Toc363725079)

# Abstract

(iii) a summary of the work completed.

The problem is that the Northern Ireland Statistics and Research Agency (NISRA) receive several datasets throughout the year which contain thousands of addresses on which critical statistical data outputs rely on such as Census Data. It has been noticed that a large percentage of those address cannot be validated as correct address. This is due to the addresses being captured through forms completed by people which lead to human error in the addresses. If an address cannot be confirmed it is not then reliable enough to be used in key statistical data. The current process of confirming these addresses is expensive in both time and money.

It was then proposed to create a less expensive application that would carry out the current process more quickly. The application would make use of the current software Microsoft Access and the programming language Visual Basic with Applications (VBA). This is because current applications that NISRA use are built using these utilities and therefore they are recognised by the staff and also reduces training to a minimum. Another request was the ability to carry out the process through Microsoft SQL Server as this is where the data for processing is held for cleansing. This was acceptable as MS Access and MS SQL Server can communicate quite effectively. The finished solution is to carry out the majority of the processes within MS SQL Server as Access allows this via pass through queries which allows for quicker process times and also incorporates the current security procedures within NISRA. The purpose of MS Access is to provide a user friendly front end User Interface that the staff can use effectively.

To date an application was created within MS Access to act as a front end user interface for the address validation process which is carried out in the back end within MS SQL Server. Many improvements have been made over the development of the project as will be outlined in this report but the original idea has been followed with the addition of some useful features. The application works as required and this report will outline some possible future versions of this project.

This current version of the project is being used successfully by staff at NISRA.

# Introduction

The Introduction is effectively an expanded version of the abstract. It again summarises the problem, solution and work undertaken, but this time in more detail. The Introduction can build directly on the Analysis Report that you produced in the first semester. The chapter should conclude with a summary of the remainder of the dissertation.

Length: ~4 pages

# Analysis

The Analysis Chapter documents your work in finding a suitable solution to the problem being tackled. The nature of this investigation will vary from project to project, but may involve the use of questionnaires, interviewing stakeholders, researching relevant background material and solution options, including the creation of experimental prototypes, and so on.

It is valuable to include a discussion of any ideas that were investigated but subsequently dropped, to illustrate the problem-solving process involved.

The investigation of the problem and its potential solution will have led to the identification of a set of system requirements, which should be presented in the Analysis Chapter. The requirements should be documented using whatever descriptive techniques are appropriate for the type of system and type of requirements involved. For example, functional requirements might be documented as user stories or use cases.

Material that you previously prepared for the Analysis and Design reports in the first semester should be a good starting point for both this chapter and the next one.

Length: ~12 pages

# Design

Design covers the user interface, software architecture, data definitions, algorithms and other high-level descriptions of the system you have created. Ideally, a good system design document is one that can be passed to someone else to implement.

It is expected that during the design phase various options will have been considered before any final decision was taken. These options should be identified and the rationale for each decision presented.

You are encouraged to use descriptions and models suitable for your own circumstances. For example,

* In describing how the user will interact with your system you may want to present a block diagram identifying key parts of the user interface in addition to showing screen shots. If you considered HCI guidelines, you should explain how these influenced your design
* If you are using web pages, you may want to include a map of your site, indicating how users navigate through it
* If you are using a database, you will need to describe the design schema, including details of any normalisation involved. You may also want to include an entity-relationship diagram
* For a Java program, you will need to start with the class design, identifying your main classes and indicating their purpose
* For some processing it may be necessary to make use of complex algorithms, which should be described and illustrated appropriately

Length: ~14 pages

# Implementation, Testing and Evaluation

The content of the implementation, testing and evaluation chapter is largely self-explanatory.

First, under implementation, you should describe the approach you took to development, indicating what was learned from the prototype and successive versions produced, identifying any other incremental steps followed to achieve the final system.

Secondly, you should describe any specific tools used to support the development process, including details of any new languages that you needed to learn.

Also under implementation, you should include details of any significant aspects of the code you have produced, including the use of specific algorithms. If you have used code from other sources, this should also be identified, indicating how it was integrated into your system.

Your approach to testing should then be described, identifying the test cases that you have used to verify the correctness of the software. These should demonstrate that your testing has been appropriate and thorough.

Under evaluation, you should assess the perceived value of your system to its intended users against the specified requirements. You are not expected to create a ‘perfect’ system so marks are awarded more for the thoroughness of your validation than any praise obtained. Indeed criticism gives you an opportunity to explain the issue raised and suggest improvement.

Note that when planning your project, sufficient time should be included after implementation for testing and evaluation as these activities have a significant impact on the final quality of your system and the write up in your dissertation.

Length: ~16 pages

# Conclusions

In the final chapter you should summarise your project work overall and assess it critically. This should indicate what lessons you have learned and so clarify what you might do differently if faced with the same situation again. In particular, you should identify and discuss how the project plan evolved as the project progressed.

The limited time available for implementation means that you are likely to have ideas for further work. These should also be included in the conclusions.

The Conclusions chapter, like the Introduction, should be freestanding, allowing the reader to understand what the project has achieved without studying other chapters.

Length: ~4 pages

# References

Present the references used in the dissertation in Harvard format.

# Appendices

The appendices are an opportunity to provide secondary material in support of the description in the body of the dissertation. In principle, the reader need not look at the appendices and no specific marks are awarded for this section.

Sample content:

## A1. Analysis Models

e.g. SSM models

## A2. Design Models

e.g. database schema

## A3. Code

Code developed through the project

## A4. Test Suite

Full set of tests applied to the software

## A5. Questionnaire Results

Results of questionnaires used to evaluate the software or identify requirements