

**ANGLIA POLYTECHNIC UNIVERSITY**

**ABSTRACT**

DEPARTMENT OF COMPUTING  
FACULTY OF SCIENCE AND TECHNOLOGY

MASTER OF SCIENCE

CREATING A CUSTOMISABLE NAME MATCHING ALGORITHM FOR  
HISTORICAL COMPUTING BY REFACTORING

By Gillian Hazel Newton  
May 2005

This project investigates how a customisable tool for matching names that sound alike but are spelled differently can be produced by refactoring and extending an existing phonetic matching algorithm. The project aims to produce an extensible, portable and easy to use name matching system for use in historical research computing. The motivation behind such research lies in the need for a tool that is more reliable than the Soundex algorithm for the computerised linkage of records by personal names.

The main body of the project is split into four major stages. At each stage the methods are described and the outcomes evaluated. The first stage concerns the development of a user interface for the system. The second, third and fourth stages concern sequential steps in refactoring and extending the double metaphone algorithm to fit it to the purposes of the project.

Refactoring design patterns applicable to the project are critically evaluated, and it is shown that a flexible refactoring process strictly controlled by regression testing is an effective means of reusing and extending expert knowledge bound up in existing code. It is further shown that specific low-level refactorings can be applied to introduce object orientated structure to improve the logical organisation of code, in addition to the more commonly understood application of refactoring to code that is already object orientated. The reliability of the new customised algorithm is evaluated, and indicators of its superiority to Soundex are given.

Full source code for the system is provided in the appendices, together with a worked example of how to use it to match names stored in a Microsoft Access database (Appendix G).