University of London

Computing and Information Systems/Creative Computing

CO2226 Software engineering, algorithm design and analysis

Coursework assignment 1 2018–19

Data Mining Tools and Techniques

***Note**: You may need to undertake some background reading on Data Mining (DM) as part of this coursework assignment. This is factored into the time you are expected to spend on thinking, developing, writing and checking your answers to the questions. It is recommended that you devote between one fifth and one third of the time you devote to this coursework assignment on targeted background reading, primarily focused on the answer to Task 3.*

Submission details

Please submit **one** PDF document which is named using the following convention:

YourName_SRN_COxxxxcw#.pdf (*e.g*. MarkZuckerberg_920000000_CO2226cw1.pdf)

- **YourName** is your full name as it appears on your student record (check your student portal)
- **SRN** is your Student Reference Number, for example 920000000
- **COXXXX** is the course number, for example CO2226, and
- **cw#** is either cw1 (coursework 1) or cw2 (coursework 2).

It is important that your submitted assignment is your own individual work and, for the most part, written in your own words. You must provide appropriate in-text citation for both paraphrase and quotation, with a detailed reference section at the end of your assignment. It is important that your submitted assignment is your own individual work and, for the most part, written in your own words. Copying, plagiarism and unaccredited and/or wholesale reproduction of material from books online sources, *etc.* is unacceptable, and will be penalised (see How to avoid plagiarism).

The scenario: BooksWeLove is a new project, funded by a public-private partnership, aiming to provide an easy, affordable, secure and quick way to finding a new book to read (or comment on/recommend one just read) for everyone with an internet connection. The aim is to develop a web application that anyone can run from their browser, as well as apps running on all mobile platforms drawing data from the same database (securely hosted and run by a separate company). The app will provide information on current and forthcoming publications, as well as allowing customers to provide feedback on books they have read, access personalised offers, register their book genre interests and even pay for any purchases made either through a loyalty account or through existing payment gateways. The app should also serve as an "online get-together" space where book presentations or, simply, book club meetings can be held between its members.

The BooksWeLove project will be implemented by Cloud Solutions Limited, with Sanjeev

Patel as the senior executive manager. Patel has been working in IT for about fifteen years now, switching from an accounting background. He is the senior partner and head of Business IT Services of Cloud Solutions Limited, employing around a hundred people (on different work patterns – full-time, part-time, on- demand, flexible contracts), each with different skills. Cloud Solutions Limited also employs a full-time administrative manager to manage the Business IT Services division, Fiona Evans, and another one to manage the project, Pawel Kowalski.

Fiona is liaising with Pawel in order to ensure the quality of service delivery for BooksWeLove, making sure that the expectations of everyone involved are met. She is also checking that extensive testing is being carried out, so that the application runs without problems and meets all functional and non-functional requirements. The system should be able to allow customers to:

- search and find a book using any appropriate way of searching (*e.g*. title, author, genre, publishing house, year of publication);
- pay using credit cards, Paypal, bitcoins or a customer/corporate account;
- review any books they have read;
- organise online meetings, if they are on the system with the special status that allows them to do this.

The system should also enable new authors and/or publishing houses who wish to join the scheme to create an account and input the necessary information.

The application should also operate a forum as well as a real-time chat application, where users could use any form of authentication for signing in (*e.g.* local account, social media account, *etc*.), and use the chat facility to also hold private conversations with any members who are online at the same time.

The system should also keep track of customers' contact details as well as invoices and payments. Both the general public as well as publishing houses and bookshops should be able to open a customer or merchant account respectively. BooksWeLove wants to give all its users (customers and publishing houses) the option of setting up a "hold account" where customers can top-up money and pay for their books this way or for publishing houses and bookshops to build up their payments and transfer them to their bank account whenever they want to increase security. Since BooksWeLove is a growing and ambitious project, it is increasingly taking on staff, both full-time, part-time and on specific short term contracts.

After graduating with first class honours, you have recently joined Ubiquitous Computing Ltd. (UCL), a small software house, specialising in test-driven software development. UCL has been approached by BooksWeLove to estimate the cost of development and deployment of an integrated computer-based system for running the project. BooksWeLove is in the process of growing, so that Mr Kowalski will be the office manager of a team of twelve – at the moment including himself and Ms Evans. The ten other staff to be hired will handle customer relations, marketing, HR administration, project and event support and finances. Each of these areas should have access to their own dedicated 'area' within the proposed system.

Senior full-time employees have greater permission to review and make changes in all sections, while junior employees can only look at their section and are limited to what they can change. The junior employees need to seek permission for larger, more elaborate, changes, or for changes in the areas that they do not have permission to access. Part-time employees have no permission for changes, and have to request any

changes they may need from a full-time employee with the relevant permissions.

Your manager at UCL, Dr Sussanah Rodriguez, has opted for a test-driven development model for the BooksWeLove project, since UCL has a great deal of experience and expertise in this area. She has asked you to develop an initial set of black box test cases.

Task 1

Identify the stakeholders and the relationships between them. Explain, briefly, how each will interface to the BooksWeLove system.

[10 marks]

Task 2

Design a set of black box test cases for the BooksWeLove project that cover the critical areas of the business. Your test cases should, at least, include the input, expected output, prerequisites, steps for conducting the test and an explanation of the purpose of each test case. Please also provide a rationale for the overall choice of test cases.

[50 marks]

Task 3

Dr Rodriguez is looking for ways the system can be made more "intelligent" and as helpful and assistive for its customers (*e.g.* offering personalised choices, identifying associations between purchase patterns, *etc*., while at the same time observing the new GDPR legislation). She recently attended the ICDM 2017 conference, IEEE International Conference on Data Mining Series, where she heard presentations about Data Mining (DM).

She has asked you to write a short report about DM techniques and tools. She is interested to learn how DM could be applied to the BooksWeLove project and in which parts, such that the maximum value will be gained while at the same time any customers' private data will be kept in line with the new regulations.

Write a report for Dr Rodriguez, explaining the concept of Data Mining, presenting tools and techniques that can be used for this, discussing which areas from the project system it might be applicable to and illustrating how the techniques could be applied to the BooksWeLove project.

[40 marks]

[Total 100 marks]

[END OF COURSEWORK ASSIGNMENT 1]