Northumbria University
NEWCASTLE

Department of Computer & Information Sciences

ASSESSMENT BRIEF	
Module Title:	Program Design and Development
Module Code:	KF5008
Academic Year / Semester:	2021-22 / Semester 1
Module Tutor / Email (all queries):	Ossama Alshabrawy ossama.alshabrawy@northumbria.ac.uk
% Weighting (to overall module):	60%
Assessment Title:	Algorithms
Date of Handout to Students:	10th October 2021
Mechanism for Handout:	Module Blackboard Site & Online session
Deadline for Attempt Submission by Students:	13th December 2021 11.59pm GMT
Mechanism for Submission:	All the code files and template solution document should be compressed (.zip format) and uploaded to Module Blackboard Site
Submission Format / Word Count	Please upload your files as a zip format only.
Date by which Work, Feedback and Marks will be returned:	24th January 2022
Mechanism for return of Feedback and Marks:	Mark and individual written feedback sheet will be uploaded to the Module Site on Blackboard. For further queries please email module tutor.

LEARNING OUTCOMES

The learning outcomes (LOs) for this module are:-

Knowledge & Understanding

LO1 A critical understanding of collection classes and be able to implement data structures such as lists, queues and trees together with algorithms to process these structures (e.g., searching and sorting).

LO2 A critical application of algorithm complexity and the big-O notation, including algorithms for data security, and use this to select suitable data structures for a particular scenario.

Intellectual / Professional skills & abilities

LO3 Analyse a problem, produce an object-oriented design for its solution and evaluate the object-oriented design documentation.

LO4 Employ object-oriented programming and apply use of collection classes and their algorithms.

Personal Values Attributes (Global / Cultural awareness, Ethics, Curiosity) (PVA)
LO5 Employ Ability to work in a team to produce a design model.

This assessment addresses learning outcomes LO1, LO2, LO3 and LO4.

Nature of the submission required:

The solution document and supporting code should be attached and submitted in a zip file. The zip file should identify student by name and ID. Only .zip files will be accepted.

Instructions to students:

This is an individual piece of work, and you must not work with others to construct your work. During the semester there are numerous opportunities to seek and get advice and support on your work, from tutors and peers but you must ensure you do not do work for others or copy work from others.

Academic Conduct:

You must adhere to the university regulations on academic conduct. Formal inquiry proceedings will be instigated if there is any suspicion of misconduct or plagiarism in your work. Refer to the University's regulations on assessment if you are unclear as to the meaning of these terms. The latest copy is available on the university website.

https://www.northumbria.ac.uk/about-us/university-services/academic-registry/quality-and-teaching-excellence/assessment/guidance-for-students/

If you need an extension:

Contact ask4Help. Tutors and Module tutors cannot change deadlines.

Make sure that your report is submitted on time. University regulations state that assignments submitted late without approval will incur a 10% reduction for the first 24hours then a zero mark after this.

You may apply for an extension of time to complete assessed coursework if there are personal circumstances which are unforeseen and unpreventable and have a serious effect on your ability to submit the work by the published hand-in deadline. You must submit an 'Application for Authorisation for Late Submission of Assessed Work' before the hand-in deadline. Appropriate medical certification, or other relevant evidence confirming the circumstances, must be provided. Information regarding this policy and procedure can be accessed below:

https://northumbria-cdn.azureedge.net/-/media/corporate-website/new-sitecore-gallery/services/academic-registry/documents/qte/assessment/guidance-for-students/pl,-d-,008-v004-late-submission-of-work-and-extension-requests-policy.pdf?modified=20200803152930

Disabled students

Contact the module lead tutor about reasonable adjustments. For example, dyslexic students might submit question 1 as a power point with an audio recording of material.

Errors

If any errors are found in this document, changes will be posted to the eLP (Blackboard). Versions will be clearly stated. All versions will be accepted.

INSTRUCTION OF ASSESSMENTS

Introduction

You have been hired to work for a company called GenetHoc.

You have been hired to look through the code and try to improve

- The overall speed of the program
- The overall software engineering of the solution.

GenetHoc Lts make a program that stores the DNA of a large number of criminals. The program has two facilities. The first is to take a DNA sample taken for a forensic analysis of the victim or site. It then searches the database for a match to the DNA sequence. In this software, DNA is stored in a string as a series of four letters.

As the database has grown, several problems have emerged. The first is the speed of checking samples. As the database has grown, the number of samples processed per hour has dropped. This is partly blamed by the team on the number of duplicate samples. So, the system was built to scan the database looking for duplicates before processing. This was coded by some gene technology experts who are experts in the field, not programming.

You have been given access to the source code but not the database, as this would break UK general data protection regulation (GDPR). Instead, you need to use a randomly generated test dataset. Scientists at GenetHoc have written several unit tests to ensure that your code is working to their standards.

Your first task is to improve the overall speed of the program. The engineers at the company believe code cannot go any faster than they have done it with it not working correctly. Speed improvements must be made to operate on their computers (which, for security reasons, they won't tell you what kinds of computers, so your solution must be general and not assume multiple processors). They have set up a bonus incentive system so that the faster your process, the more you get paid. Suppose your code does not process any more quickly than the code you are not get paid. How much you get paid depends upon how much faster than theirs your code goes.

STUDENT FAQ

Q1 DOCTOR MY CODE SUDDENLY STOPPED WORKING ! WILL I FAIL ?

1. Download the original project
2. Apply the changes you noted in section 6 back to the old code.
3. IF this fails See Q3

Q2. THIS IDE IS DRIVING ME MAD CAN I USE ANOTHER ONE?

You need to do the profiling in Netbeans. You can use another IDE for everything else. Professionals, while they have preferences, generally can switch IDEs. Many companies have a companywide IDE preference. Netbeans was selected because it was on the Image and had a working profiler with out plug-ins.

Q3. I SUCK AT PROGRAMMING JAVA WILL I FAIL THE COURSE?

A .YOU CAN SUBMIT PSUEDO CODE IN SECTION 6 OR NON-WORKING CODE.

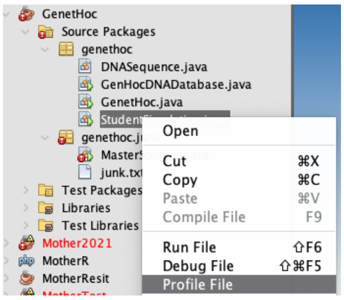
If you can demonstrate you can understand how to use a profiler and know what in principle you need to do, then you get partial marks. For Q6 if you say 'I would use faster code' then zero marks but if you say 'I would switch to a linked list as it has the ability to add at the front efficiently' then you can get a reasonable but not perfect mark.

Your tasks are as follows.

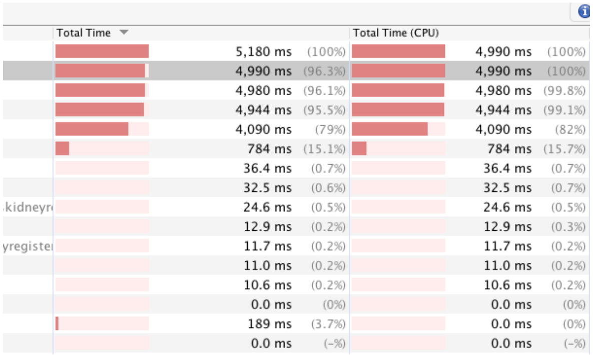
1. Look at the questions on blackboard. This is there to help you get through the exercise and make sure you have done everything.
2. Put your name and ID at the top of Question 3.
3. Based on **the last DIGIT** of your STUDENT ID name pick one of the following pairs
    - 2,6,0,
        i. Look at ArrayList and LinkedList
    - 3,7,1
        i. Look at TreeMap and HashMap
    - 4,8
        i. Look at TreeSet and HashSet
    - 9,5,
        i. Look at HashMap and ArrayList
    - IN A MAXIMUM OF **300** of YOUR OWN WORDS (300 each, 600 in total) – Pick a method from each class which is faster than the same method in the other class. Say which method that is and explain why it is faster (you might need to refer to how the class works internally).
    - If you have problems writing use things which have no word penalty
        i. Diagrams
        ii. Example code
    - For each data method write an additional 100 words on what real world circumstances might be faster for which purpose. You can make the real-world example up but it must be plausible.
    - Going over is permissible but if you go far over you will lose marks.
    - If you want to have references in a particular format use Harvard
    - IF YOU HAVE PROBLEMS GETTING STARTED GO TO THE LIBRARY AND GET ONE OF THE ALGORITHM BOOKS OUT.
    - This section will be checked for plagiarism

[15 MARKS]

4. Download the NetBeans project.
5. Investigate the current classes – check the documentation.
6. Run GeneSimApplication.
    - Do a performance analysis of the performance test in Netbeans.
    - To do this right click on StudentSimulation.java and select 'Profile File'



- This is your time to be a detective.
  i. Ask yourself what is slowing this code down?
  ii. Dig wide and deep. Use the profiler to look for clues.
- 6.1 this tests your ability to use and read a profiler. In your solution document (see the word template document), paste in a screen dump of your performance analysis when running (it should look something like the following figure but include the method names). You should unfold the profiler where necessary. Note profiling something other than StudentSimulation.java will get zero marks.



	Total Time		Total Time (CPU)	
	5,180 ms	(100%)	4,990 ms	(100%)
	4,990 ms	(96.3%)	4,990 ms	(100%)
	4,980 ms	(96.1%)	4,980 ms	(99.8%)
	4,944 ms	(95.5%)	4,944 ms	(99.1%)
	4,090 ms	(79%)	4,090 ms	(82%)
	784 ms	(15.1%)	784 ms	(15.7%)
	36.4 ms	(0.7%)	36.4 ms	(0.7%)
	32.5 ms	(0.6%)	32.5 ms	(0.7%)
kidneyr	24.6 ms	(0.5%)	24.6 ms	(0.5%)
	12.9 ms	(0.2%)	12.9 ms	(0.3%)
yregiste	11.7 ms	(0.2%)	11.7 ms	(0.2%)
	11.0 ms	(0.2%)	11.0 ms	(0.2%)
	10.6 ms	(0.2%)	10.6 ms	(0.2%)
	0.0 ms	(0%)	0.0 ms	(0%)
	189 ms	(3.7%)	0.0 ms	(0%)
	0.0 ms	(-%)	0.0 ms	(-%)

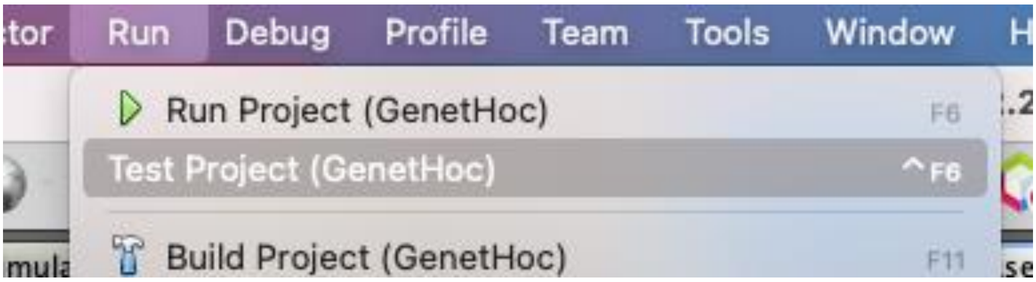
[**5 Marks**]

- 6.2 In your solution document identify which top 5 methods from the project code are the most time consuming (most slow the program down). Again, the function must be responsible for actually slowing your program down. The code which is causing the problem must be in the function. This is a test of your understanding of the profiler and telling the difference between methods/functions which are actually slowing the program down and not being distracted by things which are not slowing the program down. [**1 Mark each up to 5**]
- 6.3 Now examine the code inside the project. Which methods should receive priority for your attention? Put down what and why you choose them. You don't have to pick the same methods as in 6.2. You might pick another method for some reason. For example, one method might speed up two other methods which call it. Provided you have a good reason that's fine**. Put down your top 5 methods to change and why.**
  [2 Mark each up to 10]
- 6.4 For each of your top 5 targets, enter into your solution document what is your estimate of the O notation (e.g. O(n), O(n^2) , O(log(n)), etc.) time of this method and/or any methods it calls
  [**2 Marks each up to 10**]
- 6.5 Estimate the overall O-Notation for the project (that is what is the worst case for the worst code) ? [**5 marks**]
- 6.6 What is O-Notation for
  i. Simuation. removeDuplicateDNASequnces ( )

ii. Simuation. **containsWholeDNASequence** ( )
iii. Simulation. **searchForDNAFragment** ( )

[5 marks]

7. You should now alter the code to improve the performance. There is a convenient UNIT TEST called 'which takes your improved class as an argument'. It will print out what kind of speed increase you're getting. Use this to get an idea of how much you have achieved. Remember projects which fail to compile on the campus machines or don't pass the unit tests will score 0. Campus machines are running Netbeans version 11.3


tor
Run
Debug
Profile
Team
Tools
Window
H
Run Project (GenetHoc) F6
Test Project (GenetHoc) ^F6
Build Project (GenetHoc) F11
mula
se

- **NOTE: the unit test takes about 2-5 minutes to run be patient.**
- From the output you should see 'WARP-FACTOR ####' where #### is a number. You can convert this to your mark via the table on blackboard next to this assignment. Note this table may be revised only the final version is correct.
- Note the mark for this question is approximate – We may raise the score if we feel your code shows you know what you're doing, but you don't get the speed you expect. The top result is against the gold-standard version. You might be able to do better than the sample. If so up to 5 optional marks may be added if you exceed the gold standard in a genuine performance improving way. These additional marks won't be added if you go beyond 100 marks for this assessment.
- **Warning** - if you're submitted code fails to compile against Netbeans 11.3 on a Northumbria cluster machine, it fails (zero marks).
- **If** your code breaks unit tests, you lose ALL your marks for this question (zero marks).
- Again, the marks for this section will are scaled against your performance increase. That is, the faster you make the code, the more marks you get.
- **Warning.** Don't change the unit tests.
- **Warning.** Don't change the number of iterations.
- **Warning. For your** convenience**, try to** confine your changes to StudentSimulation.java – this is the only file that **needs** to be changed for maximum marks. Other changes are not forbidden but thought unnecessary and can still gain marks.
    i. **Note inside the code for StudentSimulation there is the code for an internal subclass of DNASequence – this is a convenience if you want to subclass DNASequence just uncomment the code.**
- SUBMIT JUST YOUR JAVA CODE
    i. (**Just the StudentSimulation.java file you changed is fine**).

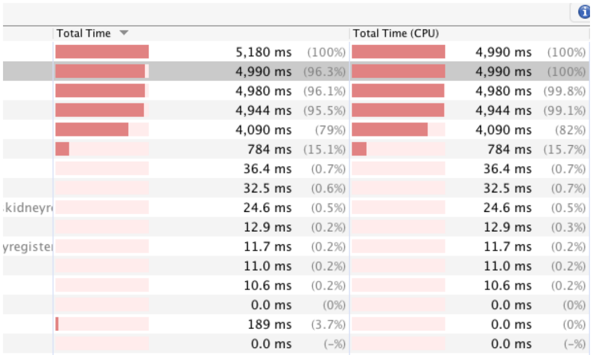
[**20 Marks**]

8. What did you change and why? **Code not working? See I can't compile note below**. Each row in this table will have three columns
    - FOR EACH CHANGE YOU MAKE
        i. Each row will have the original code before
        ii. Each row will have an after (your change) code copied from your source.

iii. The final column should contain your reasons for making that change. This can be short as long as it is clear. For example, 'Changed TreeSet to HashSet because HashSet has log(1) on deletion'.

[**5 marks for each row up to 20 Marks**]

9. In your solution document (the word document - see template) paste in a screen dump of your performance analysis AFTER YOUR NEW CODE. Note – many students didn't submit this last year and lost 5 marks



	Total Time		Total Time (CPU)	
	5,180 ms	(100%)	4,990 ms	(100%)
	4,990 ms	(96.3%)	4,990 ms	(100%)
	4,980 ms	(96.1%)	4,980 ms	(99.8%)
	4,944 ms	(95.5%)	4,944 ms	(99.1%)
	4,090 ms	(79%)	4,090 ms	(82%)
	784 ms	(15.1%)	784 ms	(15.7%)
	36.4 ms	(0.7%)	36.4 ms	(0.7%)
	32.5 ms	(0.6%)	32.5 ms	(0.7%)
kidneyr	24.6 ms	(0.5%)	24.6 ms	(0.5%)
	12.9 ms	(0.2%)	12.9 ms	(0.3%)
yregiste	11.7 ms	(0.2%)	11.7 ms	(0.2%)
	11.0 ms	(0.2%)	11.0 ms	(0.2%)
	10.6 ms	(0.2%)	10.6 ms	(0.2%)
	0.0 ms	(0%)	0.0 ms	(0%)
	189 ms	(3.7%)	0.0 ms	(0%)
	0.0 ms	(-%)	0.0 ms	(-%)

[**5 Marks**]

10. List out the classes you would restructure the code to.

Submission of materials

- **Your solution (word document**) should be submitted to blackboard via the link attached to the assessment area.
  - You can submit more than once but if you do check your mark reflects the last version when published.
- All code you modify should contain
  - Your name In a // comment.
  - Your student ID. In a // comment.
- **The Zipped file** of your java project should be submitted to blackboard.
  - Your zip file should contain your last name and first initial and id code i.e.
    - DaltonNW124030202022.zip
  - The file should not contain any .class files.
  - Any submissions which **only** contain .class files will not be accepted score 0
  - The zip file SHOULD ONLY include the files you have altered. So, you don't have to include the whole project
    - E.g., StudentSolution.java ☑

Assessment

- This assessment accounts for 60% of the marks for this module.
- There is a form to help you get all of the elements in the right place and allow for efficient marking and feedback.
- The unit test will indicate your mark for the speed test. This is a rough guide only. After all work is submitted this mark may be changed reflecting your tutors views of your understanding of what you did.
- Feedback
  - Feedback will be against your submission document on the blackboard.

FAQ

Help my code will not compile for some reason.

You can still get Marks for section 9. Simply put on which code you would fix. Feel free to put down some pseudocode to explain what you would do to fix it. If your reasoning is clear, it should be quite possible to get a passing mark for this part of the assessment - in other words part marks will be awarded.

Marking criteria Question 3

Grade	Criteria
12-15	Explanations are consistent with the classed. There is a reference to external sources to support the explanation. Both diagrams, code and text are used to support each other. More than one method is explained and the time implications for each method given. Real world examples given to support the explanation.
10-12	The overall work is good but is not outstanding. There are some flaws in the charts and reasoning, or there is a in fault logic. No real-world example(s) are given.
6-9	Work contains a number of problems, and there may be some good elements however there are sufficient problems. One, but not both explanations are incorrect.
0 - 5	Work in incomplete or contains significant flaws.

For Question 6

Grade	Criteria
2	The O notation correctly reflects the code in question and any method the code calls.
1	O notation is wrong by 1 (so N^3 when it should have been N^4)
1	The O notation is correct, but the code has no impact on the performance. Unit test code, system code like *println* relates to irrelevant N values such as length of player names.
0	Missing or incorrect. Code would not compile like this.

For Question 8 (for each area identified)

Grade	Criteria
5	The code identified is problematic. The replacement code is valid, or the pseudo code offered clearly expresses the solution. The explanation of the thinking for the change is clear.
4	The code identified is problematic. The replacement code is valid, or the pseudo code offered clearly expresses the solution. The explanation of the thinking for the change is clear.
3	The code identified is problematic. The explanation of the thinking for the change is clear. Replacement code is missing or incorrect.
2	The code identified is problematic. The replacement code is valid, or the pseudo code offered clearly expresses the solution. Explanation is unclear
1	The code identified is problematic.
0	There is no answer, or the code identified is not problematic.