DEPARTMENT OF PHYSICS & ASTRONOMY
PHAS3459 Final Examination
09:45 - 12:45: 11th December 2013

Please read the exam guidelines, rules, instructions and marking criteria at
`https://moodle.ucl.ac.uk/mod/wiki/view.php?pageid=15228`
(linked from the *Examinations* page).

This exam is worth 50% of your final mark for the course. The duration of the exam is 3 hours.

Students should upload the Java source code files for their solution using Moodle under the section headed "Exam II".

The code you upload must be self-contained: the marker must be able to compile and run it using only the classes uploaded and the Java API. If you use your own classes from earlier modules, make sure you upload them as well as any new classes you create during the exam.

Note: no knowledge of ecology or geography is required to complete this exam.

In this examination you will be processing simulated data from a survey of species on a fictional island. You will read data from a web directory, store the data in suitable containers and perform statistical analyses. You are strongly encouraged to read the entire exam paper before starting coding. Some formulas and definitions that may be useful are provided at the end of this question paper.

The input data files are provided in the following web directory:
`http://www.hep.ucl.ac.uk/undergrad/3459/exam-data/2013-14/.`

- The file `survey-plants.txt` contains records of individual specimens found during the survey. Each line of the file contains the following fields:
  - The latitude in degrees of the location of the specimen.
  - The longitude in degrees of the location of the specimen.
  - An identification code for the species of the specimen.
  - The height of the specimen in centimetres.
- The file `species-plants.txt` contains details of the species in the survey. Each line of the file contains the following fields:
  - The identification code for the species.
  - The full scientific name of the species, which has two parts separated by a space.

Part 1: 20/50 marks

Write a program to do the following:

- Read the data from `survey-plants.txt` and `species-plants.txt` and store it in one or more appropriate data structures.
- For each of the species, print the scientific name along with the number of specimens found and their mean height.
- Print the scientific names of the species with the highest and lowest mean height.

Part 2: 20/50 marks

- Create an interface with a method that will take a container of specimens and return a container that only includes the subset of the specimens that were found within some specific region.
- Create an implementation that will select the specimens that are found between a given minimum and maximum latitude.
- Use this to calculate the mean height of species *Urtica dioica* for specimens found
  - north of latitude -30° (30° south)
  - south of latitude -30°

- Create an implementation that will select the specimens found within a given distance of a specific point.
- Use this to calculate the mean height of species *Solanum carolinense* for specimens found within 50 km of the summit of the island's highest mountain, which is located at latitude -30.967° and longitude 75.430°.

Part 3: 10/50 marks

The files `survey-animals.txt` and `species-animals.txt` contain similar records of animal species and specimens that were identified in the survey, but only the location and species of each specimen were recorded, not its height.

By reusing or adapting your existing code, or otherwise, write a program to list the scientific names of all animal species that are *exclusively* found within 100 km of the mountain summit at latitude -30.967° and longitude 75.430°.

Formulas and definitions

For the purposes of this exam the Earth can be taken to be a sphere of radius 6371 km.

The position of any point on the surface of the Earth can be specified using its *latitude* and *longitude*. The latitude is the angle measured northward from the equator, so takes values from -90° (at the south pole) to +90° (at the north pole). The longitude is the angle measured eastward from the Prime Meridian, a north-south line that was chosen for historical reasons to pass through Greenwich. The longitude takes values from -180° to +180°: positive values are east of Greenwich and negative values are west, but +180° and -180° are equivalent.

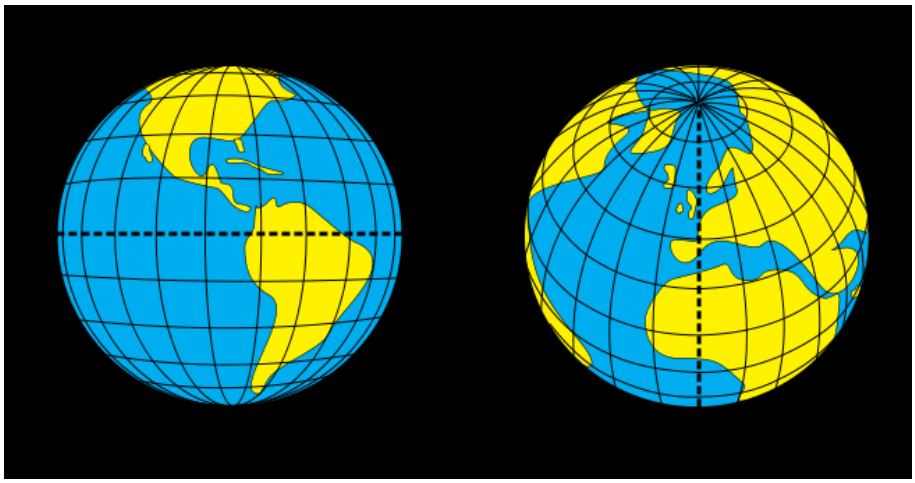

Figure 1: Latitude and longitude.

The distance d between two points on a sphere can be calculated using the *haversine formula*:

$$d = 2r \arcsin(\sqrt{h})$$

Here r is the radius of the sphere and h is defined as

$$h = \text{haversin}(\phi_2 - \phi_1) + \cos(\phi_1)\cos(\phi_2)\text{haversin}(\lambda_2 - \lambda_1)$$

where ϕ_1 and ϕ_2 are the latitudes of the points and λ_1 and λ_2 are their longitudes, and haversin is the *haversine* function, defined as

$$\text{haversin}(\theta) = \sin^2\left(\frac{\theta}{2}\right) = \frac{1 - \cos(\theta)}{2}$$

END OF PAPER