4. Analyzing words in a text file

The purpose of this project is to write a program that

1. Reads a text from a file.
2. Separates words from the text and converts them to lowercase.
3. Sorts the words into a binary tree (so that for each node words before it are on the left branch of the tree and those after are on the right) and stores the number of times the word occurs in the text.
4. Prints out the words and their occurrences in alphabetical order.

A word in the text is a string containing

- letters, digits or hyphens

separated by

- spaces, tabs, commas, periods or line feeds.

The example program `binarytree_gv.f90`[9] can be used as a basis for the binary tree operations. It also contains commands to generate a png image of the tree[10] which may be useful for debugging the code. When the word is extracted from the text it is either inserted as a new node in the tree or the occurrence count of an existing node is increased.

Note that you need not recognize different forms of a single word (i.e. different inflections). So, you may assume that *cat* and *cats* are two different words.

9 http://www.courses.physics.helsinki.fi/fys/tilaII/files/binarytree_gv.f90

10 The Linux package `graphviz` is needed to generate the image.