Artificial Intelligence 1 (3BA2)
Practical 3, Michaelmas Term Lecture Week 4
November 3, 2004

Text Processing

Building on the ideas we developed in the last practical, build a complete Prolog program to analyse the text of a book and to output a list of every word, along with its frequency of use, and ordered so that the most frequently used word is listed first, with the least frequently used word listed last. You can get the texts of lots of [older] books on the web, for example, Project Gutenberg at `http://promo.net/pg/` has a very large selection.

Apart from the logic and control aspects of this program, there are *efficiency* and *implementation* issues as well. So, start with a short 'book' until you get an algorithmically efficient program working. You then need to make sure the program runs well in the context of the Prolog implementation. You should use 'cuts' judicially to remove choice points that you are certain will not be used. This gives the memory manager a much better opportunity to recover garbage, enabling it to run faster and on larger problems.