

1994 Paper 3 Question 1

Programming Language Compilation

It is required to write a program to gather information about the lexical structure of a source file. The file contains words composed of letters and digits starting with a letter, decimal integers composed of digits, and various other single- and multi-character symbols such as `:=`, `+`, `-`, `[`, `]`, and `->`. About 50 of the words are reserved, the others being identifiers.

It is required to count how often each symbol and reserved word occurs, the number of distinct identifiers used, and the mean value of all the integers that occur in the source file. Suggest how you would design such a program with the aid of Lex, paying particular attention to the overall structure of the program, and giving a detailed account of how you would deal with the identifiers. [20 marks]