LR(k) Sparse-Parsers and their Optimisation

by

J.M. Rushby

Department of Computer Science
University of Newcastle upon Tyne

Ph.D Thesis
September 1977

Acknowledgements

It is a pleasure to record my appreciation of the assistance and encouragement which I have received from my Parents, friends, and colleagues during the preparation of this thesis.

I am especially grateful to Dr. Tom Anderson and Dr. James Eve for their interest in this work, their advice, and their scrupulous and valuable criticism of preliminary drafts. l am also pleased to acknowledge the guidance of my supervisor, Mr. L.B. Wilson.

I am grateful to Mrs. L.C. Woodcock for her speedy and cheerful typing of a difficult manuscript and to Gill Nyfield for typing an earlier draft.

Financial support for the research described here was received from the Science Research Council.

ABSTRACT

A method of syntactic analysis is developed which is believed to surpass all known competitors in all major respects.

The method is based upon that associated with the LR(k) grammars but is faster because it bypasses all reduction steps concerned with 'chain' productions. These are freely selected productions which are considered semantically irrelevant and whose right parts consist of just a single symbol. The parses produced by the method are 'sparse' in that they contain no references to chain productions - they are termed 'chain-free' parses.

The CFLR(k) grammars are introduced as the largest class which can be Chain-Free parsed from Left to Right while looking k symbols ahead of the current point of the parse. The properties of these grammars are examined in detail and their relationship to the conventional LR(k) grammars is explored. Techniques are presented for testing grammars for the CFLR(k) property and for constructing chain-free parsers for those grammars possessing the property. Methods are also presented for converting ordinary LR(k) parsers into chain-free parsers.

CFLR(k) parsers are more widely applicable than their LR(k) counterparts, are faster and provide the same excellent detection of syntactic errors. Unfortunately they also tend to be rather larger. A simple optimization is presented which completely overcomes this single disadvantage without sacrificing any of the advantages of the method.

These theoretical techniques are adapted to provide truly practical chain-free parsers based on the conventional SLR and LALR parsing methods. Detailed consideration is given to use of 'default reductions' and related techniques for achieving compact representations of these parsers. The resulting chain-free parsers are not only faster than their ordinary counterparts, but probably smaller too. We believetheir advantages are such that they should substantially replace other parsing methods currently used in programming language compilers.

Contents