The systematic construction of Earley Parsers: Application to the production of an $O(n^6)$ Earley Parser for Tree Adjoining Grammars

Bernard Lang
INRIA
B.P. 105
78153 Le Chesnay, CEDEX, France
lang@margaux.inria.fr

Logic Programming languages (Prolog) were originally introduced as an extension of context-free (CF) languages. Conversely CF grammars may be seen as logic programs (or Horn clauses) where the predicates are the CF grammar symbols, and where these predicates have arguments corresponding to the boundaries of the input string fragments they derive into.

Earley's parsing algorithm can be generalized to Horn Clauses as a dynamic programming evaluation technique. Keeping in mind the relation between Horn clauses and CF grammars, we suggest encoding similarly in Horn Clauses other syntactic formalisms so as to take advantage of this generalisation of Earley's algorithm to obtain for free efficient parsers for these encoded formalisms.

As an example, we consider the problem of TAG parsing. We show that any TAG can be encoded into a logic program for which there is an evaluation in time $O(n^6)$. We show on this example how the general dynamic programming procedure can be adapted to conform the constraint that sentences be parsed from left to right, even in the presence of interleaved constituents as is the case for TAGs.

The Valid Prefix Property and Parsing Tree Adjoining Grammars

Yves Schabes
Department of Computer and Information Science
R-555 Moore School
University of Philadelphia
220 South Street 33rd Street
Philadelphia, PA 19104-6389, USA
schabes@linc.cis.upenn.edu

The valid prefix property (VPP), capability of a left to right parser to detect errors as soon as possible, is often unobserved in parsing CFGs. Earley's parser for CFG maintains the VPP and obtains a worst case complexity ($O(n^3)$) as good as parsers that do not maintain VPP (as the CKY parser). Contrary of CFGs, maintaining the valid prefix property for TAGs seems costly.

The aim of talk was to informaly explain why the VPP for TAGs seems expensive to maintain and also to introduce a new Earley-style parser for TAGs which has $O(n^6)$ worst case time complexity. The new parser does not maintain VPP but it can