Bottum-up tree acceptors

_Citation for published version (APA):_

_Document status and date:_
Published: 01/01/1988

_Publisher Version:_
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

_Please check the document version of this publication:_
- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

_General rights_
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

_Take down policy_
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
Bottom-up tree acceptors

by

C.Hemerik and J.P. Katoen

October, 1988
This is a series of notes of the Computing Science Section of the Department of Mathematics and Computing Science Eindhoven University of Technology. Since many of these notes are preliminary versions or may be published elsewhere, they have a limited distribution only and are not for review. Copies of these notes are available from the author or the editor.
Abstract

This paper deals with the formal derivation of an efficient tabulation algorithm for table-driven bottom-up tree acceptors. Bottom-up tree acceptors are based on a notion of match sets. First we derive a naive acceptance algorithm using dynamic computation of match sets. Tabulation of match sets leads to an efficient acceptance algorithm, but tables may be so large that they cannot be generated due to lack of space. Introduction of a convenient equivalence relation on match sets reduces this effect and improves the tabulation algorithm.

1 Introduction

Nowadays, many parts of a compiler can be generated automatically. For instance, automatic generation of lexical and syntactic analyzers using notations based on regular expressions and context-free grammars is commonly used (see e.g. [Aho]). However, much research is still going on in the field of universal code generator-generators, which take a description of a machine as input and deliver a (good) code generator for that machine.

Code generation forms an important subject in developing a compiler. Requirements traditionally imposed on a code generator are severe: the generated code must be correct and must utilize the resources of the machine (such as registers) efficiently. A particular and relevant issue in code generation is instruction selection. This forms the subject of the remainder of this section. Nature of the instruction set and addressing modes of the target machine determine the difficulty of instruction selection. As an example, we illustrate the instruction selection of an expression for a register machine with a very simple instruction set. First the addressing modes are presented.
Example 1.1

<table>
<thead>
<tr>
<th>addressing mode</th>
<th>format</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>immediate</td>
<td>#c</td>
<td>c</td>
</tr>
<tr>
<td>register</td>
<td>Ri</td>
<td>Ri</td>
</tr>
<tr>
<td>indexed</td>
<td>c(Ri)</td>
<td>M(c + Ri)</td>
</tr>
<tr>
<td>indirect</td>
<td>*Ri</td>
<td>M(Ri)</td>
</tr>
</tbody>
</table>

Here, \( M(a) \) denotes the contents of address \( a \); \( c \) is a constant and \( R_i \) is a register. Suppose our target machine supports the following instructions.

Example 1.2

<table>
<thead>
<tr>
<th>instruction</th>
<th>definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOV #c, Ri</td>
<td>( R_i := c )</td>
</tr>
<tr>
<td>MOV *Rj, Ri</td>
<td>( R_i := M(R_j) )</td>
</tr>
<tr>
<td>MOV c(Rj), Ri</td>
<td>( R_i := M(c + R_j) )</td>
</tr>
<tr>
<td>ADD Ri, Rj</td>
<td>( R_i := R_i + R_j )</td>
</tr>
</tbody>
</table>

Now consider the expression \( R_1 := c_1 + M(c_2 + R_2) \). Using the instructions given above we may derive an instruction sequence for this expression as follows. At each step the selected subexpression to code is underlined.

Example 1.3

<table>
<thead>
<tr>
<th>expression</th>
<th>instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R_1 := c_1 + M(c_2 + R_2) )</td>
<td>MOV #c_2, R_3</td>
</tr>
<tr>
<td>( R_1 := c_1 + M(R_3 + R_2) )</td>
<td>ADD R_2, R_3</td>
</tr>
<tr>
<td>( R_1 := c_1 + M(R_2) )</td>
<td>MOV *R_2, R_2</td>
</tr>
<tr>
<td>( R_1 := c_1 + R_2 )</td>
<td>MOV #c_1, R_1</td>
</tr>
<tr>
<td>( R_1 := R_1 + R_2 )</td>
<td>ADD R_1, R_2</td>
</tr>
</tbody>
</table>

Observe that the derivation above looks like the parsing of a string. By replacing the definition of an instruction, which is of the form \( R_i := \ldots \), by a production rule of the form \( R_i \rightarrow \ldots \), definitions (1)-(4) of example 1.2 may be considered as a code generation grammar.

The above suggests to use traditional parsing techniques for code generation. For instance, Graham & Glanville use LR-parsing for instruction selection (see [Graham]). Main problem with this approach is the solution of the large number of parsing conflicts caused by the fact that code generation grammars are inherently highly ambiguous. For example, an alternative instruction sequence may be derived for the expression \( R_1 := c_1 + M(c_2 + R_2) \) as follows:
Example 1.4

expression  
\[ R_1 := c_1 + M(c_2 + R_2) \]
\[ R_1 := c_1 + R_2 \]
\[ R_1 := R_1 + R_2 \]

instruction  
MOV \( c_2(R_2) \), \( R_2 \)
MOV \( #c_1, R_1 \)
ADD \( R_1, R_2 \)

A way to overcome this problem could be to use a more general parsing method like Earley parsing, as suggested in [Christopher a], but the resulting space and time complexity is unacceptable for practical use in code generators.

Neither of these methods just mentioned takes into account a special property of code generation grammars, viz. that every operator symbol has a fixed rank. Using this fact leads to the idea of considering the tree representation of an expression, rather than its string representation. To this end, the code generation (string) grammar becomes a so-called tree grammar. For the instructions of example 1.2 the production rules, represented by trees, become:

Example 1.4

instruction  
MOV \( #c, R_i \)
MOV \( *R_j, R_i \)
MOV \( c(R_j), R_i \)
ADD \( R_i, R_j \)

tree representation  
\[ R_i \rightarrow c \]
\[ R_i \rightarrow R_j M \]
\[ R_i \rightarrow c + R_j \]
\[ R_i \rightarrow R_j + R_j \]

Several code generation algorithms based on tree grammars have been described [Aho, Christopher b, Turner], but a theoretical framework is painfully missing. This is the more remarkable as a well-developed theory of tree grammars and tree automata already exists for some twenty years [Brainerd, Doner, Rounds, Hoffmann]. A systematic treatment of this theory, aimed at code generation applications, is given in [Van Dinther]; a survey paper is in preparation [Hemerik].

In this paper we consider a particular class of tree acceptors, called deterministic bottom-up tree acceptors, which have a time complexity proportional to the size of the tree to be analyzed. They can easily be extended to bottom-up tree parsers. Our main purpose is to present algorithms
for the efficient generation of compressed parse tables, and to show how these rather complex programs can be derived systematically.

The organization of this paper is as follows: in sections 2 and 3 we present a simplified treatment of the theory of tree grammars and deterministic bottom-up tree acceptors. Section 4 shows how the transition functions of the acceptor can be tabulated, thus leading to a linear time acceptance algorithm. In practical applications the size of the resulting tables may be prohibitive however. Therefore in section 5, using ideas of [Chase], an improved algorithm is described which generates compressed transition tables. Finally, section 6 contains some concluding remarks.

2  Tree grammars

In this section we define the basic concepts of the theory of tree grammars. Readers familiar with context-free (string-) grammars will notice that tree grammars are a generalization of context-free (string-) grammars.

Definition 2.1 (ranked alphabet 1)
A ranked alphabet is a pair \((V, r)\) such that \(V\) is a finite set and \(r : V \rightarrow \mathbb{N}\)

Elements of \(V\) are called symbols and \(r(a)\) is called the rank of symbol \(a\). In the following the set \(V_n\) denotes the set of symbols with rank \(n\), that is, \(V_n = \{v \in V \mid r(v) = n\}\).

Definition 2.2 (Tree\((V, r)\), trees over a ranked alphabet)
The set \(\text{Tree}(V, r)\) of trees over a ranked alphabet \((V, r)\) is the smallest set \(X\) such that:

- \(V_0 \subseteq X\)
- \(\forall n : 1 \leq n : \forall a \in V_n : \forall t_1, \ldots, t_n \in X : a(t_1, \ldots, t_n) \in X\)

Definition 2.3 (tree grammar)
A tree grammar \(G\) is a 5-tuple \((N, V, r, P, S)\) such that:

- \((N \cup V, r)\) is a ranked alphabet such that \(\forall A \in N : r(A) = 0\)
- \(N \cap V = \emptyset\)
- \(P\) is a finite subset of \(N \times \text{Tree}(N \cup V, r)\)
- \(S \in N\)

1 throughout this paper \(\mathbb{N}\) denotes the set of natural numbers
Elements of $N$, $V$, $P$ are called nonterminals, terminals, and production rules, respectively. $S$ is called the start symbol of $G$. Notational remark: upper-case letters are used to denote nonterminals, lower-case letters stand for terminals. An element $(A, t) \in P$ is usually written as $A \rightarrow t$; $A$ is sometimes called the left-hand side, $t$ the right-hand side of the production rule.

**Definition 2.4** \{ $\longrightarrow$, derivation-step (informal) \}

Let $(N, V, r, P, S)$ be a tree grammar. $\forall t_1, t_2 \in \text{Tree}(N \cup V, r)$:

\[
\begin{align*}
  t_1 & \rightarrow t_2 \text{ if } \exists (A \rightarrow \alpha) \in P, \text{ such that } t_2 \text{ can be obtained from } \\
  t_1 & \text{ by substituting } \alpha \text{ for one occurrence of } A \text{ in } t_1.
\end{align*}
\]

$\overset{\longrightarrow}{\rightarrow}$ is the reflexive and transitive closure of $\rightarrow$. We say that $t$ is derivable from $A$ if $A \overset{\longrightarrow}{\rightarrow} t$.

The set of trees, containing only terminals, derivable from the start symbol constitute the language that is generated by the corresponding tree grammar.

**Definition 2.5** \{ $\mathcal{L}$, language generated by a tree grammar \}

Let $G = (N, V, r, P, S)$ be a tree grammar.

- the function $\mathcal{L} : N \rightarrow P(\text{Tree}(V, r))$ is defined by:
  \[
  \forall A \in N : \mathcal{L}(A) = \{ t \in \text{Tree}(V, r) \mid A \overset{\longrightarrow}{\rightarrow} t \}\n  \]
- the language generated by $G$ is $\mathcal{L}(S)$

The tree grammar defined in the following example shall be used as a running example throughout this paper.

**Example 2.6**

Let $G = (N, V, r, P, A)$ be a tree grammar, where $N = \{ A, B \}$; $V = \{ a, b, c, d \}$;

\[
\begin{align*}
  r(a) &= 2 ; \\
  r(b) &= 1 ; \\
  r(c) &= 0 ; \\
  r(d) &= 0 ;
\end{align*}
\]

\[
\begin{align*}
P &= \{ (1) & A \rightarrow a(b(c), B), \\
  (2) & A \rightarrow a(B, d), \\
  (3) & A \rightarrow c, \\
  (4) & B \rightarrow b(B), \\
  (5) & B \rightarrow A, \\
  (6) & B \rightarrow d \} \n\end{align*}
\]

An alternative presentation of the elements of $P$ is given in figure 1. Some examples of derivations are :
3 TREE ACCEPTORS

Figure 1: Production rules represented as trees

\[
(1) \quad A \rightarrow a \\
(2) \quad A \rightarrow a(b(c)) \\
(3) \quad A \rightarrow a(b(c), B) \\
(4) \quad B \rightarrow b \\
(5) \quad B \rightarrow A \\
(6) \quad B \rightarrow (b(c), d)
\]

Some elements of \( \mathcal{L}(G) \) are: \( c, a(b(c), d), a(a(b(c), b(d)), d) \).

3 Tree acceptors

A tree acceptor is a tree automaton which, given a tree grammar \( G = (N, V, r, P, S) \) and a tree \( t \in \text{Tree}(V, r) \), establishes whether \( t \in \mathcal{L}(G) \). In this section we consider a particular kind of tree acceptors, viz. deterministic bottom-up tree acceptors, although we shall not stress the automata-theoretic concepts. The basic idea underlying this kind of acceptor is to extract from the grammar \( G \) a set \( PS \) of patterns, i.e. subtrees of right-hand sides of production rules, and to compute for a tree \( t \) the match set \( MS_1(t) \) of patterns from which it may be derived. The tree \( t \) is accepted if and only if \( S \in MS_1(t) \). As the match set of a computed tree can be simply computed from the match sets of its direct subtrees, an acceptor can be obtained which operates in time proportional to the size of the tree.

The following definitions are all relative to a given tree grammar \( G = (N, V, r, P, S) \). We assume that \( G \) has no useless terminals and nonterminals, i.e. every (non)terminal occurs in some tree derivable from \( S \) and for all \( A \in N : L(A) \neq \emptyset \).

**Definition 3.1 { sub, subtree relation }**

\[ \forall n : 1 \leq n : \forall \alpha \in V_n : \forall t_1, \ldots, t_n \in \text{Tree}(N \cup V, r) : \forall i : 1 \leq i \leq n : t_i \text{ sub } a(t_1, \ldots, t_n) \]
is the reflexive and transitive closure of \( \text{sub} \).

**Definition 3.2** \( PS, \) **pattern set**
\[
PS = \{ t \in \text{Tree}(N \cup V, r) \mid \exists A, t': A \in N \land t' \in \text{Tree}(N \cup V, r) : (A \rightarrow t') \in P \land t \overset{\text{sub}^*}{\longrightarrow} t' \}
\]

Notice that \( N \subseteq PS \) holds, since every nonterminal occurs in some tree derivable from \( S \). The closure of a pattern \( s \) is the set of patterns containing \( s \) and the nonterminals from which \( s \) is derivable. Similarly, the closure of a set of patterns is defined as follows.

**Definition 3.3** \( \text{closure} \)

- the function \( \text{closure} : \mathcal{P}(PS) \rightarrow \mathcal{P}(PS) \) is defined by:
  - \( \forall s \in \mathcal{P}(PS) : \text{closure}(s) = s \cup \{ A \in N \mid \exists \alpha \in s : A \xrightarrow{\alpha} \} \)

Notice that for all \( s \in \mathcal{P}(PS) : \text{closure}(s) \subseteq PS \). There are various ways to handle the acceptance problem. One possible way, commonly known as the bottom-up method (or bottom-up pattern matching), is to derive the start symbol \( S \) starting with a given tree \( t \). The bottom-up method relies on the notion of **match sets**, sets of subpatterns that match at a particular tree node. These sets are defined recursively as:

**Definition 3.4** \( MS_1, \) **match set**

- the function \( MS_1 : \text{Tree}(V, r) \rightarrow \mathcal{P}(PS) \) is defined by:
  - \( \forall a \in V_0 : MS_1(a) = \text{closure}\{a\} \)
  - \( \forall n : 1 \leq n : \forall a \in V_n : \forall t_1, \ldots, t_n \in \text{Tree}(V, r) : 
    MS_1(a(t_1, \ldots, t_n)) = \text{closure}\{a(p_1, \ldots, p_n) \in PS \mid \forall i : 1 \leq i \leq n : p_i \in MS_1(t_i)\} \)

The relevance of match sets is stated in the following lemma.

**Lemma 3.5** \( \forall t \in \text{Tree}(V, r) : MS_1(t) = \{ t' \in PS \mid t' \overset{\text{sub}^*}{\longrightarrow} t \} \)

**Proof** : by structural induction over Tree\((V, r)\).

1. base step : let \( a \in V_0,
   \[
   MS_1(a) = \{ \text{definition 3.4} \} 
   \]
3 TREE ACCEPTORS

\[
\text{closure}\{a\} \\
= \{\text{definition 3.3}\} \\
= \{a\} \cup \{A \in N \mid A \xrightarrow{*} a\} \\
= \{N \subseteq PS\} \{a \in PS \land a \xrightarrow{*} a\} \\
= \{t' \in PS \mid t' \xrightarrow{*} a\}
\]

2. induction step : let \(a \in V_n, 1 \leq n\), then

\[
MS_1(a(t_1, \ldots, t_n)) \\
= \{\text{definition 3.4}\} \\
= \text{closure}\{a(p_1, \ldots, p_n) \in PS \mid \forall i : 1 \leq i \leq n : p_i \in MS_1(t_i)\} \\
= \{\text{induction hypothesis}\} \\
= \text{closure}\{a(p_1, \ldots, p_n) \in PS \mid \forall i : 1 \leq i \leq n : p_i \xrightarrow{*} t_i\} \\
= \{\text{set calculus}\} \\
= \text{closure}\{a(p_1, \ldots, p_n) \in PS \mid a(p_1, \ldots, p_n) \xrightarrow{*} a(t_1, \ldots, t_n)\} \\
= \{\text{definition 3.3}\} \{\forall s \in P(PS) : \text{closure}(s) \subseteq PS\} \\
= \{t' \in PS \mid t' \xrightarrow{*} a(t_1, \ldots, t_n)\}
\]

\[\square\]

**Lemma 3.6** \(\forall t \in \text{Tree}(V, r) : S \in MS_1(t) \iff t \in \mathcal{L}(S)\)

**Proof** : use definition 2.5 and lemma 3.5. \(\square\)

So, by computing \(MS_1(t)\) for a given tree \(t\), it is rather simple to decide whether \(t\) belongs to the language generated by \(G\) or not.

**Example 3.7**

Consider the grammar of example 2.6. Its pattern set is :

\(PS = \{a(b(c), B), b(c), c, B, a(B, d), d, b(B), A\}\).

For \(t = a(b(c), b(a(d), d))\) bottom-up computation of \(MS_1(t)\) is depicted in figure 2, where each node of \(t\) is annotated with the match set of the tree rooted at that node.

It shows that \(A \in MS_1(t)\), hence (see lemma 3.6), \(t \in \mathcal{L}(A)\). \(\square\)

Notational remark : elements of a match set added by a closure operation are separated from other elements by a semicolon.
4 TABULATION OF MATCH SETS

A program computing $MS_1$ is easy implementable, but very inefficient. At each determination of the acceptance of a tree, match sets (and closures) must be recalculated. Fortunately, since $\mathcal{P}(PS)$ is a finite set (due to the fact that $PS$ is finite) the number of match sets is finite. This gives the possibility of tabulation of match sets. Observe that $MS_1(a(t_1, \ldots, t_n))$ is of the form $fa(\mathcal{M}_{s1}(a), \ldots, \mathcal{M}_{sn}(a))$, where $fa$ is defined as follows.

**Definition 4.1** \{ $fa$, transition function for symbol $a$ \}

- $\forall a \in V_0 : f_a \in \mathcal{P}(PS)$, where $f_a = \text{closure}\{a\}$
- $\forall n : 1 \leq n : \forall a \in V_n : \forall s_1, \ldots, s_n \in \mathcal{P}(PS) : f_a \in \mathcal{P}(PS)^n \rightarrow \mathcal{P}(PS)$, where $f_a(s_1, \ldots, s_n) = \text{closure}\{a(p_1, \ldots, p_n) \in PS | \forall i : 1 \leq i \leq n : p_i \in s_i\}$

$\square$

Obviously all $f_a$'s have a finite domain (since the number of match sets is finite), so we may tabulate $f_a$. This means that we will have an $n$-dimensional table for a symbol of rank $n$. We do not have to tabulate $f_a$ for the entire powerset $\mathcal{P}(PS)$, but only for its reachable part, i.e. the smallest set $Z \subseteq \mathcal{P}(PS)$ closed under all $f_a$'s.

To compute the reachable part $Z$ and the tabulation of $f_a$, $T_a \in Z^n \rightarrow Z$, where $n = r(a)$, the standard reachability algorithm (see e.g. [Rem]) is used. This leads to the following algorithm, which is called $A_1$. 

![Figure 2: An example of dynamic computation of match sets](image)
4 TABULATION OF MATCH SETS

\[
\begin{array}{l}
\text{\textbf{var} } x, y : \mathcal{P}(PS) \\
\quad | Z, W, G := \emptyset, \mathcal{P}(PS), \emptyset \\
\quad \text{; for all } a \in V_0 \\
\quad \text{do } y := f_a; W, G := W \setminus \{y\}, G \cup \{y\}; T_a := y \text{ od} \\
\quad \text{; do } G \neq \emptyset \rightarrow x \in G \\
\quad \text{; for all } a \in V \setminus V_0 \\
\quad \text{do } \| \text{ var } n : \mathcal{N} \\
\quad \quad | n := r(a) \\
\quad \quad \| \text{; for all } (s_1, \ldots, s_n) \in (Z \cup \{x\})^n \setminus Z^n \\
\quad \quad \text{do } y := f_a(s_1, \ldots, s_n) \\
\quad \quad \quad \text{; if } y \in W \rightarrow W, G := W \setminus \{y\}, G \cup \{y\} \\
\quad \quad \quad \quad | y \notin W \rightarrow \text{skip} \\
\quad \quad \quad \| \text{; } T_a(s_1, \ldots, s_n) := y \\
\quad \quad \text{od} \\
\quad \text{od} \\
\quad \text{; } G, Z := G \setminus \{x\}, Z \cup \{x\}
\end{array}
\]

To allow indexing of transition tables with numbers rather than match sets, we introduce an enumeration \( E \in \mathcal{N} \rightarrow_p Z \) of match sets\(^2\). \( E \) is an injection. The definition of transition tables is stated in terms of the enumeration as follows.

**Definition 4.2** \( \{ T_a, \text{ transition table for symbol } a \} \)

- \( \forall a \in V_0 : T_a \in \text{dom}(E), \text{ where } E(T_a) = f_a \)
- \( \forall n : 1 \leq n \land \forall a \in V_n : \forall s_1, \ldots, s_n \in \text{dom}(E) : T_a \in \text{dom}(E)^n \rightarrow \text{dom}(E), \text{ where } E(T_a(s_1, \ldots, s_n)) = f_a(E(s_1), \ldots, E(s_n)) \)

\[\square\]

The corresponding definition of match set is now.

\(^2\) we use \( \rightarrow_p \) to denote a partial function
4 TABULATION OF MATCH SETS

Definition 4.3 \{ MS_2, match set \}

- the function $MS_2 \in \text{Tree}(V, r) \rightarrow \text{dom}(E)$ is defined by :
- $\forall a \in V_0 : MS_2(a) = T_a$
- $\forall n : 1 \leq n : \forall a \in V_n : \forall t_1, \ldots, t_n \in \text{Tree}(V, r) :$
  $MS_2(a(t_1, \ldots, t_n)) = T_a(MS_2(t_1), \ldots, MS_2(t_n))$

The correspondence between $MS_1$ and $MS_2$ is stated in the following lemma.

Lemma 4.4 $\forall t \in \text{Tree}(V, r) : MS_1(t) = E(MS_2(t))$

Proof: by structural induction over $\text{Tree}(V, r)$, using definitions 3.4 and 4.1-4.3.

Match sets containing the start symbol $S$ have a special meaning (see lemma 3.6) and are called accepting states.

Definition 4.5 \{ F, accepting states \}

$F = \{ n \in \text{dom}(E) | S \in E(n) \}$

Lemma 4.6 $\forall t \in \text{Tree}(V, r) : t \in L(S) \Leftrightarrow MS_2(t) \in F$

Proof: use definition 4.5 and lemmata 4.4 and 3.6.

As can be observed from definition 3.3, taking the closure of a set of patterns just consists of deriving some left-hand sides of production rules. This calculation can be simplified by tabulating the closure of nonterminals in a table Nclosure. Formally:

Definition 4.7 \{ Nclosure \}

- $Nclosure \in N \rightarrow \mathcal{P}(N)$
- $\forall A \in N : Nclosure(A) = \{ B \in N | B \xrightarrow{*} A \}$

The Nclosure of a nonterminal is nothing more than determining the reflexive and transitive closure of $P \cap N \times N$. We use Warshall's algorithm to calculate the transitive closure.

The relation between closure and Nclosure is stated in lemma 4.8.

Lemma 4.8 $\forall s \in \mathcal{P}(PS) : closure(s) = s \cup (\bigcup A, \alpha | (A \rightarrow \alpha) \in P \land \alpha \in s : Nclosure(A))$

Proof: use definitions 3.3, 2.4, and 4.7.

The elaborated version of algorithm $A_1$ (named $A_2$) is presented below. Here, the sets $Z, G,$ and $W$ are characterized by $\{ E(i) | 0 \leq i < p \}, \{ E(i) | p \leq i < q \},$ and $\mathcal{P}(PS) \setminus (Z \cup G)$, respectively.


4 TABULATION OF MATCH SETS

\[ G = (N, V, r, P, S) : \text{tree grammar} \]

\[
; \text{var } E : N \rightarrow \mathcal{P}(P(S))
\]

\[
; F : \mathcal{P}(N)
\]

\[
p, q : N
\]

\[
T_a : N^* \rightarrow N
\]

\[
T_0 : N^0 \rightarrow N
\]

\[
N \text{ closure} : N \rightarrow \mathcal{P}(N)
\]

\[
; \text{proc } \text{compute}_N \text{ closure} =
\]

\[
[| (* \text{ first, transitive closure by means of Warshall's algorithm } *) \]
\]

\[
\text{for all } A \in N \text{ do } \text{N closure}(A) := \{ B \in N \mid (B \rightarrow A) \in P \} \text{ od}
\]

\[
; \text{for all } B \in N
\]

\[
\text{do } \text{for all } A \in N
\]

\[
\text{do } \text{if } B \in \text{N closure}(A) \rightarrow \text{N closure}(A) := \text{N closure}(A) \cup \text{N closure}(B)
\]

\[
| B \notin \text{N closure}(A) \rightarrow \text{skip}
\]

\[
\text{fi}
\]

\[
\text{od} (* \text{ second, reflexive closure } *)
\]

\[
; \text{for all } A \in N \text{ do } \text{N closure}(A) := \text{N closure}(A) \cup \{ A \} \text{ od}
\]

\[
\]

\[
; \text{func } \text{ closure} =
\]

\[
(1 s : \mathcal{P}(P(S)) \mid \mathcal{P}(P(S))
\]

\[
| r := s
\]

\[
; \text{for all } (A \rightarrow \alpha) \in P
\]

\[
\text{do } \text{if } \alpha \in s \rightarrow r := r \cup \text{N closure}(A)
\]

\[
| \alpha \notin s \rightarrow \text{skip}
\]

\[
\text{fi}
\]

\[
\text{od}
\]

\[
| r
\]

\[
\]

\[
(* \text{ main program } *)
\]

\[
; \text{compute}_N \text{ closure}()
\]

\[
; q := 0 ; p := 0
\]

\[
; \text{for all } a \in V_0
\]

\[
\text{do } E(q) := \text{closure}(\{ a \}) (* \text{ new match set } *)
\]

\[
; T_0, q := q, q + 1
\]
Given the transition tables, the acceptance problem is easily solved: some simple table look-ups do the job. The table-driven acceptor is described by the following algorithm. Time complexity of this program is proportional to the size of the tree.
5 OPTIMIZED TABULATION

\[ \text{con } F : \mathcal{P}(N) \]
\[ ; \quad T_a : N \]
\[ ; \quad T_a : N^n \rightarrow N \]
\[ ; \quad t : \text{Tree}(V, r) \]
\[ \text{var accepted : bool} \]
\[ : \text{func } \text{ms}_2 = \]
\[ (| t : \text{Tree}(V, r) | N \]
\[ | \text{if } t :: a \rightarrow T_a \]
\[ | \quad t :: a(t_1, \ldots, t_n) \rightarrow T_a(\text{ms}_2(t_1), \ldots, \text{ms}_2(t_n)) \]
\[ \text{fi} \]
\[ | \quad \text{accepted} := \text{ms}_2(t) \in F \]
\[ \] 

Example 4.9

Consider the grammar of example 2.6. The tables generated by algorithm A2 are:

<table>
<thead>
<tr>
<th>( E )</th>
<th>match set</th>
<th>( T_a )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>( T_b )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>{c; A, B}</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>{d; B}</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>\emptyset</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>{a(B, d); A, B}</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>{b(c), b(B); B}</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>{b(B); B}</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>{a(b(c), B); A, B}</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>{a(B, d), a(b(c), B); A, B}</td>
<td>7</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>7</td>
</tr>
</tbody>
</table>

\( T_c = 0, T_d = 1 ; F = \{0, 3, 6, 7\} \)

Table-driven bottom-up accepting of an input tree, for instance \( t = a(b(c), b(a(d, d))) \) proceeds as demonstrated in figure 3. Each node of \( t \) is now annotated with a number corresponding to the match set as it was annotated with in figure 2.

Since \( 6 \in F, t \in \mathcal{L}(G) \).

\[ \Box \]

5 Optimized tabulation

In practice, code generation tree grammars (see introduction) are rather extensive. This means that transition tables may be very large. Compression of these tables can be applied after com-
The optimization is based on an equivalence relation on match sets. The basic idea of the equivalence relation is the observation that some patterns only occur as j-th subtree of a tree labelled with a symbol of rank n (n ≥ j). Main advantage is that with generation of match sets one can iterate over the equivalence classes instead of the match sets. This is quite lucrative, provided that the mapping of match sets on equivalence classes is not (nearly) a bijection (then no improvement is made).

The idea of this optimization originated with David Chase, but he only gave an informal treatment of his ideas (see [Chase]). Here, we derive an improved algorithm for the generation of match sets based on the material presented in [Chase].

The j-th childset of a symbol, say a, 1 ≤ j ≤ r(a), is the set of patterns that appear as j-th subtree of a tree in PS labelled with a. Formally:

**Definition 5.1** \{ CS_{a,j}, j-th childset of symbol a \}

∀n : 1 ≤ n : ∀a ∈ V_n : ∀j : 1 ≤ j ≤ n : ∀t_1, ..., t_n ∈ Tree(N ∪ V, r) :

CS_{a,j} = \{ t ∈ Tree(N ∪ V, r) | ∃t' ∈ PS : t' :: a(t_1, ..., t_n) ∧ t_j = t \}

The childsets of the symbols of the grammar of example 2.6 are : CS_{a,1} = \{b(c), B\}, CS_{a,2} = \{B, d\}, and CS_{b,1} = \{c, B\}.

**Lemma 5.3** ∀n : 1 ≤ n : ∀a ∈ V_n : ∀j : 1 ≤ j ≤ n : CS_{a,j} ⊆ PS

**Proof** : PS is closed under taking subtrees. □
Using the childsets we may refine the definition of match set as follows.

**Definition 5.4** \{ **MS**₁, **match set** \}

- the function \( MS₁ \in \text{Tree}(V, r) \rightarrow \mathcal{P}(PS) \) is defined by :
  - \( \forall a \in V₀ : MS₁(a) = \text{closure}({a}) \)
  - \( \forall n : 1 \leq n : \forall a \in Vₙ : \forall t₁, \ldots, tₙ \in \text{Tree}(V, r) : MS₁(a(t₁, \ldots, tₙ)) = \text{closure}({a(p₁, \ldots, pₙ) \in PS | \forall i : 1 \leq i \leq n : pᵢ \in MS₁(tᵢ) \cap CSₐᵢ}) \)

The only difference with definition 3.4 is the intersection with \( CSₐᵢ \). This does not affect the value of \( MS₁(a(t₁, \ldots, tₙ)) \), because the only patterns missing from \( MS₁(tᵢ) \cap CSₐᵢ \) are those patterns that do not appear as the \( i \)-th subtree of a tree in \( PS \) labelled with \( a \).

For the same reasons as mentioned in section 4 a program computing \( MS₁ \) is easy implementable, but rather inefficient. Again, tabulation is possible. Observe that \( MS₁(a(t₁, \ldots, tₙ)) \) is now of the form \( fₐ(gₐ₁(MS₁(t₁)), \ldots, gₐₙ(MS₁(tₙ))) \) where \( fₐ \) is defined as before (see definition 4.1) and \( gₐᵢ \) is defined as follows.

**Definition 5.5** \{ \( gₐᵢ \), map function for \( j \)-th child of symbol \( a \) \}

- the function \( gₐᵢ \in \mathcal{P}(PS) \rightarrow \mathcal{P}(PS) \), is defined by :
  - \( \forall s \in \mathcal{P}(PS) : gₐᵢ(s) = s \cap CSₐᵢ \)

For practical reasons we use an enumeration \( E \in \mathcal{N} \rightarrow \mathcal{P}(PS) \) of match sets. The definition of transition tables is stated in terms of the map function and the enumeration \( E \) as follows.

**Definition 5.6** \{ \( Tₐ \), transition table for symbol \( a \) \}

- \( \forall a \in V₀ : Tₐ \in \text{dom}(E) \), where \( E(Tₐ) = fₐ \)
- \( \forall n : 1 \leq n : \forall a \in Vₙ : \forall s₁, \ldots, sₙ \in \text{dom}(E) : Tₐ \in \text{dom}(E) \rightarrow \text{dom}(E) \), where \( E(Tₐ(s₁, \ldots, sₙ)) = fₐ(gₐ₁(E(s₁)), \ldots, gₐₙ(E(sₙ))) \)

Apparently, the transition tables are similar with those defined in definition 4.2. An important observation is that intersection of a match set with some childset \( CSₐᵢ \), for some symbol \( a \in Vₙ, 1 \leq n \) and \( j : 1 \leq j \leq n \), induces an equivalence relation over match sets.
Definition 5.7 \{ equivalence relation \( \approx_{a,j} \) \}
\[ \forall s, s' \in \mathcal{P}(PS) : \forall n : 1 \leq n : \forall a \in V_n : \forall j : 1 \leq j \leq n : \; s \approx_{a,j} s' \Leftrightarrow s \cap CS_{a,j} = s' \cap CS_{a,j} \]
\[
\]

Definition 5.8 \{ equivalence class \( \varepsilon_{\approx_{a,j}} \) \}
\[ \forall s \in \mathcal{P}(PS) : \forall n : 1 \leq n : \forall a \in V_n : \forall j : 1 \leq j \leq n : \; \varepsilon_{\approx_{a,j}}(s) = \{ s' \in \mathcal{P}(PS) \mid s \approx_{a,j} s' \} \]
\[
\]

In other words: the equivalence class \( \varepsilon_{\approx_{a,j}}(s) \) is the set of all match sets that are equivalent (under \( \approx_{a,j} \)) to \( s \). An equivalence class \( \varepsilon_{\approx_{a,j}}(s) \) is represented by \( s \cap CS_{a,j} \), which is called the \textit{representer-set} of \( \varepsilon_{\approx_{a,j}}(s) \).

Tabulation of representer-sets is possible, since the number of equivalence classes is finite. We introduce for all \( n : 1 \leq n \), for all \( a \in V_n \), for all \( j : 1 \leq j \leq n \) an enumeration \( R_{a,j} \in \mathcal{N} \rightarrow \mathcal{P}(PS) \) of representer-sets. The mapping of (the enumeration of) match sets on (the enumeration of) representer-sets is performed by an index map table \( \mu_{a,j} \).

Definition 5.9 \{ \mu_{a,j}, index map table for \( j \)-th child of symbol \( a \) \}
\[ \forall n : 1 \leq n : \forall a \in V_n : \forall j : 1 \leq j \leq n : \]
\[ \bullet \quad \text{the function} \; \mu_{a,j} \in \text{dom}(E) \rightarrow \text{dom}(R_{a,j}) \; \text{is defined by} : \]
\[ \bullet \quad \forall s \in \text{dom}(E) : R_{a,j}(\mu_{a,j}(s)) = E(s) \cap CS_{a,j} \]
\[
\]

Notice that \( \mu_{a,j} \) is, in fact, nothing else than the tabulation of \( g_{a,j} \), defined above. This means that transitions have to be tabulated for representer-sets only (instead of match sets). This is reflected in the following definition.

Definition 5.10 \{ \( T'_a \), transition table for symbol \( a \) \}
\[ \bullet \; \forall a \in V_0 : T'_a \in \text{dom}(E), \; \text{where} \; E(T'_a) = f_a \]
\[ \bullet \; \forall n : 1 \leq n : \forall a \in V_n : \forall j : 1 \leq j \leq n : \forall s_1, \ldots, s_n \in \text{dom}(E) : \]
\[ \quad - \; \mu_{a,j} \in \text{dom}(E) \rightarrow \text{dom}(R_{a,j}) \]
\[ \quad - \; T'_a \in (\text{dom}(R_{a,1}) \times \ldots \times \text{dom}(R_{a,n})) \rightarrow \text{dom}(E) \]
\[ \quad - \; E(T'_a(\mu_{a,1}(s_1), \ldots, \mu_{a,n}(s_n))) = f_a(g_{a,1}(E(s_1)), \ldots, g_{a,n}(E(s_n))) \]
\[
\]

The corresponding definition of match set is changed into:

Definition 5.11 \{ \( MS_2 \), match set \}
\[ \bullet \; MS_2 \in \text{Tree}(V, r) \rightarrow \text{dom}(E) \]
5 OPTIMIZED TABULATION

- \( \forall a \in V_0 : M_{S_2}(a) = T'_a \)
- \( \forall n : 1 \leq n : \forall a \in V_n : \forall t_1, \ldots, t_n \in Tree(V, r) : M_{S_2}(a(t_1, \ldots, t_n)) = T'_a(\mu_{a,1}(M_{S_2}(t_1)), \ldots, \mu_{a,n}(M_{S_2}(t_n))) \)

Notice that \( M_{S_2} \) is still related to \( M_{S_1} \) by lemma 4.4.

Using the definitions above and the invariants given below we may derive a tabulation algorithm (named \( A_3 \)), which is presented here. First, we give the invariants of the program.

\[ P_1 : \ (\forall n : 1 \leq n : \forall a \in V_n : \text{new}_a = (\exists j : 1 \leq j \leq n : p_{a,j} < q_{a,j}) \]  

\[ P_2 : \ 0 \leq p \leq q \wedge (\forall n : 1 \leq n : \forall a \in V_n : (\forall j : 1 \leq j \leq n : 0 \leq p_{a,j} \leq q_{a,j}) \wedge (\forall i : 0 \leq i < q \wedge 0 \leq \mu_{a,j}(i) < q_{a,j} : R_{a,j}(\mu_{a,j}(i)) = E(i) \cap C_{S_{a,j}})) \wedge (\forall i_1, \ldots, i_n : :)  

\[ E(T'_a(\mu_{a,1}(i_1), \ldots, \mu_{a,n}(i_n))) = f_a(g_{a,1}(E(i_1)), \ldots, g_{a,n}(E(i_n)))) \]

\[ \text{con} \ G = (N, V, r, P, S) : \text{tree grammar} \]

\[ \text{var} \ E : N \rightarrow \mathcal{P}(PS) \]
\[ \quad F : \mathcal{P}(N) \]
\[ \quad T'_a : N 
\[ \quad T'_a : N^n \rightarrow N \]
\[ \quad \mu_{a,j} : N \rightarrow N \]
\[ \quad R_{a,j} : N \rightarrow \mathcal{P}(PS) \]
\[ \quad p_{a,j}, q_{a,j} : N \]
\[ \quad C_{S_{a,j}} : \mathcal{P}(PS) \]
\[ \quad \text{new}_a : \text{bool} \]
\[ \quad \text{closure} : N \rightarrow \mathcal{P}(N) \]

\[ \text{proc} \ \text{compute\_closure} \ (* \text{see algorithm A}_2 *) \]

\[ \text{func} \ \text{closure} \ (* \text{see algorithm A}_2 *) \]

\[ \text{proc} \ \text{compute\_childsets} = \]
\[ \quad \text{var} \ j, n : N \]
\[ \quad \text{for all} \ a \in V \setminus V_0 \]
do \( j, n := 1, r(a) \)

\( ; \) do \( j \neq n + 1 \rightarrow CS_{a,j} := \emptyset; j := j + 1 \) od

od

; for all \( a(t_1, \ldots, t_n) \in PS \)

do \( j := 1 \)

\( ; \) do \( j \neq n + 1 \rightarrow CS_{a,j} := CS_{a,j} \cup \{t_j\} \) od

od

; proc compute_reprsets =

\( \downarrow p : \mathcal{N} \)

\( \| \) (* compute representer-sets of matchset \( E(p) \) for all \( a \in V \setminus V_0 \) *)

(* and fill \( \mu_{a,j} \)-tables, \( 1 \leq j \leq r(a) \), for \( E(p) \) *)

for all \( a \in V \setminus V_0 \)

do \( \| \) var \( j, n : \mathcal{N} \)

\( \| \) \( j, n := 1, r(a) \)

; do \( j \neq n + 1 \rightarrow R_{a,j}(q_{a,j}) := E(p) \cap CS_{a,j} \)

\( \| \) var \( k : \mathcal{N} \)

\( \| \) \( k := 0 ; \) do \( R_{a,j}(k) \neq R_{a,j}(q_{a,j}) \rightarrow k := k + 1 \) od

\( \| \) if \( k \neq q_{a,j} \rightarrow \) skip

\( \| \) \( k = q_{a,j} \rightarrow (* \text{new representer-set} *) \)

\( \| \quad q_{a,j}, \text{new}_{a} := q_{a,j} + 1, \text{true} \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)

\( \| \)
The table-driven acceptor is described by algorithm $A_4$ which is presented below.

```plaintext
od
; \textcolor{red}{q := 0}; \textcolor{blue}{p := 0}
; \textcolor{red}{\text{for all}} \ a \ \in \ V_0
\textcolor{green}{\text{do}} \ E(q) := \text{closure}\{\textcolor{red}{a}\}\n; \textcolor{blue}{T_a',q := q,q + 1}
od
; \textcolor{red}{\text{do}} \ p \neq q \rightarrow \text{compute_reprsets}\textcolor{red}{(p)}\;\textcolor{blue}{p := p + 1} \od
; \textcolor{red}{\text{do}} \ (\exists a \in V \setminus V_0 : \textcolor{blue}{\text{new}_a})
    \rightarrow \textcolor{red}{\text{for all}} \ a \ \in \ V \setminus V_0 : \textcolor{blue}{\text{new}_a}
    \textcolor{green}{\text{do}} \ [\textcolor{red}{\text{var}} \ j,n : \mathcal{N}
        \textcolor{blue}{\mid n := r(a)}
        ; \textcolor{red}{\text{for all}} \ (p_1, \ldots, p_n) \in \{0..q_a,1-1\} \times \ldots \times \{0..q_a,n-1\}\n            \{0..p_a,1-1\} \times \ldots \times \{0..p_a,n-1\}
        \textcolor{green}{\text{do}} \ E(q) := \text{closure}\{(a(t_1, \ldots, t_n) \in PS \mid \forall i : 1 \leq i \leq n : t_i \in R_a(p_i))\}
        \textcolor{blue}{\mid [\textcolor{red}{\text{var}} \ k : \mathcal{N}}
            \textcolor{blue}{\mid k := 0}; \textcolor{green}{\text{do}} \ E(k) \neq E(q) \rightarrow k := k + 1 \od
            \textcolor{blue}{\mid \text{if}} k = q \rightarrow q := q + 1 \ (*\text{new match set}*)
            \textcolor{blue}{\mid \text{if}} k \neq q \rightarrow \text{skip}
        \text{fi}
        \textcolor{blue}{\mid T_a'(p_1, \ldots, p_n) := k}
\text{fi}
\textcolor{red}{\text{od}}
; j := 1; \textcolor{green}{\text{do}} \ j \neq n + 1 \rightarrow p_{a,j} := q_{a,j}; j := j + 1 \od
; \textcolor{blue}{\text{new}_a := false}
od
; \textcolor{red}{\text{do}} \ p \neq q \rightarrow \text{compute_reprsets}(p)\;\textcolor{blue}{p := p + 1} \od
od
; p := 0; \textcolor{red}{F := \emptyset \ (*\text{determine accepting states}*)}
; \textcolor{red}{\text{do}} \ p \neq q \rightarrow \text{if} \ S \in E(p) \rightarrow F := F \cup \{p\}
    \textcolor{blue}{\mid \text{if} S \notin E(p) \rightarrow \text{skip}
\text{fi}
\textcolor{red}{\text{od}}
; p := p + 1
\text{od}
```

The table-driven acceptor is described by algorithm $A_4$ which is presented below.
5  Optimized Tabulation

\[
\text{con } F : \mathcal{P}(N)
\]

\[
T'_a : N^a \rightarrow N
\]

\[
\mu_{a,j} : N \rightarrow N
\]

\[
t : \text{Tree}(V,r)
\]

var accepted : bool

::func \( m(s) = \)

\[
(\t : \text{Tree}(V,r) | N)
\]

\[
\begin{align*}
| \text{if } t :: a & \rightarrow T'_a \\
| t :: a(t_1, \ldots , t_n) & \rightarrow T'_a(\mu_{a,1}(m(t_1)), \ldots , \mu_{a,n}(m(t_n)))
\end{align*}
\]

fi

\[
\text{accepted} := m(s(t)) \in F
\]

\]

Compare the results given in the following example with those of example 4.9.

Example 5.12
Consider again our running example and consider the childsets of our grammar as given in example 5.2. The tables generated by algorithm A3 are:

<table>
<thead>
<tr>
<th>( E )</th>
<th>match set</th>
<th>( \mu_{a,1} )</th>
<th>( \mu_{a,2} )</th>
<th>( \mu_{b,1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>{c; A, B}</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>{d; B}</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>\emptyset</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>{a(B, d); A, B}</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>{b(c), b(B); B}</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>{b(B); B}</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>{a(b(c), B); A, B}</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>{a(B, d), a(b(c), B); A, B}</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

\[
T'_a = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 2 & 3 & 2 \\ 1 & 2 & 2 & 2 \\ 2 & 6 & 7 & 2 \\ \end{bmatrix}
\]

\[
T'_b = \begin{bmatrix} 0 & 4 \\ 1 & 5 \\ 2 & 2 \\ \end{bmatrix}
\]

\[
R_{a,1} = \begin{bmatrix} 0 & \{B\} \end{bmatrix} \\
R_{a,2} = \begin{bmatrix} 0 & \{B\} \\ 1 & \{d, B\} \end{bmatrix} \\
R_{b,1} = \begin{bmatrix} 0 & \{c, B\} \\ 1 & \{B\} \\ 2 & \emptyset \end{bmatrix}
\]

\( T'_c = 0, T'_d = 1; F = \{0, 3, 6, 7\} \)

For example, take \( E(4) \), which equals \( \{b(c), b(B), B\} \).

\( E(4) \cap C_{S_{a,1}} = E(4) \cap \{b(c), B\} = \{b(c), B\} = R_{a,1}(2) \), so \( \mu_{a,1}(4) = 2 \)
6 CONCLUDING REMARKS

\[ E(4) \cap CS_{a,2} = E(4) \cap \{B, d\} = \{B\} = R_{a,2}(0), \text{ so } \mu_{a,2}(4) = 0 \]
\[ E(4) \cap CS_{b,1} = E(4) \cap \{c, B\} = \{B\} = R_{b,1}(1), \text{ so } \mu_{b,1}(4) = 1 \]

Bottom-up accepting of \( t = a(b(c), b(a(d, d))) \) proceeds now as depicted in figure 4.

\[
\begin{align*}
T_a'(\mu_{a,1}(0)) &= 4 && T_a'(\mu_{a,2}(5)) = 6 \\
T_a'(\mu_{a,1}(1)) &= 3 && T_a'(\mu_{a,1}(3)) = 5
\end{align*}
\]

Figure 4: Example of matching using compressed tables

Since \( 6 \in F, t \in \mathcal{L}(G) \).

\[ \Box \]

6 Concluding remarks

We derived a rather efficient tabulation algorithm for table-driven bottom-up tree acceptors. The base idea for the optimized tabulation is quite simple, nevertheless it leads to a complex algorithm. The presented algorithms are derived by step-wise refinement: starting with the well-known reachability algorithm we elaborate this towards an algorithm for the efficient generation of compressed parse tables. Our opinion is that such a systematical derivation gives us more insight in a complex algorithm.

Experiments with an implementation (in Pascal) of the algorithm have demonstrated a considerable improvement in table generation. For example, a code generation grammar with 33 production rules (representing a part of the Intel 8085 instruction set) gave an improvement of 9208 table entries to only 635, of which 468 index map table entries. Index map tables take up most of the space, but traditional compression techniques can be used to reduce that space since index maps are inherently sparse.

Although the tabulation algorithm has an exponential time complexity we believe that for code generation grammars the mapping of match sets on equivalence classes is such that a passable improvement is made. Though much work remains to be done in the field of universal code generator-generators, we hope to have made a contribution to bottom-up parsing algorithms.
Acknowledgements
Thanks are due to Huub ten Eikelder, Berry Schoenmakers and Pieter Struik (all of the Eindhoven University of Technology) for reading and commenting on draft versions of this paper.

References


REFERENCES


In this series appeared:

<table>
<thead>
<tr>
<th>No.</th>
<th>Author(s)</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>85/01</td>
<td>R.H. Mak</td>
<td>The formal specification and derivation of CMOS-circuits</td>
</tr>
<tr>
<td>85/02</td>
<td>W.M.C.J. van Overveld</td>
<td>On arithmetic operations with M-out-of-N-codes</td>
</tr>
<tr>
<td>85/03</td>
<td>W.J.M. Lemmens</td>
<td>Use of a computer for evaluation of flow films</td>
</tr>
<tr>
<td>85/04</td>
<td>T. Verheoff</td>
<td>Delay insensitive directed trace structures satisfy the foam rubber wrapper postulate</td>
</tr>
<tr>
<td>86/01</td>
<td>R. Koymans</td>
<td>Specifying message passing and real-time systems</td>
</tr>
<tr>
<td>86/02</td>
<td>G.A. Bussing, K.M. van Hee, M. Voorhoeve</td>
<td>ELISA, A language for formal specifications of information systems</td>
</tr>
<tr>
<td>86/03</td>
<td>Rob Hoogerwoord</td>
<td>Some reflections on the implementation of trace structures</td>
</tr>
<tr>
<td>86/04</td>
<td>G.J. Houben, J. Paredaens, K.M. van Hee</td>
<td>The partition of an information system in several parallel systems</td>
</tr>
<tr>
<td>86/05</td>
<td>Jan L.G. Dietz, Kees M. van Hee</td>
<td>A framework for the conceptual modeling of discrete dynamic systems</td>
</tr>
<tr>
<td>86/06</td>
<td>Tom Verheoff</td>
<td>Nondeterminism and divergence created by concealment in CSP</td>
</tr>
<tr>
<td>86/07</td>
<td>R. Gerth, L. Shira</td>
<td>On proving communication closedness of distributed layers</td>
</tr>
<tr>
<td>86/09</td>
<td>C. Huizing, R. Gerth, W.P. de Roever</td>
<td>Full abstraction of a real-time denotational semantics for an OCCAM-like language</td>
</tr>
<tr>
<td>86/10</td>
<td>J. Hooman</td>
<td>A compositional proof theory for real-time distributed message passing</td>
</tr>
<tr>
<td>86/11</td>
<td>W.P. de Roever</td>
<td>Questions to Robin Milner - A responder's commentary (IFIP86)</td>
</tr>
<tr>
<td>86/12</td>
<td>A. Boucher, R. Gerth</td>
<td>A timed failures model for extended communicating processes</td>
</tr>
<tr>
<td>Year</td>
<td>Author(s)</td>
<td>Title</td>
</tr>
<tr>
<td>------</td>
<td>-----------</td>
<td>-------</td>
</tr>
<tr>
<td>86/14</td>
<td>R. Koymans</td>
<td>Specifying passing systems requires extending temporal logic</td>
</tr>
<tr>
<td>87/01</td>
<td>R. Gerth</td>
<td>On the existence of sound and complete axiomatizations of the monitor concept</td>
</tr>
<tr>
<td>87/02</td>
<td>Simon J. Klaver, Chris F.M. Verberne</td>
<td>Federatieve Databases</td>
</tr>
<tr>
<td>87/03</td>
<td>G.J. Houben, J.Paredaens</td>
<td>A formal approach to distributed information systems</td>
</tr>
<tr>
<td>87/04</td>
<td>T. Verhoeff</td>
<td>Delay-insensitive codes - An overview</td>
</tr>
<tr>
<td>87/05</td>
<td>R. Kuiper</td>
<td>Enforcing non-determinism via linear time temporal logic specification.</td>
</tr>
<tr>
<td>87/06</td>
<td>R. Koymans</td>
<td>Temporele logica specificatie van message passing en real-time systemen (in Dutch).</td>
</tr>
<tr>
<td>87/07</td>
<td>R. Koymans</td>
<td>Specifying message passing and real-time systems with real-time temporal logic.</td>
</tr>
<tr>
<td>87/08</td>
<td>H.M.J.L. Schols</td>
<td>The maximum number of states after projection.</td>
</tr>
<tr>
<td>87/10</td>
<td>T. Verhoeff</td>
<td>Three families of maximally nondeterministic automata.</td>
</tr>
<tr>
<td>87/11</td>
<td>P. Lemmens</td>
<td>Eldorado ins and outs. Specifications of a data base management toolkit according to the functional model.</td>
</tr>
<tr>
<td>87/12</td>
<td>K.M. van Hee, A. Lapinski</td>
<td>OR and AI approaches to decision support systems.</td>
</tr>
<tr>
<td>87/13</td>
<td>J.C.S.P. van der Woude</td>
<td>Playing with patterns, searching for strings.</td>
</tr>
<tr>
<td>87/14</td>
<td>J. Hooman</td>
<td>A compositional proof system for an occam-like real-time language</td>
</tr>
<tr>
<td>Volume</td>
<td>Author(s)</td>
<td>Title</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>-------</td>
</tr>
<tr>
<td>87/15</td>
<td>C. Huizing, R. Gerth, W.P. de Roever</td>
<td>A compositional semantics for statecharts</td>
</tr>
<tr>
<td>87/16</td>
<td>H.M.M. ten Eikelder, J.C.F. Wilmont</td>
<td>Normal forms for a class of formulas</td>
</tr>
<tr>
<td>87/17</td>
<td>K.M. van Hee, G.-J. Houben, J.L.G. Dietz</td>
<td>Modelling of discrete dynamic systems framework and examples</td>
</tr>
<tr>
<td>87/18</td>
<td>C.W.A.M. van Overveld</td>
<td>An integer algorithm for rendering curved surfaces</td>
</tr>
<tr>
<td>87/19</td>
<td>A.J. Seebregts</td>
<td>Optimalisering van file allocatie in gedistribueerde database systemen</td>
</tr>
<tr>
<td>87/20</td>
<td>G.J. Houben, J. Paredaens</td>
<td>The $R^2$-Algebra: An extension of an algebra for nested relations</td>
</tr>
<tr>
<td>87/21</td>
<td>R. Gerth, M. Codish, Y. Lichtenstein, E. Shapiro</td>
<td>Fully abstract denotational semantics for concurrent PROLOG</td>
</tr>
<tr>
<td>88/01</td>
<td>T. Verhoeff</td>
<td>A Parallel Program That Generates the Möbius Sequence</td>
</tr>
<tr>
<td>88/02</td>
<td>K.M. van Hee, G.J. Houben, L.J. Somers, M. Voorhoeve</td>
<td>Executable Specification for Information Systems</td>
</tr>
<tr>
<td>88/03</td>
<td>T. Verhoeff</td>
<td>Settling a Question about Pythagorean Triples</td>
</tr>
<tr>
<td>88/04</td>
<td>G.J. Houben, J. Paredaens, D. Tahon</td>
<td>The Nested Relational Algebra: A Tool to handle Structured Information</td>
</tr>
<tr>
<td>88/05</td>
<td>K.M. van Hee, G.J. Houben, L.J. Somers, M. Voorhoeve</td>
<td>Executable Specifications for Information Systems</td>
</tr>
<tr>
<td>88/06</td>
<td>H.M.I.L. Schols</td>
<td>Notes on Delay-Insensitive Communication</td>
</tr>
<tr>
<td>88/07</td>
<td>C. Huizing, R. Gerth, W.P. de Roever</td>
<td>Modelling Statecharts behaviour in a fully abstract way</td>
</tr>
<tr>
<td>88/08</td>
<td>K.M. van Hee, G.J. Houben, L.J. Somers, M. Voorhoeve</td>
<td>A Formal model for System Specification</td>
</tr>
<tr>
<td>88/09</td>
<td>A.T.M. Aerts, K.M. van Hee</td>
<td>A Tutorial for Data Modelling</td>
</tr>
<tr>
<td>Reference</td>
<td>Authors</td>
<td>Title</td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>88/10</td>
<td>J.C. Ebergen</td>
<td>A Formal Approach to Designing Delay Insensitive Circuits</td>
</tr>
<tr>
<td>88/11</td>
<td>G.J. Houben, J. Paredaens</td>
<td>A graphical interface formalism: specifying nested relational databases</td>
</tr>
<tr>
<td>88/12</td>
<td>A.E. Eiben</td>
<td>Abstract theory of planning</td>
</tr>
<tr>
<td>88/13</td>
<td>A. Bijlsma</td>
<td>A unified approach to sequences, bags, and trees</td>
</tr>
<tr>
<td>88/14</td>
<td>H.M.M. ten Eikelder, R.H. Mak</td>
<td>Language theory of a lambda-calculus with recursive types</td>
</tr>
<tr>
<td>88/15</td>
<td>R. Bos, C. Hemerik</td>
<td>An introduction to the category theoretic solution of recursive domain equations</td>
</tr>
<tr>
<td>88/16</td>
<td>C. Hemerik, J.P. Katoen</td>
<td>Bottom-up tree acceptors</td>
</tr>
<tr>
<td>88/17</td>
<td>K.M. van Hee, G.J. Houben, L.J. Somers, M. Voorhoeve</td>
<td>Executable specifications for discrete event systems</td>
</tr>
</tbody>
</table>