CYK Derivation

by Fares Hedayati

Let $w = a_1a_2 \cdots a_n$ and let $G = (V, T, S, P)$. CYK algorithm takes the string w and the grammar G as input and says that if G can generate w or not. The algorithm is as follows:

Let $w_{i,j} = w_iw_{i+1} \cdots w_j$ and let $X_{i,j}$ be the set of variables $A \in V$ such that when you start from A after applying some production rules from P, the substring $w_{i,j}$ can be achieved. This means that $X_{i,j} = \{A \in V : A \Rightarrow^* w_{i,j}\}$. Obviously w is generated by G if and only if S is a member of $X_{1,n}$ because $X_{1,n}$ is the set of variables that when you start from any of them, then you can derive the whole string (note that $w_{1,n} = w$). Now if S is in $X_{1,n}$ then by starting from S the whole string can be derived.

Now note $X_{i,i}$ is the set of variables that can derive $w_{i,i}$ (note that $w_{i,i}$ is a_i). Hence $X_{i,i}$ is the set of variables from V like A that have production rules of the form $A \to a_i$ in P. Furthermore $X_{i,j}$ can be built recursively by finding all A in V such that there is a production $A \to BC$ with $B \in X_{i,k}$ and $C \in X_{k+1,j}$ for some k. The CYK algorithm is as follows:

```
Let X_{i,i} = {} for 1i ≤ n (initially they are all empty)
For i = 1 to n
    For A ∈ V
        If A → a_i ∈ P
            Add A to X_{i,i}

For s = 1 to n − 1
    For i = 1 to n − s
        X_{i,i+s} = {}
        For k = i to i + s − 1
            For B ∈ X_{i,k}
                For C ∈ X_{k+1,i+s}
                    For A ∈ V
                        If A → BC ∈ P
                            Add A to X_{i,i+s}
```

If $S \in X_{1,n}$ output *yes* otherwise output *no*.

Now we change the CYK algorithm in the following way to find the derivation :

We introduce a new data structure $Y_{i,j}$ for each $X_{i,j}$. Each member of $Y_{i,j}$ is a set of four elements, something of the form $[A, k, B, C]$ where A and B and C are variables from V and k is a number between i and $j-1$. If $[A, k, B, C] \in Y_{i,j}$, this means that $A \to BC \in P$ and $B \in X_{i,k}$ and $C \in X_{k+1,j}$

Let $X_{i,i} = \{\}$ for $1i \leq n$ (initially they are all empty)
For $i = 1$ to n
    For $A \in V$
        If $A \to a_i \in P$
            Add A to $X_{i,i}$ and Add $[A, null, null, null]$ to $Y_{i,i}$

For $s = 1$ to $n-1$
    For $i = 1$ to $n-s$
        $X_{i,i+s} = \{\}$ and $Y_{i,i+s} = \{\}$
        For $k = i$ to $i+s-1$
            For $B \in X_{i,k}$
                For $C \in X_{k+1,i+s}$
                    For $A \in V$
                        If $A \to BC \in P$
                            Add $[A, k, B, C]$ to $Y_{i,i+s}$ and Add A to $X_{i,i+s}$
If $S \in X_{1,n}$
    Output *yes*
    Find a set in $Y_{1,n}$ such that S is its first element,
    let this set be $[S, k, R, T]$ call Derive($[S, k, R, T]$, 1 , n).
Else
    Output *no*

Derive($[A, k, B, C]$, i, j)
    If $i == j$
        Print $A \to a_i$
    Else
        Print $A \to BC$

Call Derive(Find(B, i, k) , i, k)
Call Derive(Find(C, k+1, j) , $k+1$, j)

Find(A,i,j)
Return the set in $Y_{i,j}$ that starts with A, let this set be $[A, k, B, C]$.
If such a set does not exist in $Y_{i,j}$ return $\{\}$