

Assignment 1

180001043

Roorpaj B S

Aim:-

Write a parallel program to parse a string of symbols. The inputs are a context-free grammar G in Chomsky Normal Form and a string of symbols.

Given:-

We are given a context free Grammar $G = (V, X, R, S)$ and a string w , where:

- 1) V is a finite set of variables or non-terminal symbols,
- 2) X is a finite set of terminal symbols,
- 3) R is a finite set of rules,
- 4) S is the start symbol, a distinct element of V , and
- 5) V and X are assumed to be disjoint sets.

A Context Free Grammar G is in Chomsky Normal Form (CNF) if each rule of G is of the form:

- $A \rightarrow BC$, (with at most two non-terminal symbols on RHS)
- $A \rightarrow a$, (one terminal symbol on the RHS)
- $S \rightarrow \text{null string}$, (NULL string)

Idea:-

Here to solve this problem I have parallelized the CYK algorithm. Basically there are four loops in the code, and I have parallelized the outermost and next to outermost loop. The Entire code is in openmp, and by just doing this step i am getting a significant time difference.

Algorithm:-

CYK algorithm is a parsing algorithm for context free grammar. In order to apply the CYK algorithm to a grammar, it must be in Chomsky Normal Form. It uses a dynamic programming algorithm to tell whether a string is in the language of a grammar. It is used to solve the membership problem using a dynamic programming approach. The algorithm is based on the principle that the solution to problem $[i,j]$ can be constructed from solution to subproblem $[i,k]$ and the solution to sub problem $[k,j]$. The algorithm requires the Grammar G to be in Chomsky Normal Form(CNF). Note that any Context-Free Grammar can be systematically converted to CNF. This restriction is employed so that each problem can be divided into two subproblems and not more to bound the time complexity.

The Implementation is done in the code file.

Example:

ALPHABET:

a, b;

VARIABLES:

S, A, B;

PRODUCTIONS:

$S \rightarrow AB$

$A \rightarrow BC$

$A \rightarrow BA$

$A \rightarrow a$

$B \rightarrow CC|b$

$C \rightarrow AB|a$

The string to be checked:

Baaba

The Output are:

```
Assignment1 — Roopraj — -bash — 132x39
ROOPRAJs-MacBook-Pro:Assignment1 rooprajbs$ g++ -Xpreprocessor -fopenmp -lomp assig1.c
clang: warning: treating 'c' input as 'c++' when in C++ mode, this behavior is deprecated [-Wdeprecated]
ROOPRAJs-MacBook-Pro:Assignment1 rooprajbs$ ./a.out
Part of language
ROOPRAJs-MacBook-Pro:Assignment1 rooprajbs$
```

```
/* Checks if last element of first array contains a string variable
of length 10 (str.length() = 10, find(str) == str.length() = 10, length()
    and is "Part of language")
else
    and is "Not part of language")
}

//Program passed
return 0;
```