AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA W KRAKOWIE

Wybrane metody automatycznej generacji kodu na procesory wielordzeniowe

Student realizujący: Piotr Listkiewicz
Opiekun pracy: dr inż. Marcin Pietroń

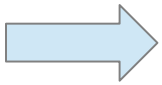
Cel pracy

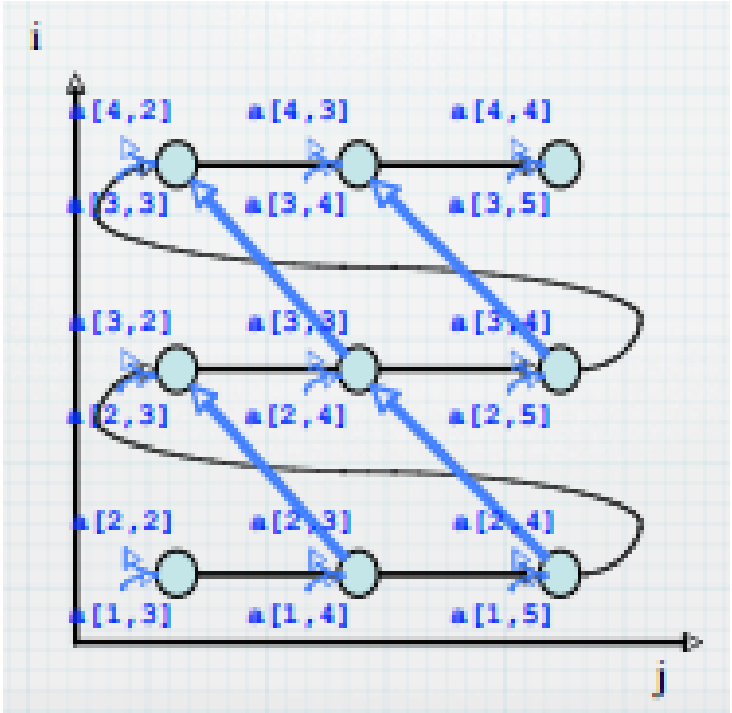
- **Analiza metod automatycznej detekcji kodu który można zrównoleglić**
- **Wytworzenie narzędzia umożliwiającego automatyczną generację kodu w wersji równoległej**
- **Obecnym celem jest implementacja narzędzia dla języka C , które przekształca go do postaci OpenMP**
- **Sprawdzenie działania narzędzia na algorytmach Machine Learningu**

Testowanie zależności

- **Wykrywanie zależności występujących w pętli pomiędzy iteracjami i instrukcjami**
- **Jest kilka możliwych rodzajów zależności pomiędzy instrukcjami**
- **Na ich podstawie buduje się graf zależności**

```
  for i = 2 to 4
    for j = 2 to 4
1:    a[i,j] = a[i-1,j+1];
```

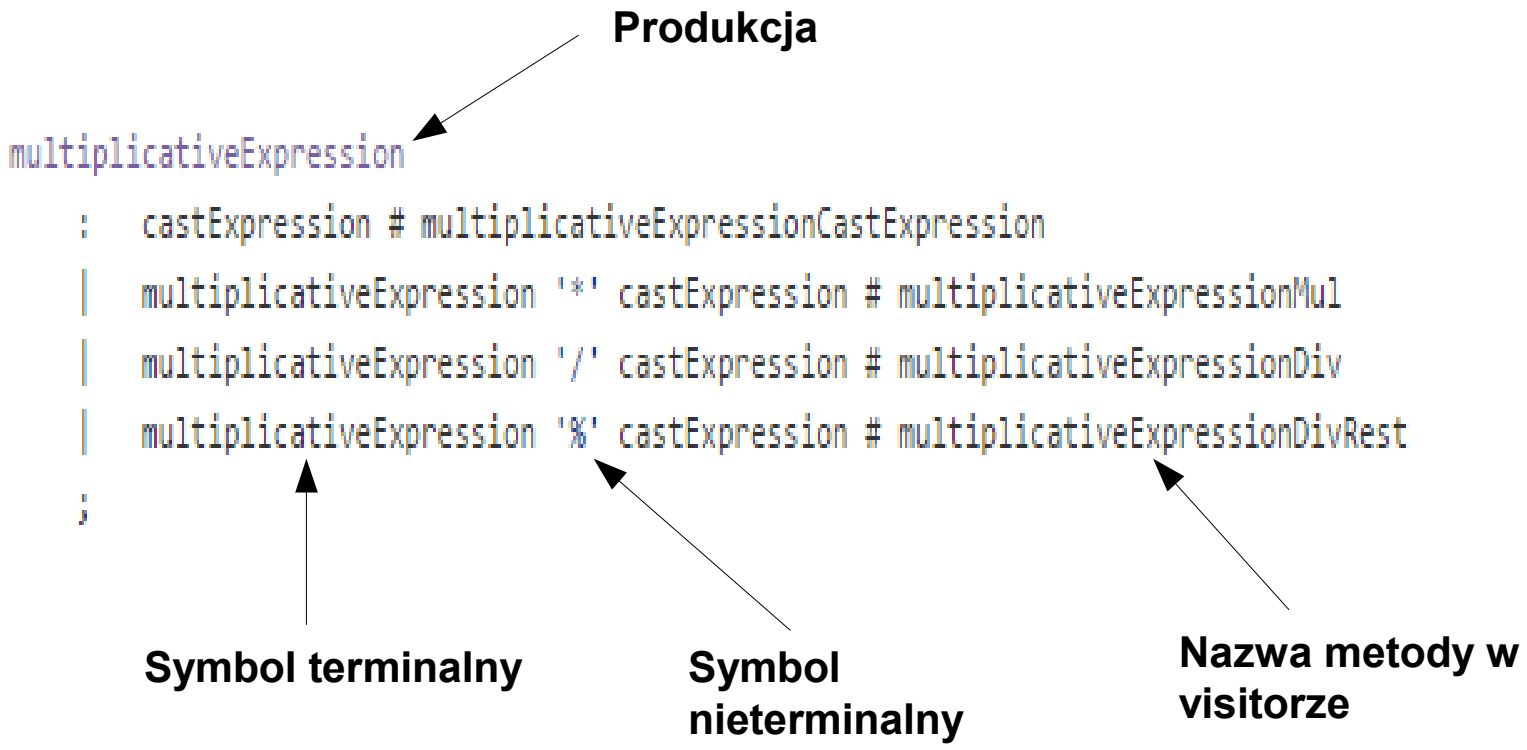




Gramatyka języka C

- Według specyfikacji C11
- Jedna z oficjalnych gramatyka udostępnianych przez antlr4 na licencji BSD na ich stronie na githubie
- Lekko poprawiona z powodu niewielkich bugów
- Co zrobić z dyrektywami? Najlepiej założyć ,że plik został obrany obróbce przez preprocesor
- Plik z gramatyką ma ponad 900 linii kodu

Gramatyka języka C – przykład produkcji



Przekształcenie drzewa rozbioru na AST

- Drzewo rozbioru oparte o gramatykę, AST(Abstract Syntax Tree) własne drzewo udostępniające odpowiednie informację
- Wykorzystuje wzorzec visitor z antlr4, polega na nadpisaniu odpowiednich funkcji z utworzonego przez antlr4 domyślnego visitora
- Antlr4 jest zintegrowany z mavenem – przy każdej budowie projektu na podstawie gramatyki są tworzone odpowiednie parsery,visitory itp.
-

Visitor – przykład metody

```
@Override
public DetailAst visitMultiplicativeExpressionDivRest(C11Parser.MultiplicativeExpressionDivRestContext ctx) {
    List<DetailAst> childrens = new LinkedList<DetailAst>();
    childrens.add(visit(ctx.multiplicativeExpression()));
    childrens.add(visit(ctx.castExpression()));
    return new DivRestExpressionsAst(getText(ctx), childrens);
}
```

Postać AST

- Bazą jest klasa DetailAST w której są trzymane informację o dzieciach danego węzła drzewa
- Klasa DetailAST umożliwia przechodzenie po węźle dla wszystkich klas dziedziczących z niej
- Poszczególne instrukcje/wyrażenia z C mają postać klas dziedziczących po DetailAST takich jak BreakStatementAst, GotoStatementAst, WhileStatementAst, itd.
- Istnieją także klasy pośrednie grupujące pewne funkcjonalności dla specyficznych typów węzłów jak przypisania, statementy itp.
- Istnieje kilkadziesiąt klas dziedziczących po DetailAst

Testowanie

- Wymagana duża poprawność rozwiązania – odczytanie wywołania funkcji jako deklaracji zmiennej może totalnie wypaczyć działanie algorytmów testowania zależności
- Testy jednostkowe – w tym momencie około 200 testów
- Całość wytwarzana przy pomocy TDD
- Testy integracyjne – bierzemy kody w C , parsujemy do AST, potem z AST przekształcamy do postaci tekstu i porównujemy z plikiem wejściowym
- W tym momencie znajduje się 10 plików, ponad 7000 linii kodu w C

AST Api

- Aby AST było łatwe w obsłudze wprowadzono dodatkowe funkcje ułatwiające dokonanie operacji na drzewie
- Przekształcenie węzła do postaci tekstu
- Przeszukiwanie drzewa: uzyskanie dzieci danego węzła, uzyskanie wszystkich potomków danego węzła, wyszukiwnie wszystkich potomków będących pętlami, wskaźnikami, odwołaniami tablicowymi itp.
- Uzyskanie zależności wejściowych i wyjściowych danego węzła: zmienne które czyta i zmienna do których piszę
- Uzyskanie informacji czy dany węzeł jest pętlą, odwołaniem tablicowym, instrukcją złożoną itp

Obliczenie Data Flow Graph

- **Węzły reprezentują instrukcję**
- **Krawędź reprezentuje zależność pomiędzy instrukcjami: kierunek krawędzi reprezentuje która instrukcja następuje po której, typ krawędzi reprezentuje jakiego rodzaju jest zależność**
- **Rodzaje zależności : Read-After-Read, Read-After-Write, Write-After-Read, Write-After-Write**
- **Krawędź występuje tylko do najbliższej instrukcji ze względu na daną zmienną, w stosunku do instrukcji późniejszych można odczytać zależność „przechodnio” np.:**
  - **1 -RAW-> 2 -RAR-> 3 => 1 -RAW-> 3**

Obliczenie Data Flow Graph - cd

- Zaimplementowany algorytm generujący
- Generacja DFG jest w dalszym ciągu testowana w szczególności przypadki takie jak pętle, wyrażenia tablicowe
- Testowanie przy pomocy unit testów, w trakcie implementacji jest test automatycznie generujący graf z plików wejściowych

Następne kroki

- **Posiadając DFG będzie można podpiąć istniejące algorytmy testowania zależności po dostosowaniu ich do API**
- **Z wyników algorytmów testowania zależności i DFG będzie następowała generacja wstawek w OpenMP**
- **Testowanie rozwiązania na algorytmach machine learningu na maszynach wieloprocesorowych**

Pytania?