

Zadanie č. 3

Vo zvolenom (podľa seba) programovacom jazyku naprogramujte program, ktorý realizuje redukciu bezkontextovej gramatiky zadanej v Backus-Naurovej forme. Váš program bude pracovať nasledovným spôsobom:

1. Prvým argumentom programu pri spustení bude cesta k textovému súboru, v ktorom je uložená bezkontextová gramatika zapísaná v Backus-Naurovej forme. Popis Backus-Naurovej formy je uvedený v neskoršej časti zadania, nájdete ho aj tu:
https://en.wikipedia.org/wiki/Backus%E2%80%93Naur_form.
2. Druhým argumentom programu pri spustení bude cesta k textovému súboru, do ktorého bude uložená výsledná redukovaná bezkontextová gramatika, zapísaná tiež v Backus-Naurovej forme.

Vstupy a výstupy

Prvý argument programu pri spustení je textový súbor so vstupnou gramatikou zapísanou v Backus-Naurovej forme. Bližší popis tejto formy je uvedený neskôr v tomto dokumente. Váš program vezme túto gramatiku a vykoná na nej redukciu, t.j. odstránenie nadbytočných a nedostupných symbolov.

Druhý argument programu pri spustení je textový súbor, do ktorého sa zapíše výsledná redukovaná gramatika, ekvivalentná ku načítanej gramatike z prvého argumentu. V prípade, že vstupná gramatika generuje prázdny jazyk, do súboru sa vloží len textová informácia PRAZDNY JAZYK.

Ak by ste teda napríklad naprogramovali Váš program v jazyku Python a Váš Python-kód by bol v súbore `zadanie3.py`, meno súboru so vstupnou gramatikou by bolo `vstup.txt` a meno súboru, kam sa uloží redukovaná gramatika `vystup.txt`, program by sa spúšťal:

```
python zadanie3.py vstup.txt vystup.txt
```

Zadanie môžete vypracovať v ľubovoľnom programovacom jazyku, Python je tu uvedený len ako príklad.

Implementácia algoritmov

Požaduje sa vaša samostatná práca, t.j. výsledný program musí byť výsledkom vašej samostatnej práce. Kopírovanie zdrojových kódov z internetu, prípadne od iných študentov bude hodnotené ako plagiátorstvo a v zmysle platného študijného poriadku hodnotené známku FX. Taktiež použité algoritmy a štruktúry musia byť výsledkom vašej vlastnej implementácie, t.j. VY si musíte naprogramovať vlastnú reprezentáciu gramatiky a odstránenia nadbytočných a nedostupných symbolov.

Deadline zadania

Zdrojové kódy Vášho riešenia odovzdajte do AIS-u do príslušného miesta odovzdania. Deadline je **21. apríl 2024, 23:59:59**.

Na ďalšej strane nájdete popis formátu vstupného/výstupného súboru, resp. Backus-Naurovej formy.

FORMÁT SÚBORU S BNF POPISOM GRAMATIKY

Backus-Naurova forma (BNF) je len iná forma zápisu bezkontextovej gramatiky. BNF, resp. jej ďalšie rozšírenia ako EBNF a ABNF (ktoré však nie sú súčasťou tohto zadania), sa spravidla používa pri popise syntaxe programovacích jazykov. Budeme uvažovať BNF podľa popisu na webstránke:

https://en.wikipedia.org/wiki/Backus%E2%80%93Naur_form

- Každý neterminál je ohraničený symbolmi $\langle \rangle$, napr. $\langle \text{program} \rangle$ je neterminálny symbol $\langle \text{program} \rangle$.
- Každý terminálny symbol je ohraničený **úvodzovkami** " ", napr. "a" je terminálny symbol **a** alebo "begin" je terminálny symbol **begin**. Oproti uvedenému linku ide o **zjednodušenie**, popis na uvedenom linku pripúšťa, že terminál je ohraničený aj apostrofmi ' ' (takúto situáciu nemusíte uvažovať).
- Meno neterminálu, resp. terminál môže obsahovať viacero znakov. Pre jednoduchosť môžete uvažovať, **na rozdiel od popisu na uvedenom linku**, že mená neterminálov a terminály budú tvorené len malými/veľkými písmenami anglickej abecedy a číslicami. Meno neterminálu **musí začínať** písmenom.
- Symbol \rightarrow , štandardne používaný v klasickom popise gramatík, oddeľujúci ľavú a pravú časť, je v BNF nahradený reťazcom $:=$
- Počiatočný neterminál je ten, ktorý je uvedený ako ľavá strana pravidla z **prvého riadku popisu gramatiky**.
- Ak je v pravidle pre jeden neterminál uvedených viacero pravých strán, sú oddelené symbolom |.
- **Jeden a ten istý neterminál** môže byť uvedený ako **ľavá strana na viacerých riadkoch**. V takom prípade je potrebné samozrejme brať do úvahy **všetky pravé strany na všetkých príslušných riadkoch**.
- Prázdny reťazec ε je na pravej strane reprezentovaný 2 po sebe idúcimi úvodzovkami, "".
- Akékoľvek medzery v popise gramatiky navyše (pred/za terminálmi, neterminálmi, symbolmi |, "", $:=$) je potrebné pri spracovaní gramatiky ignorovať.
- Každý riadok vstupného súboru s pravidlom je ukončený symbolom nového riadku.

FORMÁT SÚBORU S BNF POPISOM GRAMATIKY - alternatívny popis

Alternatívne je možné popísať syntax vstupného súboru pomocou BNF nasledovne (pozrite sekcia **Further examples** na linku):

https://en.wikipedia.org/wiki/Backus%E2%80%93Naur_form

```
<syntax> ::= <rule> | <rule> <syntax>
<rule> ::= <opt-whitespace> "<" <rule-name> ">" <opt-whitespace> "::="
<opt-whitespace> <expression> <line-end>
<opt-whitespace> ::= " " <opt-whitespace> | ""
<expression> ::= <list> | <list> <opt-whitespace> "|" <opt-whitespace>
<expression>
<line-end> ::= <opt-whitespace> <EOL> | <line-end> <line-end>
<list> ::= <term> | <term> <opt-whitespace> <list>
<term> ::= <literal> | "<" <rule-name> ">"
<literal> ::= "'" <text1> "'" | "" <text2> ""
<text1> ::= "" | <character1> <text1>
<text2> ::= "" | <character2> <text2>
<character> ::= <letter> | <digit> | <symbol>
<letter> ::= "A" | "B" | "C" | "D" | "E" | "F" | "G" | "H" | "I" | "J" | "K"
| "L" | "M" | "N" | "O" | "P" | "Q" | "R" | "S" | "T" | "U" | "V" | "W" | "X" | "Y" |
"Z" | "a" | "b" | "c" | "d" | "e" | "f" | "g" | "h" | "i" | "j" | "k" | "l" | "m" |
"n" | "o" | "p" | "q" | "r" | "s" | "t" | "u" | "v" | "w" | "x" | "y" | "z"
<digit> ::= "0" | "1" | "2" | "3" | "4" | "5" | "6" | "7" | "8" | "9"
<symbol> ::= "|" | " " | "!" | "#" | "$" | "%" | "&" | "(" | ")" | "*" | "+"
| "," | "-" | "." | "/" | ":" | ";" | ">" | "=" | "<" | "?" | "@" | "[" | "]" |
"^" | "_" | "`" | "{" | "}" | "~"
<character1> ::= <character> | ""
<character2> ::= <character> | ""
<rule-name> ::= <letter> | <rule-name> <rule-char>
<rule-char> ::= <letter> | <digit> | "-"
```

Obr. 1: Prevzaté z https://en.wikipedia.org/wiki/Backus–Naur_form

Zjednodušenia oproti tomuto popisu platné pre zadanie a iné upozornenia:

- (zjednodušenie) Mená neterminálov a terminály obsahujú len veľké/malé písmená a číslice.
- (zjednodušenie) Terminál je ohraničený úvodzovkami (popis na obrázku pripúšťa aj apostrofy).
- (upozornenie) Dve úvodzovky bez symbolov medzi nimi "" znamenajú prázdny reťazec ε .
- (upozornenie) <EOL> je koniec riadku podľa toho, ako ho chápe Váš operačný systém.

Ukážka vstupu/výstupu

Ak by súbor so vstupnou gramatikou obsahoval:

```
<S> ::= <A>"a" | <B>"a"  
<A> ::= "a"<B> | "a"<S>  
<A> ::= "a"  
<B> ::= <A><B>"a" | "b"<B><C>  
<C> ::= ""
```

Potom výstupný súbor by obsahoval napríklad:

```
<S> ::= <A>"a"  
<A> ::= "a"<S> | "a"
```

Ak by súbor so vstupnou gramatikou obsahoval:

```
<program> ::= "begin" <PRIKAZY> "end"  
<PRIKAZY> ::= <PRIKAZ> <PRIKAZY>  
<PRIKAZ> ::= "p"
```

Potom výstupný súbor by obsahoval:

PRAZDNY JAZYK

Ak by súbor so vstupnou gramatikou obsahoval:

```
<program> ::= "begin" <PRIKAZY> "end"  
<PRIKAZY> ::= <PRIKAZ> <PRIKAZY>  
<PRIKAZY> ::= ""  
<PRIKAZ> ::= "p"
```

Potom výstupný súbor by obsahoval napríklad:

```
<program> ::= "begin" <PRIKAZY> "end"  
<PRIKAZY> ::= <PRIKAZ> <PRIKAZY> | ""  
<PRIKAZ> ::= "p"
```

Changelog:
2.4.2024 Zverejnená prvá verzia zadania.