

Funkcionální projekt

BKG-2-CNF

(Funkcionální a logické programování 2020/2021)

1 Úvod

Tento dokument popisuje požadavky na projekt **BKG-2-CNF** v předmětu FLP v akademickém roce 2020/2021. Obecné pokyny pro všechna zadání jsou sepsány ve zvláštním dokumentu, který si nezapomeňte důkladně pročíst.

2 Zadání

Vytvořte program převádějící bezkontextové vlastní gramatiky (dále BKG) na bezkontextové gramatiky v Chomského normální formě (dále CNF).

Vlastní gramatika je (dle Def 4.11 Opory TIN) gramatika beze zbytečných symbolů, bez cyklů a bez epsilon-pravidel.

2.1 Rozhraní programu

Program se bude jmenovat **bkg-2-cnf** a bude jej možné spustit takto:

```
bkg-2-cnf volby [vstup]
```

kde

- *vstup* je jméno vstupního souboru (pokud není specifikováno, program čte standardní vstup) obsahujícího BKG ve formátu popsaném v sekci 3.
- *volby* jsou parametry ovlivňující chování programu, standardně tyto:
 - i na standardní výstup se vypíše pouze načtená a do vaší vnitřní reprezentace převedená BKG, a to ve formátu popsaném v sekci 3. Nevypisujte jen načtený řetězec; tato volba ověřuje, že váš program dokáže gramatiku syntakticky analyzovat a znovu vypsát.
 - 1 na stdout se vypíše BKG bez jednoduchých pravidel, ve formátu popsaném v sekci 4. Postupujte podle algoritmu 4.5 z opory předmětu TIN.
 - 2 na stdout se vypíše CNF ve formátu popsaném v sekci 4. Postupujte podle Algoritmu 4.7 z Opory TIN.

3 Formát vstupu

BKG $G = (N, \Sigma, P, S)$ na vstupu odpovídá standardní definici vlastní BKG (viz oporu předmětu TIN). Pro jednoduchost je abeceda N podmnožinou množiny velkých písmen $[A-Z]$ a abeceda Σ je podmnožinou množiny malých písmen $[a-z]$. Dále můžete předpokládat, že vstupní gramatika neobsahuje pravidlo $S \rightarrow \varepsilon$ (tedy jazyk, který vlastní gramatika generuje, neobsahuje prázdné slovo). Vstupní textová reprezentace BKG o N pravidlech má následující formát:

```
seznam všech neterminálů
seznam všech terminálů
počáteční neterminál
pravidlo1
...
pravidlon
```

Symbols jsou v seznamech odděleny čárkou. Levá strana pravidla je od pravé oddělena šipkou \rightarrow . Například reprezentace BKG z příkladu 4.1 Opory TIN je následující:

```
S,A,B
a,b,c,d
S
S->AB
A->aAb
A->ab
B->cBd
B->cd
```

4 Formát výstupu

Textová reprezentace BKG na výstupu má stejný formát jako BKG na vstupu ze sekce 3. Jelikož není možné rozhodnout, zda jsou dvě BKG ekvivalentní, je nutné, aby vaše řešení generovalo výstupy přesně odpovídající výstupům algoritmů 4.5 a 4.7 z opory předmětu TIN. Pořadí pravidel a pořadí symbolů v seznamech na výstupu není důležité, ale označení symbolů, počet a struktura pravidel musí být 100% shodná. Symbol X' se reprezentuje na výstupu jako X' , stav $\langle X_1 \dots X_k \rangle$ jako $\langle X_1 \dots X_k \rangle$. Například, $\langle ABCD \rangle$ bude $\langle ABCD \rangle$.