

Lexikální analýza 3

(195.113.207.171:32)

Cíl: Procvičování dalších aspektů lexikální analýzy:

- a/ Lex jako konečný automat z pohledu zdrojového kódu
- b/ Využití interních funkcí a maker

Příklad 1: Analýza komentářů

Navrhněte lexikální analyzátor, odstraňující klasické komentáře ze zdrojového kódu v jazyce C (*/ * komentář */*) a zachovávající (*ECHO*) zbytek programu. Mějte přitom na paměti, že C neumožňuje vnořené komentáře a Váš program by měl na podobné situace reagovat chybovým hlášením a ukončením své činnosti. Pro tento účel můžete využít funkci *yyterminate()*.

Úplný jazyk komentářů v jazyce C bohužel nepatří do kategorie regulárních, tudíž je třeba nějak si v této situaci pomoci. Jednou z běžně využívaných možností je skoková změna množiny aktuálně prohledávaných vnitřních stavů odpovídajícího konečného automatu pomocí direktivy *%x* nebo *%s*, definující jméno nového stavu (*%x COMMENT*). Odskok na něj se provede klíčovým slovem *BEGIN* a jménem stavu, tj. např. *BEGIN COMMENT*. Zpracování uvnitř stavu je uvozeno jménem stavu v ostrých závorkách (*<COMMENT>*). Základní cyklus zpracování, do kterého se ze speciálních stavů vracíme (= úvodní stav), se jmenuje *INITIAL*.

Komentáře v Pascalu jsou uvozeny složenými závorkami '*{ komentář }*'. Myslíte si, že *(f)lex*_ový vzor *\{(.|\n)*\}* je bude vyhledávat správně? Svůj názor zdůvodněte. Modifikujte program na analýzu komentářů v C tak, aby fungoval na komentáře v Pascalu a zároveň ke každému z nich určil počet řádků, které zabírá.

Příklad 2: Trocha počítání v rámci lexikální analýzy

Abychom si procvičili i další vestavěné funkce *(f)lex_u*, napíšeme s jeho pomocí obdobu klasické UNIX_ové utility *wc* (word counter), analyzující vstupní text a vypisující počet zde obsažených řádek, slov a znaků.

Váš výsledný program proto musí být schopen číst data ze standardního vstupu nebo ze souboru, vyskytuje-li se jeho jméno jako parametr příkazového řádku. Toho lze snadno dosáhnout změnou globální proměnné *yyin* tak, aby místo na *STDIN* ukazovala na název vstupního souboru.

Pro usnadnění práce můžete využít také nedokončená řešení jednotlivých úkolů ze souborů *comments-stub.c* a *wordcount-stub.c*. Vaši verzi *wc* aplikujte na soubor *test.txt*. Zdrojové kódy a sestavovací soubory ke všem třem příkladům uložte do Vámi vytvořeného podadresáře:

[S:\ucitel\Voracek_Jan\XPJP\cviceni\23_3\<prijemni_nebo_login>\](#)