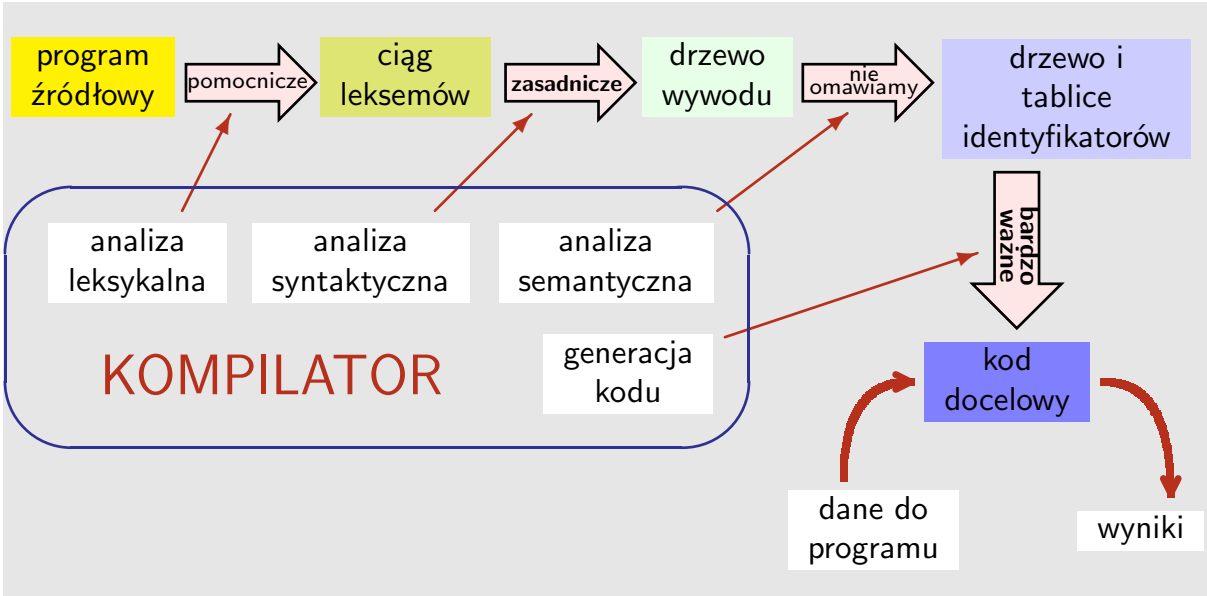
Stefan Sokołowski

JĘZYKI FORMALNE I METODY KOMPILACJI

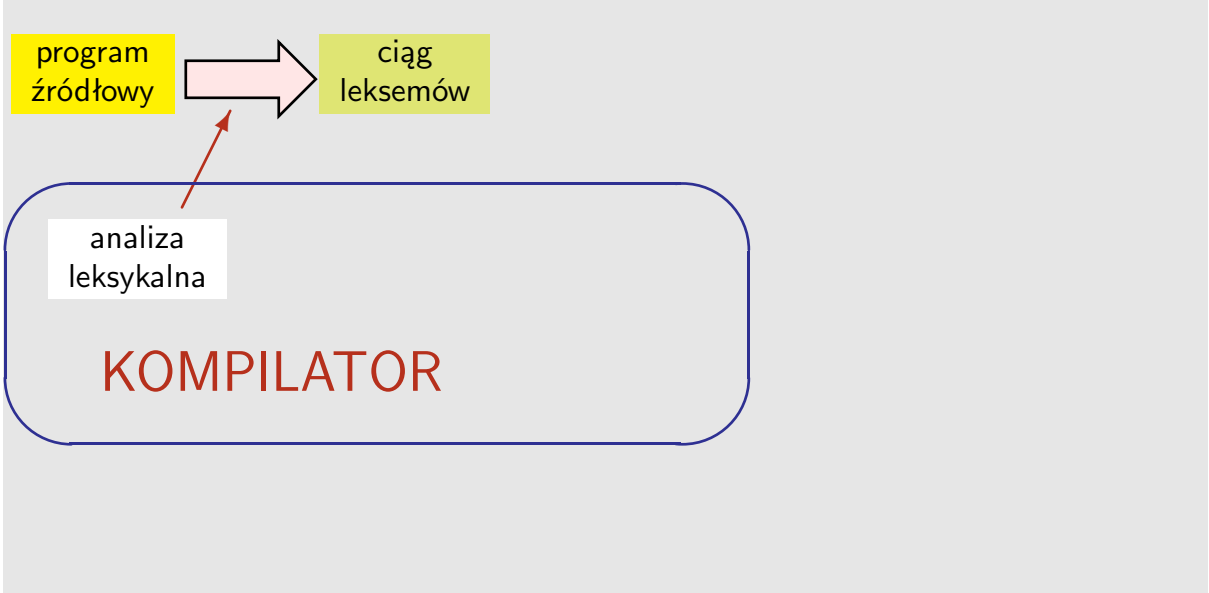
wykład 6

Inst. Informatyki Stosowanej, ANS
Elbląg, 2024/2025

Uproszczony schemat kompilacji



Uproszczony schemat kompilacji



Leksemy:
znaki ASCII, albo ich zestawienia takie jak `printf`, `else`, `3.14`, ...

Analizator leksykalny czyli skaner — przykład

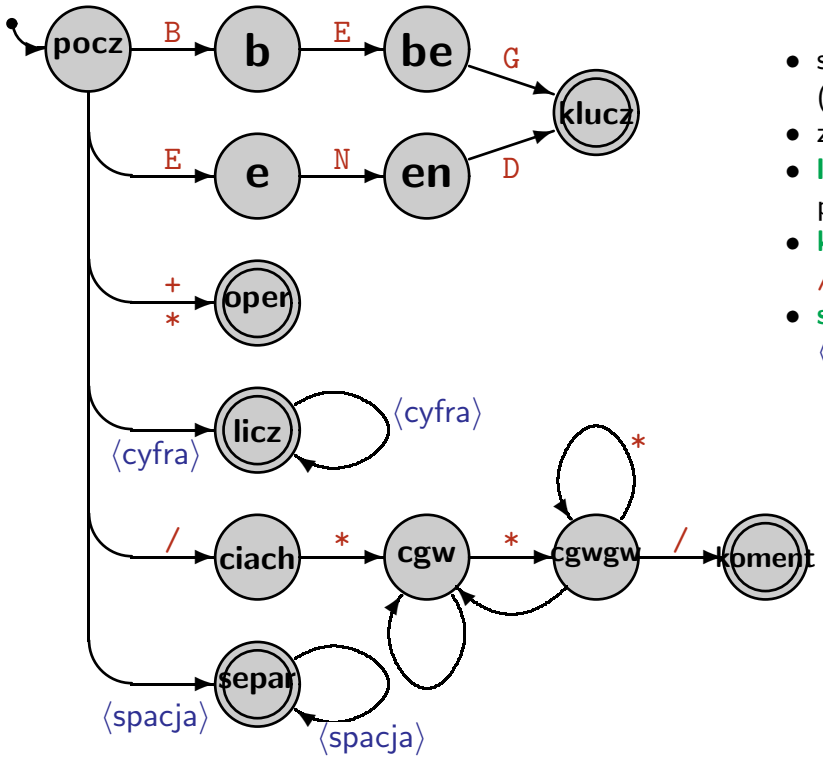
Mamy rozdzielić wejściowy ciąg symboli na *leksemy*:

- słowa **kluczowe**: `BEG`, `END` (wielkość liter nieistotna)
- znaki **operacji**: `+`, `*`
- **liczby** całkowite postaci ⟨cyfra⟩⟨cyfra⟩*
- **komentarze** postaci `/*` ⟨cokolwiek⟩ `*/`
- **separatory** postaci ⟨spacja⟩⟨spacja⟩*

PRZYKŁAD:

`bEG|12| |+|*| |/* koment * /arz */|531`

Analizator leksykalny czyli skaner — przykład



- słowa **kluczowe**: BEG, END (wielkość liter nieistotna)
- znaki **operacji**: +, *
- **liczby** całkowite postaci ⟨cyfra⟩⟨cyfra⟩*
- **komentarze** postaci /* ⟨cokolwiek⟩ */
- **separatory** postaci ⟨spacja⟩⟨spacja⟩*

Analizator leksykalny czyli skaner — przykład

Funkcja przejścia:

	b	e	g	n	d	pl	gw	cyf	ciach	sp	inny
pocz	b	e				oper	oper	licz	ciach	separ	
b		be									
be			klucz								
e				en							
en					klucz						
licz								licz			
ciach							cgw				
cgw	cgw	cgw	cgw	cgw	cgw	cgw	cgwgw	cgw	cgw	cgw	cgw
cgwgw	cgw	cgw	cgw	cgw	cgw	cgw	cgwgw	cgw	koment	cgw	cgw
separ										separ	

Uwagi:

- funkcja przejścia automatu skończonego stany × znaki → stany
- zbiór znaków rozszerzony o znak inny : „jakikolwiek inny”
- puste miejsca oraz brakujące wiersze oznaczają „stan śmietnikowy”

Po co wydziela się analizę leksykalną?

Włączenie analizy leksykalnej do analizy składniowej jest nietrudne; po co więc jest wydzielona?

1. Analiza leksykalna jest **prostsza** niż składniowa

	leksyka	**syntaktyka**
gramatyka	prawoliniowa	bezkontekstowa
akceptor	automat	maszyna stosowa

2. Analizator leksykalny czyta tekst, więc jego działanie zajmuje dużo czasu. Warto więc go optymalizować — a to robi się inaczej niż w składniowym.

3. Czytanie tekstu jest zależne od platformy, analiza składniowa jest niezależna. Rozdzielenie ich poprawia więc przenośność (*portability*) kompilatorów.