AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA W KRAKOWIE

Wprowadzenie: języki, symbole, alfabety, łańcuchy

Języki formalne i automaty

Dr inż. Janusz Majewski
Katedra Informatyki

Literatura

- Aho A. V., Sethi R., Ullman J. D.: Compilers. Principles, Techniques and Tools, Addison-Wesley, 1986 (tłumaczenie polskie: Kompilatory. Reguły, metody i narzędzia, WNT 2002).
- Homenda W.: Elementy lingwistyki matematycznej i teorii automatów, Oficyna Wydawnicza Politechniki Warszawskiej 2005.
- Hopcroft J. E., Motwani R., Ullman J. D.: Wprowadzenie do teorii automatów, języków i obliczeń, PWN, 2005.
- Waite W. M., Goos G.: Konstrukcja kompilatorów, WNT, 1989.

Lingwistyka, języki (1)

Dziedzina wiedzy:

Lingwistyka matematyczna

Czym będziemy się zajmować:

Opisywać języki

Opis języka (semiotyka)

(a) syntaktyka (opis składni)

(b) semantyka (znaczenie)

(c) pragmatyka (użyteczność)

Jaki aspekt języka będziemy opisywać?

Składnia – zasady budowy poprawnych wypowiedzi w języku (tylko poprawnych; nie bierze się pod uwagę aspektu znaczeniowego i pragmatycznego, nie analizuje się znaczenia wypowiedzi, a tym bardziej prawdziwości lub fałszu wypowiedzi)

Lingwistyka, języki (2)

Jakie języki będziemy opisywać?

Języki:

(a) naturalne (np. polski, angielski, chiński,...)

(b) formalne (sztuczne, np. Pascal, C, Java, Algol,...)

Zajmujemy się wyłącznie opisem języków formalnych.

Lingwistyka, języki (3)

Języki formalne:	**Języki naturalne:**
symbol	litera
	wyraz – sekwencja liter
łańcuch (słowo) – sekwencja symboli	zdanie – sekwencja wyrazów
	tekst – sekwencja zdań
język – zbiór łańcuchów	język – zbiór tekstów

Definiowanie języków formalnych

- Wyliczenie wszystkich poprawnych napisów w danym języku (bezużyteczne, gdy język zawiera nieskończenie wiele napisów).
- Podanie zasad **budowy** wszystkich poprawnych napisów w danym języku (**gramatyka języka**).
- Podanie algorytmu orzekającego, czy dany napis jest poprawnym napisem w danym języku (**automat abstrakcyjny**).
- Podanie „ogólnego wzoru” określającego postać wszystkich poprawnych napisów w danym języku (**wyrażenie regularne** wykorzystywane do definiowania szczególnie prostych języków zwanych językami regularnymi).

Symbol, alfabet, łańcuch (1)

<u>Symbol</u>

Symbol jest to pojęcie niedefiniowane (synonimy: znak, litera)

<u>Alfabet</u>

Alfabet jest to niepusty, skończony zbiór symboli, np.:

Σ_1 = {a, b, c, ..., x, y, z } - alfabet łaciński

Σ_2 = {0, 1 } - alfabet binarny

Symbol, alfabet, łańcuch (2)

<u>Łańcuch</u>

Łańcuch (synonim słowo) to skończony ciąg zestawionych razem symboli z alfabetu.

Rekurencyjna definicja łańcucha nad alfabetem Σ:

1) ε jest łańcuchem nad Σ (ε - łańcuch pusty, nie zawierający żadnego znaku),
2) jeśli x jest łańcuchem nad Σ i $a \in \Sigma$ to xa jest łańcuchem nad Σ,
3) nic innego nie jest łańcuchem poza tym, co wynika z punktów (1) i (2).

Złożenie (konkatenacja) łańcuchów (1)

Konkatenacja (złożenie) łańcuchów

Jeśli x i y są łańcuchami

$x = x_1x_2...x_m$

$y = y_1y_2...y_n$

to

$xy = x_1x_2...x_my_1y_2...y_n$

jest nazywane konkatenacją (złożeniem) łańcuchów x i y.

Przykład:

$x = ab$

$y = cde$

$xy = abcde$ ale: $yx = cdeab$

Złożenie (konkatenacja) łańcuchów (2)

Dla dowolnego łańcucha x zachodzi

$x\varepsilon = \varepsilon x = x$

Konkatenacja łańcuchów jest:

- nieprzemienna (na ogół $xy \neq yx$)
- łączna ($xyz = (xy)z = x(yz)$)
- posiada element neutralny ε będący łańcuchem pustym ($x\varepsilon = \varepsilon x = x$)

Notacja potęgowa

Notacja „potęgowa”

Niech *a* będzie symbolem. Niech *x* będzie łańcuchem. Wtedy łańcuchy złożone z wielokrotnych powtórzeń symbolu *a* lub łańcucha *x* oznaczamy w uproszczeniu:

$a^0 = \varepsilon$	$x^0 = \varepsilon$
$a^1 = a$	$x^1 = x$
$a^2 = aa$	$x^2 = xx$
$a^3 = aaa$, ... , itd.	$x^3 = xxx$, ..., itd.

Długość łańcucha

Długość łańcucha

Długość $|x|$ łańcucha *x* to liczba symboli tworzących ten łańcuch.

Definicja rekurencyjna długości łańcucha:

1) $|\varepsilon| = 0$,
2) jeśli *x* jeśli łańcuchem, zaś *a* jest symbolem $(a \in \Sigma)$, to $|ax| = 1 + |x|$.

Rewers łańcucha

Rewersem (odbiciem zwierciadlanym) x^R łańcucha $x = x_1x_2...x_n$ nazywamy łańcuch $x^R = x_nx_{n-1}...x_1$, w którym symbole zapisane są w odwrotnym porządku, przy czym $\varepsilon^R = \varepsilon$.

Definicja rekurencyjna odbicia zwierciadlanego:

1) $\varepsilon^R = \varepsilon$,
2) jeśli x jest łańcuchem nad alfabetem Σ, zaś a jest symbolem $(a \in \Sigma)$, to $(ax)^R = x^Ra$,
3) nic innego nie jest odbiciem zwierciadlanym poza tym, co wynika z punktów (1) i (2).

Palindrom

Palindromem nazywamy łańcuch, który czytany od przodu oraz czytany wspak brzmi tak samo, czyli

$x^R = x$

Definicja rekurencyjna palindromu:

1) ε jest palindromem,
2) jeśli a jest symbolem $(a \in \Sigma)$, to łańcuch zbudowany z pojedynczego symbolu a jest palindromem,
3) jeśli x jeśli palindromem, zaś a jest symbolem $(a \in \Sigma)$, to łańcuch axa jest palindromem,
4) nic innego nie jest palindromem poza tym, co wynika z punktów (1), (2) i (3).

Przykłady palindromów: oko, kajak, kobyłamamałybok.