

Gramatyki. Notacja BNF

Aspekty języków:

- > Syntaktyka (składnia) - zbiór reguł określający formalnie poprawne konstrukcje językowe
- > Semantyka - opisuje znaczenie konstrukcji językowych
- > Pragmatyka - opisuje wszystkie inne aspekty języka

Na tych zajęciach zajmiemy się syntaktyką, czyli sposobem określania reguł poprawnych konstrukcji dla zadanych języków (w tym np. języków programowania).

Każdy język zbudowany jest na bazie alfabetu:

Alfabet $V = \{v_0, \dots, v_n\}$ - dowolny skończony i niepusty zbiór symboli (liter), z których będą zestawiane słowa języka.

Słowo nad alfabetem V - wszystkie skończone, uporządkowane ciągi złożone z elementów alfabetu.

Słowo puste ϵ - słowo nie mające żadnej litery.

Język uniwersalny V^* - zbiór wszystkich słów nad alfabetem V .

Język - dowolny podzbiór języka uniwersalnego V^* nad V .

Syntaktyka (składnia) języka - reguły budowy zdań w języku.

Przykład: Jeśli $V = \{a, b\}$, to $V^* = \{\epsilon, a, b, aa, ab, bb, aaa, aab, \dots\}$.

Gramatyką formalną nazywamy sposób opisu języka formalnego, czyli podzbiór zbioru wszystkich słów skończonej długości nad danym alfabetem.

Formalnie gramatykę określamy jako: $G = \langle V, T, P, S \rangle$

- V - zbiór symboli terminalnych - skończony, niepusty zbiór symboli końcowych, z których budowane są słowa generowane przez gramatykę (zwany czasem alfabetem końcowym),
- T - zbiór symboli nieterminalnych - skończony, niepusty zbiór symboli pomocniczych,
- P - lista produkcji - reguły gramatyki,
- S - symbol startowy (aksjomat) - jest to wyróżniony symbol pomocniczy, z niego wyprowadzane są wszystkie generowane przez gramatykę G napisy.

Produkcją nazwiemy regułę zastępowania symbolu pomocniczego przez słowa, w skład których mogą wchodzić zarówno symbole pomocnicze, jak i końcowe. Proces wyprowadzania słów języka zdefiniowanego przez gramatykę zaczynamy od symbolu S . Do poszczególnych symboli pomocniczych stosujemy którąś z produkcji ze zbioru P tak długo, aż wszystkie symbole pomocnicze znikną.

Przykład:

Dla języka postaci $V^* = \{1, 11, 111, \dots\}$, reguły go definiujące mają postać:

Symbol początkowy: S

Produkcje (reguły zastępowania): $S \rightarrow 1$ (1)

$S \rightarrow S1$ (2)

Wywód : $1 : S \xrightarrow{1} 1$ (3)

$111 : S \xrightarrow{2} S1 \xrightarrow{2} S11 \xrightarrow{1} 111$ (4)

Słowo złożone z samych symboli końcowych, nazywamy słowem wyprowadzalnym z gramatyki. Jednym ze sposobów opisu składni języka jest np. „notacja BNF”.

Notacja BNF

Notacja Backusa-Naura (ang. Backus-Naur form) jest sposobem zapisu reguł gramatyki bezkontekstowej, czyli sposobem opisu języków formalnych. Notacja ta jest powszechnie używana w informatyce do zapisu składni języków programowania i protokołów komunikacyjnych. Została wymyślona przez Johna Backusa w latach 50. w czasie pracy nad językiem Fortran, a następnie zmodyfikowana przez Petera Naura i użyta do zdefiniowania składni języka Algol.

Elementy notacji BNF:

Symbole nieterminalne, symbole terminalne, produkcje

$\langle \rangle$ - symbol pomocniczy
 \rightarrow zastępowana jest przez $::=$ - symbol czytany „jest równe z definicji”, łączy strony produkcji
 $|$ - symbol alternatywy
 $\{\}_0^n$ - powtórzenie 0 lub więcej razy
 $[]$ - elementy opcjonalne

Przykład 1: Aby przy pomocy notacji BNF określić liczbę naturalną, można użyć następujących reguł:

$\langle \text{zero} \rangle ::= 0$
 Przykład wartości: 0
 $\langle \text{cyfra niezerowa} \rangle ::= 1|2|3|4|5|6|7|8|9$
 Przykład wartości: 1, 2, 3
 $\langle \text{cyfra} \rangle ::= \langle \text{zero} \rangle | \langle \text{cyfra niezerowa} \rangle$
 Przykład wartości: 0, 1, 2, 3
 $\langle \text{ciąg cyfr} \rangle ::= \langle \text{cyfra} \rangle | \langle \text{cyfra} \rangle \langle \text{ciąg cyfr} \rangle$
 Przykład wartości: 0, 1, 01, 23, 45, 99, 10023, 000001
 $\langle \text{liczba naturalna} \rangle ::= \langle \text{cyfra} \rangle | \langle \text{cyfra niezerowa} \rangle \langle \text{ciąg cyfr} \rangle$
 Przykład wartości: 0, 1, 2, 3456, 406, 556066

Przykład 2: Poniżej przedstawiono prosty przykład gramatyki bezkontekstowej, zapisanej w notacji BNF, generującej liczby binarne:

$\langle \text{liczba_binarna} \rangle ::= \langle \text{liczba_binarna} \rangle \langle \text{cyfra} \rangle$
 $\langle \text{liczba_binarna} \rangle ::= \langle \text{cyfra} \rangle$
 $\langle \text{cyfra} \rangle ::= 0|1$

Gramatyka ta ma więc postać ogólną:

$G = \{V, T, P, S\}$
 $V = \{0, 1\}$,
 $T = \{\langle \text{liczba_binarna} \rangle, \langle \text{cyfra} \rangle\}$,
 $P = \{\langle \text{liczba_binarna} \rangle ::= \langle \text{liczba_binarna} \rangle \langle \text{cyfra} \rangle;$
 $\quad \langle \text{liczba_binarna} \rangle ::= \langle \text{cyfra} \rangle ;$
 $\quad \langle \text{cyfra} \rangle ::= 0|1\}$,
 $S = \langle \text{liczba_binarna} \rangle$.

Zadania na ćwiczenia

Zadanie 1

Podaj przykład gramatyki generującej język $\{a^n b^n : n \in \mathbb{N}\}$

Zadanie 2

Podaj przykład gramatyki generującej język $\{w : w \in \{a, b\}^* \wedge w = w^R\}$, który jest językiem wszystkich palindromów (wyrazów, które wyglądają tak samo niezależnie od tego, czy czytamy je normalnie, czy od tyłu) nad alfabetem $\{a, b\}$.

Zadanie 3

Podaj przykład gramatyki generującej język $\{a^k b^l c^n : k, l, n \in \mathbb{N}\}$, oraz takiej samej w przypadku, gdy $k = m$.

Zadanie 4

Zdefiniować gramatykę generującą liczby binarne parzyste.

Zadanie 5

Dany jest język $L(G)$ nad gramatyką:

$G = \langle V, T, P, S \rangle$

$V = \{a, b\}$

$T = \{S, A\}$

Produkcje:

$S \rightarrow aAS \mid a$

$A \rightarrow SbA \mid SS \mid ba$

Czy słowo $aabbaa$ należy do języka? Aby słowo należało do języka należy udowodnić, że a) Słowo jest wyprowadzalne z S . b) Słowo to należy do T .

Zadanie 6

Dana jest gramatyka: $G = (V, T, P, S)$

$V = \{a, b\}$,

$T = \{A, B, S\}$,

$P = \{P : S \rightarrow aSB \mid aAS \mid abS \mid Sba \mid ab$

$A \rightarrow aa$

$B \rightarrow bb\}$,

$S = C$

Czy słowa: $aababbabb$, $aaaaabbb$, $aaaabbba$ są słowami poprawnymi języka $L(G)$ (języka wywiedzionego z tej gramatyki)?

Informacje końcowe

Jedno zadanie na kolokwium będzie obejmowało zadanie takie jak zadania 1 – 4 i jedno takie jak zadania 5 – 6.