

FORMÁLNÍ LINGVISTIKA

ZÁKLADNÍ POJMY

Abeceda je konečná množina symbolů. Např. { , }.

Slovo je libovolná konečná posloupnost prvků Σ , např. .

Délka slova je velikost této posloupnosti, např. | | = 5.

Prázdné slovo je slovo nulové délky, značíme je .

ZÁKLADNÍ POJMY

Σ^* je množina všech slov nad abecedou Σ , např. $\{a, b\}^* = \{\epsilon, a, b, aa, ab, ba, bb, aaa, \dots\}$.

Operace zřetězení slov, značíme tečkou (\cdot), je pro dvě slova x a y definována jako $x \cdot y = xy$, např. $a \cdot b = ab$.

Mocnina slova je pro slovo x a přirozené číslo n značena x^n a je definována induktivně: $0 = \epsilon$, $x^{n+1} = x \cdot x^n$. Např. $(ab)^3 = ababab$.

Jazyk je množina L slov nad danou abecedou, tedy pro každý jazyk L platí $L \subseteq \Sigma^*$

FORMÁLNÍ GRAMATIKA

Formální gramatika přepisovací systém, jímž lze vygenerovat slova jazyka.

Formálně, gramatika je čtveřice (Q, Σ, P, q_0) , kde:

Q je množina neterminálů (značíme nejčastěji Q mi písmeny)

Σ je množina terminálů (symbolů abecedy, značíme Σ mi písmeny), je disjunkt ní s množinou Q a $Q \cup \Sigma$ označujeme jako (množina všech symbolů)

P je množina pravidel, tj. množina dvojic, kde prvním prvkem je řetězec obsahující alespoň jeden neterminál a druhým prvkem je libovolný řetězec.

q_0 je počáteční symbol gramatiky

GRAMATIKA

Pravidla gramatiky jako dvojice řetězců (slov nad množinou Σ) (α, β) zapisujeme jako $\alpha \rightarrow \beta$. α nazýváme levou stranou pravidla a musí obsahovat alespoň jeden neterminál. β nazýváme pravou stranou pravidla.

Gramatika je modelem, kterým lze generovat slova jazyka:

začneme z počátečního symbolu gramatiky S a pravidla gramatiky používáme jako přepisovací systém, to znamená, že v jednom kroku přepisu můžeme nahradit některý řetězec terminálů a neterminálů, který je současně na levé straně nějakého pravidla, pravou stranou tohoto pravidla. Tento postup opakujeme, dokud nedostaneme řetězec terminálních symbolů (čili slovo nad Σ).

Tomuto procesu říkáme odvození slova z gramatiky.

GRAMATIKA

Řekneme, že gramatika generuje jazyk , pokud existuje odvození každého slova jazyka z gramatiky . Jazyk generovaný gramatikou značíme většinou ().

Příklad!

Mějme gramatiku (, Σ , ,), kde

$$\Sigma = \{ , \}$$

$$= \{ , \}$$

$$= \{ \rightarrow , \rightarrow , \rightarrow \}$$

$$\Rightarrow \Rightarrow$$

$$\Rightarrow \Rightarrow \Rightarrow \Rightarrow \Rightarrow \Rightarrow$$

CHOMSKÉHO HIERARCHIE JAZYKŮ

Gramatika typu 0 neklade žádná omezení na množinu pravidel, libovolná gramatika je gramatikou typu 0.

Gramatika typu 1 neboli kontextová gramatika klade na všechna pravidla \rightarrow podmínku $| \alpha| \leq | \beta|$, tedy levá strana každého pravidla musí být kratší než jeho pravá strana. Přímkou z tohoto omezení je pravidlo \rightarrow , které může být přítomno.

Gramatika typu 2 neboli bezkontextová gramatika má všechna pravidla ve tvaru \rightarrow (tak, že \in), tedy na levé straně je vždy právě jeden neterminál a α je neprázdné. Přímkou z tohoto omezení je pravidlo \rightarrow , které může být přítomno.

Gramatika typu 3 neboli regulární gramatika má všechna pravidla ve tvaru \rightarrow nebo \rightarrow , kde α jsou neterminály a β je terminál. Přímkou z tohoto omezení je pravidlo \rightarrow , které může být přítomno.