

Иерархия Хомского

Материал из Википедии — свободной энциклопедии

Иерархия Хомского — классификация формальных языков и формальных грамматик, согласно которой они делятся на 4 типа по их условной сложности. Предложена профессором Массачусетского технологического института, лингвистом Ноамом Хомским.

Содержание

Классификация грамматик

Тип 0 — неограниченные

Тип 1 — контекстно-зависимые

Тип 2 — контекстно-свободные

Тип 3 — регулярные

Классификация языков

Примечания

Литература

Классификация грамматик

Согласно Хомскому, формальные грамматики можно разделить на четыре типа. Для отнесения грамматики к тому или иному типу необходимо соответствие *всех* её правил (продукций) некоторым схемам.

Тип 0 — неограниченные

Грамматика с фразовой структурой *G* — это алгебраическая структура, упорядоченная четвёрка

(

V

T

,

V

N

,
P
,
S
)

{\displaystyle (V_{T},V_{N},P,S)}

, где^[1]:

- V_T* — *алфавит* (множество) терминальных символов — *терминалов*,
- V_N* — *алфавит* (множество) нетерминальных символов — *нетерминалов*,
- V* = *V_T* ∪ *V_N* — *словарь G*, причём *V_T* ∩ *V_N* = ∅
- P* — *конечное множество продукций* (правил) грамматики, *P* ⊆ *V*⁺ × *V*^{*}
- S* — *начальный символ* (*источник*).

Здесь *V*^{*} — множество всех строк над алфавитом *V*, а *V*⁺ — множество непустых строк над алфавитом *V*.

К типу 0 по классификации Хомского относятся неограниченные грамматики — грамматики с фразовой структурой, то есть все без исключения формальные грамматики. Правила можно записать в виде:

$$\alpha \rightarrow \beta,$$

где $\alpha \in V^+$ — любая непустая цепочка, содержащая хотя бы один нетерминальный^[2] символ, а $\beta \in V^*$ — любая цепочка символов из алфавита.

Практического применения в силу своей сложности такие грамматики не имеют.

Тип 1 — контекстно-зависимые

К этому типу относятся контекстно-зависимые (КЗ) грамматики и неукорачивающие грамматики. Для грамматики $G(V_T, V_N, P, S)$, $V = V_T \cup V_N$ все правила имеют вид^[3]:

- $\alpha A \beta \rightarrow \alpha \gamma \beta$, где $\alpha, \beta \in V^*$, $\gamma \in V^+$, $A \in V_N$. Такие грамматики относят к контекстно-зависимым.
- $\alpha \rightarrow \beta$, где $\alpha, \beta \in V^+$, $1 \leq |\alpha| \leq |\beta|$. Такие грамматики относят к неукорачивающим.

Эти классы грамматик эквивалентны. Могут использоваться при анализе текстов на естественных языках, однако при построении компиляторов практически не используются в силу своей сложности. Для контекстно-зависимых грамматик доказано утверждение: по некоторому алгоритму за конечное число шагов можно установить, принадлежит цепочка терминальных символов данному языку или нет.

Тип 2 — контекстно-свободные

К этому типу относятся контекстно-свободные (КС) грамматики. Для грамматики $G(V_T, V_N, P, S)$, $V = V_T \cup V_N$ все правила имеют вид:

- $A \rightarrow \beta$, где $\beta \in V^+$ (для неукорачивающих КС-грамматик) или $\beta \in V^*$ (для укорачивающих), $A \in V_N$. То есть грамматика допускает появление в левой части правила только нетерминального символа.

КС-грамматики широко применяются для описания синтаксиса компьютерных языков (см. синтаксический анализ).

Тип 3 — регулярные

К третьему типу относятся регулярные грамматики (автоматные) — самые простые из формальных грамматик. Они являются контекстно-свободными, но с ограниченными возможностями.

Все регулярные грамматики могут быть разделены на два эквивалентных класса, которые для грамматики вида III будут иметь правила следующего вида:

- $A \rightarrow B\gamma$ или $A \rightarrow \gamma$, где $\gamma \in V_T^*$, $A, B \in V_N$ (для левوليнейных грамматик).
- $A \rightarrow \gamma B$; или $A \rightarrow \gamma$, где $\gamma \in V_T^*$, $A, B \in V_N$ (для правوليнейных грамматик).

Регулярные грамматики применяются для описания простейших конструкций: идентификаторов, строк, констант, а также языков ассемблера, командных процессоров и др.

Классификация языков

Формальные языки классифицируются в соответствии с типами грамматик, которыми они задаются. Однако, один и тот же язык может быть задан разными грамматиками, относящимися к разным типам. В таком случае, считается, что язык относится к наиболее простому из них. Так, язык, описанный грамматикой с фразовой структурой, контекстно-зависимой и контекстно-свободной грамматиками, будет контекстно-свободным.

Так же, как и для грамматик, сложность языка определяется его типом. Наиболее сложные — языки с фразовой структурой (сюда можно отнести естественные языки), далее — КЗ-языки, КС-языки и самые простые — регулярные языки.

Примечания

1. Кук, Бейз, 1990, с. 258,264.
2. *Серебряков В. А., Галочкин М. П., Гончар Д. Р., Фуругян М. Г.* Теория и реализация языков программирования. — М. : МЗ-Пресс, 2006. — С. 21. — ISBN 5-94073-094-9.
3. Кук, Бейз, 1990, с. 268.

Литература

- *Молчанов А. Ю.* Системное программное обеспечение. СПб.: Питер, 2006.
- *Джон Хопкрофт, Раджив Мотвани, Джеффри Ульман. ГЛАВА 5. Контекстно-свободные грамматики и языки* // Введение в теорию автоматов, языков и вычислений = Introduction to Automata Theory, Languages, and Computation. — М.: «Вильямс», 2002. — С. 528. — ISBN 0-201-44124-1.
- *Серебряков В. А., Галочкин М. П., Гончар Д. Р., Фуругян М. Г.* Теория и реализация языков программирования (http://trpl7.ru/t-books/TRYAP_BOOK_Details.htm) — М.: МЗ-Пресс, 2006 г., 2-е изд. — ISBN 5-94073-094-9
- *Робин Хантер.* Основные концепции компиляторов = The Essence of Compilers. — М.: «Вильямс», 2002. — С. 256. — ISBN 5-8459-0360-2.
- *Кук Д., Бейз Г.* Глава 8. Языки и грамматики // Компьютерная математика = Computer Mathematics. — М.: Наука. Физматлит, 1990. — 384 с. — ISBN 5-02-014216-6.

Источник — https://ru.wikipedia.org/w/index.php?title=Иерархия_Хомского&oldid=107601227

Эта страница в последний раз была отредактирована 11 июня 2020 в 07:02.

Текст доступен по лицензии Creative Commons Attribution-ShareAlike; в отдельных случаях могут действовать дополнительные условия.

Wikipedia® — зарегистрированный товарный знак некоммерческой организации Wikimedia Foundation, Inc.