

**АЛГОРИТМЫ АНАЛИЗА ТЕКСТОВ,
СФОРМИРОВАННЫЕ С ПРИМЕНЕНИЕМ
РАСШИРЯЕМОГО ВХОДНОГО ЯЗЫКА**

Хакимов Муфтах Хамидович, Кадиров Бабур

Национальный университет Узбекистана им. Мирзо Улугбека, г. Ташкент

Каршинский филиал ТУИТ, г. Карши

Определение 1. *Формальной системой* называется система, состоящая из множества специальных символов, множества понятий, баз слов/фраз и конечного множества математических моделей считающихся интерпретируемыми.

Определение 2. *Математическая модель естественного языка* – это есть способ формального описания его синтаксических и семантических конструкций. Основой синтаксических конструкций является вывод слово, а семантических конструкций правильный вывод фразы.

Утверждение 3. Математическая модель является *распознающей*, если она характеризует язык **A** или является *порождающей* если характеризует язык **B**. При этом направлением МП считается $A \rightarrow B$, а языки **A** и **B** принадлежат по классификации Н. Хомского [29] классу 0.

Определение 3. *Распознающая математическая модель*, это анализ синтаксических и семантических конструкций выводящих висячее дерево предложений языка **A**.

Определение 4. *Порождающая математическая модель*, это синтез синтаксических и семантических конструкций строящих дерево предложения языка **B**.

Определение 9. Формальная грамматика G естественного языка

это есть

$G = \{\Lambda, \Phi, \Psi, \Pi\}$, где

- Λ множество терминальных символов;

- Φ вспомогательное множество нетерминальных символов и фраз, с помощью которых определяются терминальные символы и понятия;

- множество продукций $\Psi: \chi \rightarrow \gamma$ ($\chi \neq \gamma$, $\chi \in \{\Phi\}$, $\gamma \in \{\Lambda \cup \Phi\}$);

- Π начальный символ, $\Pi = \langle \text{математическая модель ЕЯ} \rangle$.

Расширяемый входной язык для обработки естественных языков

Алфавит входного языка включает:

- все заглавные и прописные буквы латинского алфавита,
- арабские цифры от 0 до 9,
- специальные символы: \oplus , \forall , \downarrow , \$,], [,), (, -, =, и запятая.

Каждая буква, цифра и символ имеют свою суть назначения:

- “-“ разделитель нижней гран] , [,) , и (формальные скобки;
- “-“ разделитель нижней границы от верхней,

“=” формальный знак равенства, используемый при описании математической модели.

В языке имеется четыре типа операций:

\oplus - операция присоединения;

\vee - операция “или”;

\Downarrow - операция “подключения” (присоединения) или “не подключения” (не присоединения);

$\$$ - операция выбора, с синтаксисом вида: $\$_{[<i>,<1-m>]} <P_i>$

Каждое сочетание алфавитно-цифровых символов несут определенную функциональную нагрузку. В табл. приведены сочетание алфавитно-цифровых символов и выполняемые ими функции. Одной из компонентов функций является установление связей с соответствующими БД.

Таблица (часть, общ. кол-во эл. 94)

Символы	Функция	Символы	Функция
А	аффикс	М3	указательное местоимение
А1	аффиксы образующие слова	М3А	аффиксы указательного местоимения
А2	аффиксы образующие форму	М4	вопросительное местоимение
А3	аффиксы изменяющие слова	М4А	аффиксы вопросительного местоимения
А4	аффикс для других оснований	М5	определительное местоимение

*Логико-лингвистические модели
повествовательных предложений*

(число моделей – 12)

1. существительное ⊕ ↓↓ наречие ⊕ ↓↓ местоимение ⊕ ↓↓

числительное ⊕ ↓↓ прилагательное ⊕ ↓↓ глагол ⊕ ↓↓

прилагательное ⊕ существительное ⊕ ↓↓ местоимение ⊕ ↓↓

прилагательное ⊕ ↓↓ глагол ⊕ ↓↓ прилагательное ⊕ ↓↓

числительное

*Математические модели
повествовательных предложений*

$$\begin{aligned}
& E1(C, G, N, L1, P, M, F, Y, L, D(E)) = (\$_{[h31,1-h20]}C_{h31} \oplus \$_{[h33,1-h22]}G_{h33} \\
& \oplus \Downarrow \$_{[h35,1-h24]}N_{h35}) \quad \mathbf{V} \quad (\$_{[i,1-3]}L1_i \oplus \$_{[h32,1-h21]}P_{h32} \oplus \$_{[h31,1-h20]}C_{h31} \\
& \oplus \$_{[h33,1-h22]}G_{h33} \oplus \Downarrow \$_{[h321,1-h21]}P_{h321}) \quad \mathbf{V} \quad (\$_{[i,1-3]}L1_i \oplus \Downarrow \\
& \$_{[h32,1-h21]}P_{h32} \oplus \$_{[h31,1-h20]}C_{h31} \oplus \Downarrow \$_{[h33,1-h22]}G_{h33} \oplus \Downarrow \\
& \$_{[h35,1-h24]}N_{h35}) \quad \mathbf{V} \quad (\$_{[h34,1-h23]}M_{h34} \oplus \$_{[h33,1-h22]}G_{h33} \oplus \Downarrow \\
& \$_{[h31,1-h20]}C_{h31} \oplus \Downarrow \$_{[h32,1-h2]}P_{h32}) \quad \mathbf{V} \quad (\$_{[h31,1-h20]}C_{h31} \oplus
\end{aligned}$$

$$(\$_{[h33,1-h22]} \mathbf{G}_{h33} \oplus \Downarrow \$_{[h32,1-h21]} \mathbf{P}_{h32}) \mathbf{V} (\$_{[h31,1-h20]} \mathbf{C}_{h31} \oplus$$

$$\$_{[h33,1-h22]} \mathbf{G}_{h33} \oplus \Downarrow \$_{[h32,1-h21]} \mathbf{P}_{h32} \oplus \Downarrow \$_{[h311,1-h20]} \mathbf{C}_{h311} \oplus \Downarrow$$

$$\$_{[h35,1-h24]} \mathbf{N}_{h35} \oplus \Downarrow \$_{[h34,1-h23]} \mathbf{M}_{h34} \oplus \Downarrow \$_{[h331,1-h22]} \mathbf{G}_{h331}) \mathbf{V}$$

$$(\$_{[h34,1-h23]} \mathbf{M}_{h34} \quad \cdot \quad \cdot \quad \cdot)$$

Укрупненные алгоритмы анализа предложений

Алгоритм N1.

1. Выбрать существительному из EVX соответствующий ему существительное из базы CC.
2. Выбрать глаголу из EVX соответствующий ему глагол из базы GG.
3. Если имеется наречие в EVX, тогда выбрать соответствующий ему наречие из базы NN.
4. Записать выбранные слова в таблицу EE1.

Алгоритм N2.

1. Выбрать артиклу из EVX соответствующий ему слово из базы L1.
2. Если имеется прилагательное из EVX, тогда выбрать соответствующие ему прилагательное из базы PP.
3. Выбрать существительному из EVX соответствующий ему существительное из базы CC.
4. Если имеется глагол из EVX, тогда выбрать соответствующий ему глагол из базы GG.
5. Если имеется другое прилагательное из EVX, тогда выбрать соответствующее ему прилагательное из базы PP.
6. Записать выбранные слова в таблицу EE1.

СПАСИБО ЗА ВНИМАНИЕ