

# Компьютерный анализ естественно-языкового текста

Кафедра информационных систем в  
искусстве и гуманитарных науках

## СТРУКТУРА КУРСА

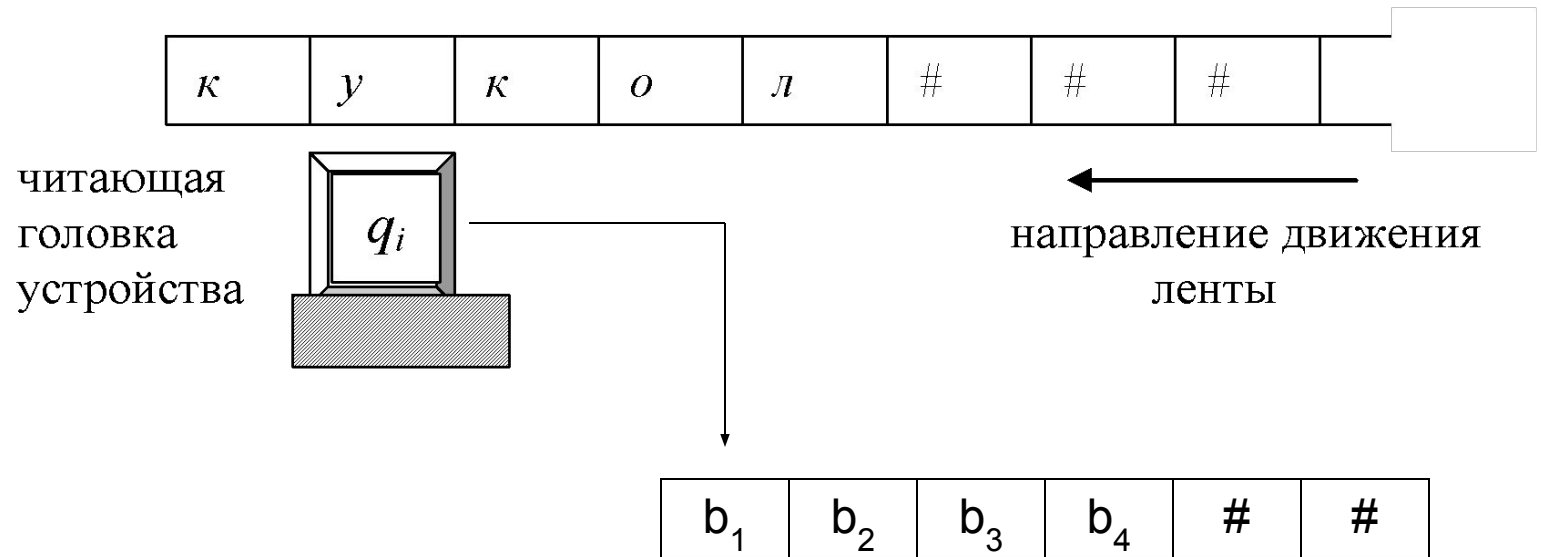
1. Введение в дисциплину
2. Автоматический анализ текста на морфологическом уровне
3. Автоматический анализ текста на синтаксическом уровне
4. Семантический компонент в системах автоматического анализа текста

## СТРУКТУРА КУРСА

### 2. Автоматический анализ текста на морфологическом уровне

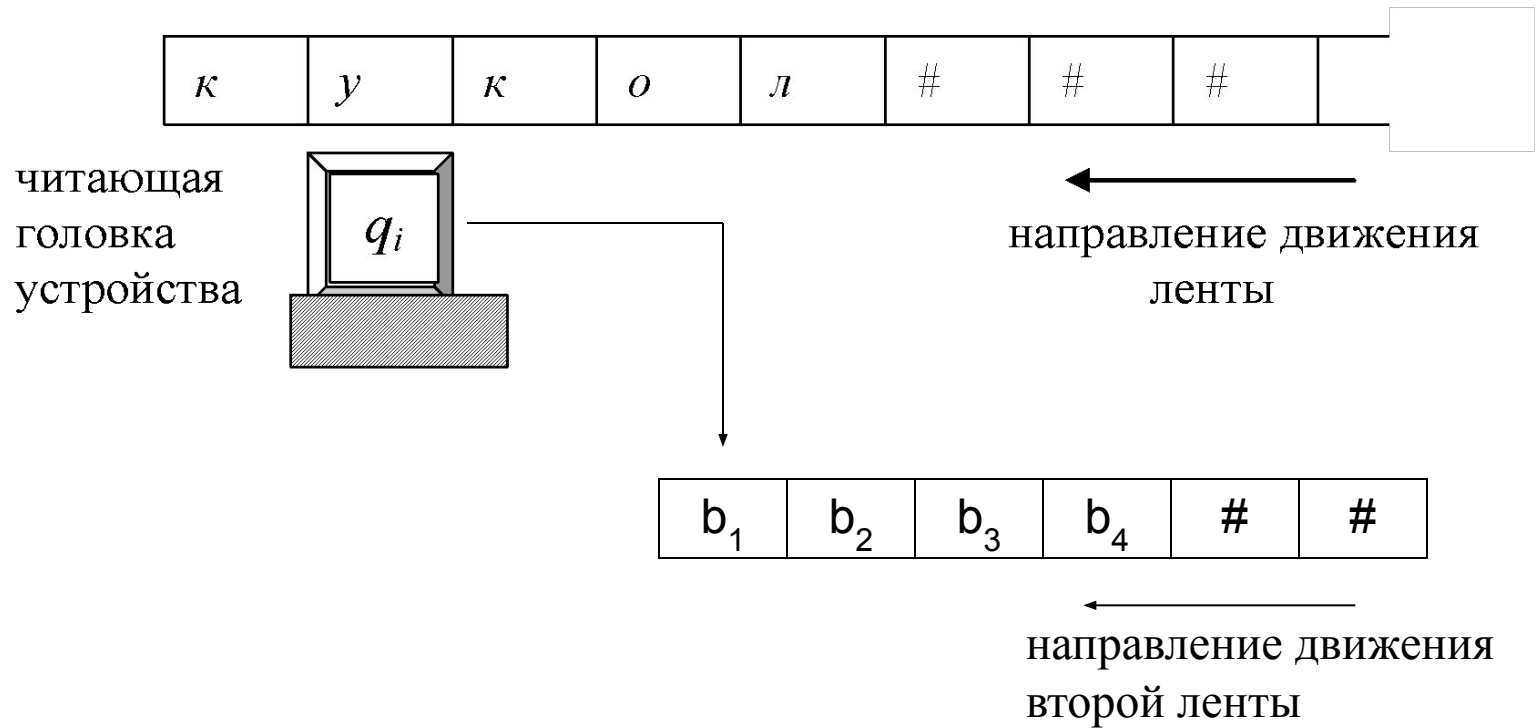
- 1. Морфологический уровень в ЛИТ*
- 2. Основные понятия морфологии в компьютерной морфологии*
- 3. Основные процедуры компьютерной морфологии*
- 4. Компьютерная морфология русского языка*
- 5. Технологии морфологического анализа*
- 6. «Предсказание» (типизация)*
- 7. Вопросы, смежные с синтаксисом*

# СТЕКОВЫЙ АВТОМАТ



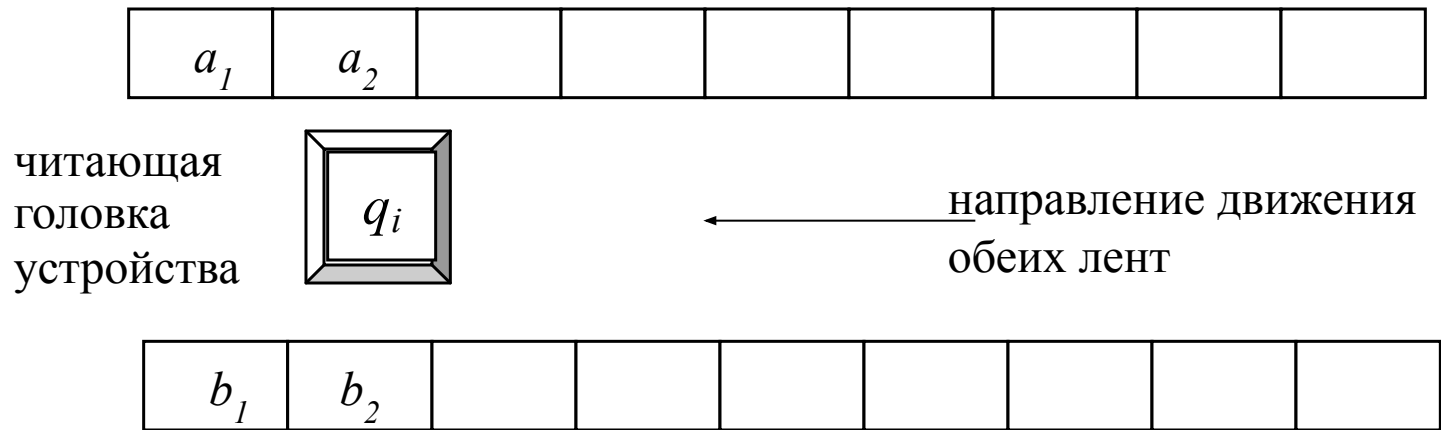
Правила вида  $q_i a_i b_i \rightarrow q_j V_j$

# КОНЕЧНЫЙ ПРЕОБРАЗОВАТЕЛЬ



Правила вида  $q_i a_i b_i \rightarrow q_j$

# КОНЕЧНЫЙ ПРЕОБРАЗОВАТЕЛЬ



Правила вида  $q_i a_i b_i \rightarrow q_j$

# КОНЕЧНЫЙ ПРЕОБРАЗОВАТЕЛЬ: определение

Шестерка  $(Q, \Sigma, T, \delta, q_0, F)$ :

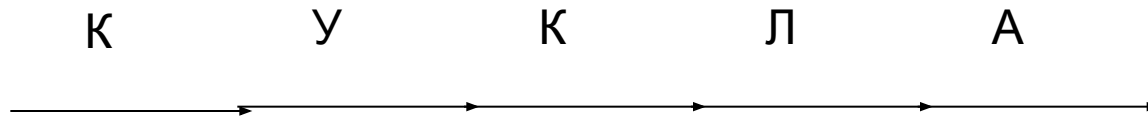
- $Q$  – конечное множество (состояния)
- $\Sigma$  – конечное множество (входной алфавит)
- $T$  – конечное множество (выходной алфавит)
- $\delta$  – отношение между  $Q \times \Sigma \times T$  и  $Q$  (правила)
- $q_0$  – элемент из  $Q$  (начальное состояние)
- $F$  – подмножество  $Q$  (заключительные состояния)

# ФУНКЦИИ КОНЕЧНОГО ПРЕОБРАЗОВАТЕЛЯ

- Распознавание (порождение) цепочек пар символов
- Перекодирование (переход от записи на языке верхней ленты к записи на языке нижней ленты и наоборот)



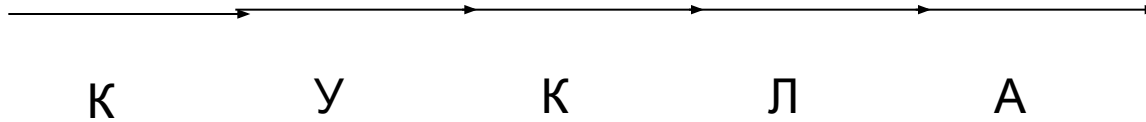
# КОНЕЧНЫЙ ПРЕОБРАЗОВАТЕЛЬ В ВИДЕ ДИАГРАММЫ



*Что записать на второй ленте?*

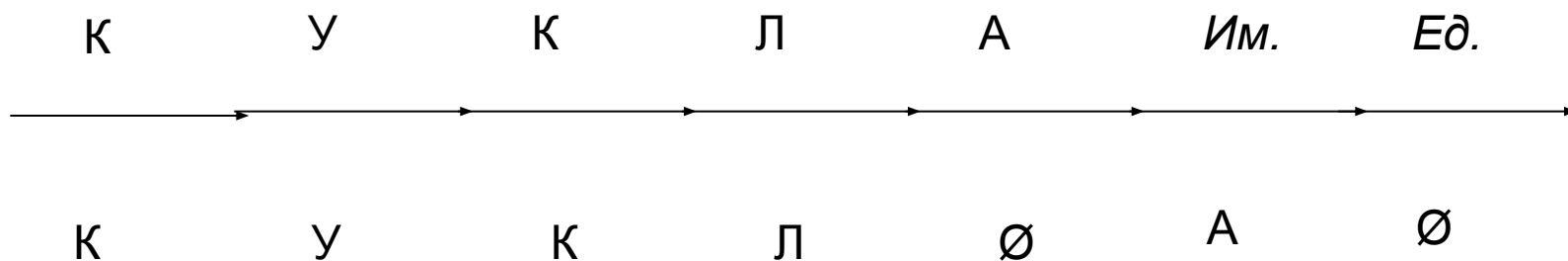
# КОНЕЧНЫЙ ПРЕОБРАЗОВАТЕЛЬ В ВИДЕ ДИАГРАММЫ

КУКЛА + Им. + Ед.



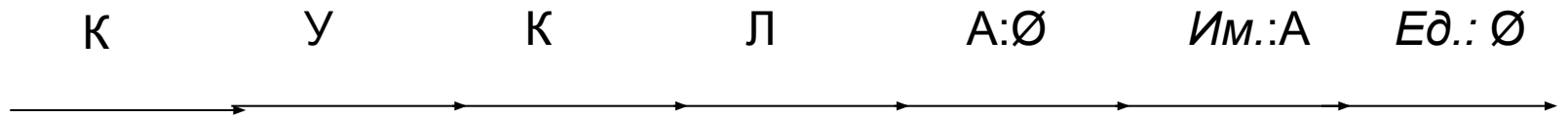
*Как расположить?*

# КОНЕЧНЫЙ ПРЕОБРАЗОВАТЕЛЬ В ВИДЕ ДИАГРАММЫ



*Нельзя ли покомпактнее?*

# КОНЕЧНЫЙ ПРЕОБРАЗОВАТЕЛЬ В ВИДЕ ДИАГРАММЫ



*Записать в виде таблицы?*

# КОНЕЧНЫЙ ПРЕОБРАЗОВАТЕЛЬ В ВИДЕ ТАБЛИЦЫ

	0	1	2	3	4	5	6
к	1	-	3	-	-	-	-
у	-	2	-	-	-	-	-
л	-	-	-	4	-	-	-
а:∅	-	-	-	-	5	-	-
<i>Им.:а</i>	-	-	-	-	-	6	-
<i>Ед.:∅</i>	-	-	-	-	-	-	F

# ФОРМАЛЬНО-ЛИНГВИСТИЧЕСКИЙ СМЫСЛ КОНЕЧНОГО ПРЕОБРАЗОВАТЕЛЯ

- Конечный автомат – язык
- Конечный преобразователь – отношение
- Язык:  $L \subseteq V^*$
- Отношение:  $R \subseteq V_B^* \times V_H^*$

# ФОРМАЛЬНО-МАТЕМАТИЧЕСКИЙ СМЫСЛ КОНЕЧНОГО ПРЕОБРАЗОВАТЕЛЯ

- Конечный автомат – регулярный язык
- Конечный преобразователь – регулярное отношение
- Язык: Результат всех возможных применений операций объединения, конкатенации, итерации символов алфавита и пустой цепочки

# ФОРМАЛЬНО-МАТЕМАТИЧЕСКИЙ СМЫСЛ КОНЕЧНОГО ПРЕОБРАЗОВАТЕЛЯ

- Конечный автомат – регулярный язык
- Конечный преобразователь –  
регулярное отношение
- Отношение: Результат произведения  
регулярных языков