Математические модели в морфологии

Введение. Теория формальных языков.

Алексей Сорокин

спецкурс, ОТИПЛ МГУ,
осенний семестр 2017–2018 учебного года

Задачи вычислительной морфологии

- Морфологический анализ (базовый случай: определение части речи);

Задачи вычислительной морфологии

- Морфологический анализ (базовый случай: определение части речи);
- Лемматизация (приведение слова к базовой форме);

Задачи вычислительной морфологии

- Морфологический анализ (базовый случай: определение части речи);
- Лемматизация (приведение слова к базовой форме);
- Морфологический синтез (построение словоформы по базовой форме и грамматической характеристике);

Задачи вычислительной морфологии

- Морфологический анализ (базовый случай: определение части речи);
- Лемматизация (приведение слова к базовой форме);
- Морфологический синтез (построение словоформы по базовой форме и грамматической характеристике);
- Автоматическое разбиение на морфемы.

Задачи вычислительной морфологии

- Морфологический анализ (базовый случай: определение части речи);
- Лемматизация (приведение слова к базовой форме);
- Морфологический синтез (построение словоформы по базовой форме и грамматической характеристике);
- Автоматическое разбиение на морфемы.
- Автоматическое построение, дополнение и извлечение парадигм.

Приложения вычислительной морфологии

Применения:

- Уточнение вероятностной модели языка (машинный перевод, классификация, исправление опечаток, ...):
  - Во многих приложениях (классификация, анализ тональности) не нужно разделять словоформы одной лексемы.

Приложения вычислительной морфологии

Применения:

- Уточнение вероятностной модели языка (машинный перевод, классификация, исправление опечаток, . . .):
    - Во многих приложениях (классификация, анализ тональности) не нужно разделять словоформы одной лексемы.
    - Данные становятся менее разреженными.

Приложения вычислительной морфологии

Применения:

- Уточнение вероятностной модели языка (машинный перевод, классификация, исправление опечаток, ...):
  - Во многих приложениях (классификация, анализ тональности) не нужно разделять словоформы одной лексемы.
  - Данные становятся менее разреженными.
- Машинный перевод (переход между поверхностным и глубинным представлением).

Приложения вычислительной морфологии

Применения:

- Уточнение вероятностной модели языка (машинный перевод, классификация, исправление опечаток, ...):
  - Во многих приложениях (классификация, анализ тональности) не нужно разделять словоформы одной лексемы.
  - Данные становятся менее разреженными.
- Машинный перевод (переход между поверхностным и глубинным представлением).
- Корпусная лингвистика (автоматическая разметка, пополнение лексических ресурсов).

Методы морфологического анализа

- Поиск по словарю.

Методы морфологического анализа

- Поиск по словарю. Недостаток “словарного” подхода:
  - Нужен очень большой словарь (в языках с развитой морфологией).

Методы морфологического анализа

- Поиск по словарю. Недостаток “словарного” подхода:
  - Нужен очень большой словарь (в языках с развитой морфологией).
  - Всё равно остаются неологизмы, производные слова.

Методы морфологического анализа

- Поиск по словарю. Недостаток “словарного” подхода:
    - Нужен очень большой словарь (в языках с развитой морфологией).
    - Всё равно остаются неологизмы, производные слова.
    - В большинстве языков развитая регулярная омонимия.

Методы морфологического анализа

- Поиск по словарю. Недостаток “словарного” подхода:
  - Нужен очень большой словарь (в языках с развитой морфологией).
  - Всё равно остаются неологизмы, производные слова.
  - В большинстве языков развитая регулярная омонимия.
- Двухуровневая морфология (конечные преобразователи).

Методы морфологического анализа

- Поиск по словарю. Недостаток "словарного" подхода:
  - Нужен очень большой словарь (в языках с развитой морфологией).
  - Всё равно остаются неологизмы, производные слова.
  - В большинстве языков развитая регулярная омонимия.
- Двухуровневая морфология (конечные преобразователи).
- Статистический анализ (на основе корпуса).

Методы морфологического анализа

- Поиск по словарю. Недостаток “словарного” подхода:
  - Нужен очень большой словарь (в языках с развитой морфологией).
  - Всё равно остаются неологизмы, производные слова.
  - В большинстве языков развитая регулярная омонимия.
- Двухуровневая морфология (конечные преобразователи).
- Статистический анализ (на основе корпуса).
- Современный подход: комбинация статистических моделей и конечных преобразователей.

Методы морфологического анализа

- Поиск по словарю. Недостаток “словарного” подхода:
  - Нужен очень большой словарь (в языках с развитой морфологией).
  - Всё равно остаются неологизмы, производные слова.
  - В большинстве языков развитая регулярная омонимия.
- Двухуровневая морфология (конечные преобразователи).
- Статистический анализ (на основе корпуса).
- Современный подход: комбинация статистических моделей и конечных преобразователей.
- Совсем современный подход: нейронные сети (с использованием вероятностных моделей и конечных преобразователей).

Регулярные выражения

Пусть зафиксирован конечный алфавит Σ.

Регулярные выражения

Пусть зафиксирован конечный алфавит Σ.

- Базовые регулярные выражения: элементы алфавита,

Регулярные выражения

Пусть зафиксирован конечный алфавит Σ.

- Базовые регулярные выражения: элементы алфавита,
- Также есть константы 0 (пустой язык) и 1 (язык, содержащий только пустое слово ε).

Регулярные выражения

Пусть зафиксирован конечный алфавит Σ.

- Базовые регулярные выражения: элементы алфавита,
- Также есть константы 0 (пустой язык) и 1 (язык, содержащий только пустое слово ε).
- Бинарные операции: $|$ (объединение) и $\cdot$ (конкатенация): $u \cdot v = uv$,

Регулярные выражения

Пусть зафиксирован конечный алфавит Σ.

- Базовые регулярные выражения: элементы алфавита,
- Также есть константы 0 (пустой язык) и 1 (язык, содержащий только пустое слово ε).
- Бинарные операции: $|$ (объединение) и $\cdot$ (конкатенация): $u \cdot v = uv$,
- Унарная операция $*$ (итерация, взять любое количество раз): L^* состоит из слов вида $u_1 \dots u_r$, где $r \in \mathbb{N}$, $u_1, \dots, u_r \in L$.

Регулярные выражения

Пусть зафиксирован конечный алфавит Σ.

- Базовые регулярные выражения: элементы алфавита,
- Также есть константы 0 (пустой язык) и 1 (язык, содержащий только пустое слово ε).
- Бинарные операции: $|$ (объединение) и $\cdot$ (конкатенация): $u \cdot v = uv$,
- Унарная операция $*$ (итерация, взять любое количество раз): L^* состоит из слов вида $u_1 \dots u_r$, где $r \in \mathbb{N}$, $u_1, \dots, u_r \in L$.
- Если α — регулярное выражение, то $L(\alpha)$ — задаваемый им язык.

Регулярные выражения

Пусть зафиксирован конечный алфавит Σ.

- Базовые регулярные выражения: элементы алфавита,
- Также есть константы 0 (пустой язык) и 1 (язык, содержащий только пустое слово ε).
- Бинарные операции: $|$ (объединение) и $\cdot$ (конкатенация): $u \cdot v = uv$,
- Унарная операция $*$ (итерация, взять любое количество раз): L^* состоит из слов вида $u_1 \ldots u_r$, где $r \in \mathbb{N}$, $u_1, \ldots, u_r \in L$.
- Если α — регулярное выражение, то $L(\alpha)$ — задаваемый им язык.
- Например, $L((a|b)^*) = \{\varepsilon, a, b, aa, ab, ba, bb, \ldots\}$.

Регулярные выражения

Пусть зафиксирован конечный алфавит Σ.

- Базовые регулярные выражения: элементы алфавита,
- Также есть константы 0 (пустой язык) и 1 (язык, содержащий только пустое слово ε).
- Бинарные операции: $|$ (объединение) и $\cdot$ (конкатенация): $u \cdot v = uv$,
- Унарная операция $*$ (итерация, взять любое количество раз): L^* состоит из слов вида $u_1 \dots u_r$, где $r \in \mathbb{N}$, $u_1, \dots, u_r \in L$.
- Если α — регулярное выражение, то $L(\alpha)$ — задаваемый им язык.
- Например, $L((a|b)^*) = \{\varepsilon, a, b, aa, ab, ba, bb, \dots\}$.
- Приоритет операций: итерация, конкатенация, объединение. При этом значок конкатенации можно опускать.

Регулярные выражения

Пусть зафиксирован конечный алфавит Σ.

- Базовые регулярные выражения: элементы алфавита,
- Также есть константы 0 (пустой язык) и 1 (язык, содержащий только пустое слово ε).
- Бинарные операции: $|$ (объединение) и $\cdot$ (конкатенация): $u \cdot v = uv$,
- Унарная операция $*$ (итерация, взять любое количество раз): L^* состоит из слов вида $u_1 \dots u_r$, где $r \in \mathbb{N}$, $u_1, \dots, u_r \in L$.
- Если α — регулярное выражение, то $L(\alpha)$ — задаваемый им язык.
- Например, $L((a|b)^*) = \{\varepsilon, a, b, aa, ab, ba, bb, \dots\}$.
- Приоритет операций: итерация, конкатенация, объединение. При этом значок конкатенации можно опускать.
- Сокращения $\alpha^+ = \alpha\alpha^*$ (положительная итерация), $\alpha? = (\alpha|1)$ (опциональное вхождение α).

Регулярные выражения

Пусть зафиксирован конечный алфавит Σ.

- Базовые регулярные выражения: элементы алфавита,
- Также есть константы 0 (пустой язык) и 1 (язык, содержащий только пустое слово ε).
- Бинарные операции: $|$ (объединение) и $\cdot$ (конкатенация): $u \cdot v = uv$,
- Унарная операция $*$ (итерация, взять любое количество раз): L^* состоит из слов вида $u_1 \dots u_r$, где $r \in \mathbb{N}$, $u_1, \dots, u_r \in L$.
- Если α — регулярное выражение, то $L(\alpha)$ — задаваемый им язык.
- Например, $L((a|b)^*) = \{\varepsilon, a, b, aa, ab, ba, bb, \dots\}$.
- Приоритет операций: итерация, конкатенация, объединение. При этом значок конкатенации можно опускать.
- Сокращения $\alpha^+ = \alpha\alpha^*$ (положительная итерация), $\alpha? = (\alpha|1)$ (опциональное вхождение α).

Язык регулярный, если он задаётся регулярным выражением.

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,
- Слова в алфавите $\{a, b, c\}$, где предпоследняя буква — b: $(a|b|c)^* b(a|b|c)$.

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,
- Слова в алфавите $\{a, b, c\}$, где предпоследняя буква — b: $(a|b|c)^* b(a|b|c)$.
- Слова в алфавите $\{a, b, c\}$, содержащие ровно 2 буквы a: $(b|c)^* a(b|c)^* a(b|c)^*$.

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,
- Слова в алфавите $\{a, b, c\}$, где предпоследняя буква — b: $(a|b|c)^* b(a|b|c)$.
- Слова в алфавите $\{a, b, c\}$, содержащие ровно 2 буквы a: $(b|c)^* a(b|c)^* a(b|c)^*$.
- Слова нечётной длины в алфавите $\{a, b\}$: $((a|b)(a|b))^*(a|b)$.

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,
- Слова в алфавите $\{a, b, c\}$, где предпоследняя буква — b: $(a|b|c)^* b(a|b|c)$.
- Слова в алфавите $\{a, b, c\}$, содержащие ровно 2 буквы a: $(b|c)^* a(b|c)^* a(b|c)^*$.
- Слова нечётной длины в алфавите $\{a, b\}$: $((a|b)(a|b))^*(a|b)$.
- Слова в алфавите $\{a, b, c\}$, содержащие чётное число букв a:

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,
- Слова в алфавите $\{a, b, c\}$, где предпоследняя буква — b: $(a|b|c)^* b(a|b|c)$.
- Слова в алфавите $\{a, b, c\}$, содержащие ровно 2 буквы a: $(b|c)^* a(b|c)^* a(b|c)^*$.
- Слова нечётной длины в алфавите $\{a, b\}$: $((a|b)(a|b))^*(a|b)$.
- Слова в алфавите $\{a, b, c\}$, содержащие чётное число букв a: $((b|c)^* a(b|c)^* a)(b|c)^*$

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,
- Слова в алфавите $\{a, b, c\}$, где предпоследняя буква — b: $(a|b|c)^* b(a|b|c)$.
- Слова в алфавите $\{a, b, c\}$, содержащие ровно 2 буквы a: $(b|c)^* a(b|c)^* a(b|c)^*$.
- Слова нечётной длины в алфавите $\{a, b\}$: $((a|b)(a|b))^*(a|b)$.
- Слова в алфавите $\{a, b, c\}$, содержащие чётное число букв a: $((b|c)^* a(b|c)^* a)(b|c)^*$.
- Слова в алфавите $\{a, b, c\}$, где перед a идёт только b:

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,
- Слова в алфавите $\{a, b, c\}$, где предпоследняя буква — b: $(a|b|c)^* b(a|b|c)$.
- Слова в алфавите $\{a, b, c\}$, содержащие ровно 2 буквы a: $(b|c)^* a(b|c)^* a(b|c)^*$.
- Слова нечётной длины в алфавите $\{a, b\}$: $((a|b)(a|b))^*(a|b)$.
- Слова в алфавите $\{a, b, c\}$, содержащие чётное число букв a: $((b|c)^* a(b|c)^* a)(b|c)^*$.
- Слова в алфавите $\{a, b, c\}$, где перед a идёт только b: $((b|c)^* ba)^*(b|c)^*$

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,
- Слова в алфавите $\{a, b, c\}$, где предпоследняя буква — b: $(a|b|c)^*b(a|b|c)$.
- Слова в алфавите $\{a, b, c\}$, содержащие ровно 2 буквы a: $(b|c)^*a(b|c)^*a(b|c)^*$.
- Слова нечётной длины в алфавите $\{a, b\}$: $((a|b)(a|b))^*(a|b)$.
- Слова в алфавите $\{a, b, c\}$, содержащие чётное число букв a: $((b|c)^*a(b|c)^*a)(b|c)^*$.
- Слова в алфавите $\{a, b, c\}$, где перед a идёт только b: $((b|c)^*ba)^*(b|c)^*$.
- Непустые слова в алфавите $\{a, b\}$, в которых одинаковые буквы не идут подряд:

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,
- Слова в алфавите $\{a, b, c\}$, где предпоследняя буква — b: $(a|b|c)^* b(a|b|c)$.
- Слова в алфавите $\{a, b, c\}$, содержащие ровно 2 буквы a: $(b|c)^* a(b|c)^* a(b|c)^*$.
- Слова нечётной длины в алфавите $\{a, b\}$: $((a|b)(a|b))^*(a|b)$.
- Слова в алфавите $\{a, b, c\}$, содержащие чётное число букв a: $((b|c)^* a(b|c)^* a)(b|c)^*$.
- Слова в алфавите $\{a, b, c\}$, где перед a идёт только b: $((b|c)^* ba)^*(b|c)^*$.
- Непустые слова в алфавите $\{a, b\}$, в которых одинаковые буквы не идут подряд: $(a(ba)^*(b|1))|(b(ab)^*(a|1))$

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,
- Слова в алфавите $\{a, b, c\}$, где предпоследняя буква — b: $(a|b|c)^*b(a|b|c)$.
- Слова в алфавите $\{a, b, c\}$, содержащие ровно 2 буквы a: $(b|c)^*a(b|c)^*a(b|c)^*$.
- Слова нечётной длины в алфавите $\{a, b\}$: $((a|b)(a|b))^*(a|b)$.
- Слова в алфавите $\{a, b, c\}$, содержащие чётное число букв a: $((b|c)^*a(b|c)^*a)(b|c)^*$.
- Слова в алфавите $\{a, b, c\}$, где перед a идёт только b: $((b|c)^*ba)^*(b|c)^*$.
- Непустые слова в алфавите $\{a, b\}$, в которых одинаковые буквы не идут подряд: $(a(ba)^*(b|1))|(b(ab)^*(a|1))$.
- Слова в алфавите $\{a, b, c\}$, в которых одинаковые буквы не идут подряд:

Примеры регулярных языков

- Все слова в алфавите $\{a, b\}$: $(a|b)^*$,
- Слова в алфавите $\{a, b, c\}$, где предпоследняя буква — b: $(a|b|c)^* b(a|b|c)$.
- Слова в алфавите $\{a, b, c\}$, содержащие ровно 2 буквы a: $(b|c)^* a(b|c)^* a(b|c)^*$.
- Слова нечётной длины в алфавите $\{a, b\}$: $((a|b)(a|b))^*(a|b)$.
- Слова в алфавите $\{a, b, c\}$, содержащие чётное число букв a: $((b|c)^* a(b|c)^* a)(b|c)^*$.
- Слова в алфавите $\{a, b, c\}$, где перед a идёт только b: $((b|c)^* ba)^*(b|c)^*$.
- Непустые слова в алфавите $\{a, b\}$, в которых одинаковые буквы не идут подряд: $(a(ba)^*(b|1))|(b(ab)^*(a|1))$.
- Слова в алфавите $\{a, b, c\}$, в которых одинаковые буквы не идут подряд: $(H|1)(cH)^*(1|c)$, где H — ответ на предыдущий пункт.

Примеры регулярных выражений

Пусть $\Sigma = \{C, V, \overline{V}, -\}$ (согласный, безударный гласный, ударный гласный, слогораздел).

- Корректное разбиение на слоги:
  - В каждом слоге ровно одна гласная: $C^*(V|\overline{V})C^*$.
  - Ровно один слог ударный.

Примеры регулярных выражений

Пусть $\Sigma = \{C, V, \overline{V}, -\}$ (согласный, безударный гласный, ударный гласный, слогораздел).

- Корректное разбиение на слоги:
  - В каждом слоге ровно одна гласная: $C^*(V|\overline{V})C^*$.
  - Ровно один слог ударный.
  - Пусть X — ударный слог, Y – безударный, тогда искомое выражение $(Y-)^*[(X-(Y-)^*Y)|X]$.

Примеры регулярных выражений

Пусть $\Sigma = \{C, V, \overline{V}, -\}$ (согласный, безударный гласный, ударный гласный, слогораздел).

- Корректное разбиение на слоги:
  - В каждом слоге ровно одна гласная: $C^*(V|\overline{V})C^*$.
  - Ровно один слог ударный.
  - Пусть X — ударный слог, Y – безударный, тогда искомое выражение $(Y-)^*[(X-(Y-)^*Y)|X]$.
  - Эквивалентно $(Y-)^*X(-Y)^* = (C^*VC^*-)^*C^*\overline{V}C^*(-C^*VC^*)^*$.

Примеры регулярных выражений

Пусть $\Sigma = \{C, V, \overline{V}, -\}$ (согласный, безударный гласный, ударный гласный, слогораздел).

- Корректное разбиение на слоги:
  - В каждом слоге ровно одна гласная: $C^*(V|\overline{V})C^*$.
  - Ровно один слог ударный.
  - Пусть X — ударный слог, Y – безударный, тогда искомое выражение $(Y-)^*[(X-(Y-)^*Y)|X]$.
  - Эквивалентно $(Y-)^*X(-Y)^* = (C^*VC^*-)^*C^*\overline{V}C^*(-C^*VC^*)^*$.
- Разбиение слова на слоги, содержащее ровно 1 открытый слог (ударность не учитывается).

Примеры регулярных выражений

Пусть $\Sigma = \{C, V, \overline{V}, -\}$ (согласный, безударный гласный, ударный гласный, слогораздел).

- Корректное разбиение на слоги:
  - В каждом слоге ровно одна гласная: $C^*(V|\overline{V})C^*$.
  - Ровно один слог ударный.
  - Пусть X — ударный слог, Y – безударный, тогда искомое выражение $(Y-)^*[(X-(Y-)^*Y)|X]$.
  - Эквивалентно $(Y-)^*X(-Y)^* = (C^*VC^*-)^*C^*\overline{V}C^*(-C^*VC^*)^*$.
- Разбиение слова на слоги, содержащее ровно 1 открытый слог (ударность не учитывается). $(C^*VC^+-)^*(C^*V)(-C^*VC^+)^*$

Примеры регулярных выражений

Пусть $\Sigma = \{C, V, \overline{V}, -\}$ (согласный, безударный гласный, ударный гласный, слогораздел).

- Корректное разбиение на слоги:
  - В каждом слоге ровно одна гласная: $C^*(V|\overline{V})C^*$.
  - Ровно один слог ударный.
  - Пусть X — ударный слог, Y – безударный, тогда искомое выражение $(Y-)^*[(X-(Y-)^*Y)|X]$.
  - Эквивалентно $(Y-)^*X(-Y)^* = (C^*VC^*-)^*C^*\overline{V}C^*(-C^*VC^*)^*$.
- Разбиение слова на слоги, содержащее ровно 1 открытый слог (ударность не учитывается). $(C^*VC^+-)^*(C^*V)(-C^*VC^+)^*$
- “Гармония гласных” (гласные типа V_1 и V_2 не встречаются вместе): $(C|V)^*(V_1(C|V_1|V)^*|V_2(C|V_2|V)^*)$

Примеры регулярных выражений

- Множественное число существительного в английском:
  - -*es* после шипящих (s, x, z, ch, sh, zh).
  - -*y* после согласных перед -*s* переходит в -*ie*.

Примеры регулярных выражений

- Множественное число существительного в английском:
  - -*es* после шипящих (s, x, z, ch, sh, zh).
  - -*y* после согласных перед -*s* переходит в -*ie*.
- Удобней разбирать $witches = witche{+}s$, $enemies = enemie{+}s$.

Примеры регулярных выражений

- Множественное число существительного в английском:
  - -*es* после шипящих (s, x, z, ch, sh, zh).
  - -*y* после согласных перед -*s* переходит в -*ie*.
- Удобней разбирать $witches = witche{+}s$, $enemies = enemie{+}s$.
- Искомое выражение $X = Ys$, где Y — выражение для основы.

Примеры регулярных выражений

- Множественное число существительного в английском:
  - -*es* после шипящих (s, x, z, ch, sh, zh).
  - -*y* после согласных перед -*s* переходит в -*ie*.
- Удобней разбирать $witches = witche{+}s$, $enemies = enemie{+}s$.
- Искомое выражение $X = Ys$, где Y — выражение для основы.
- Основа — всё, что не кончается на s, x, z, ch, sh, zh, $\boldsymbol{C}y$.

Примеры регулярных выражений

- Множественное число существительного в английском:
  - -*es* после шипящих (s, x, z, ch, sh, zh).
  - -*y* после согласных перед -*s* переходит в -*ie*.
- Удобней разбирать $witches = witche{+}s$, $enemies = enemie{+}s$.
- Искомое выражение $X = Ys$, где Y — выражение для основы.
- Основа — всё, что не кончается на s, x, z, ch, sh, zh, $\mathbf{C}y$.
- Хочется задать отрицание условия...

Примеры регулярных выражений

- Множественное число существительного в английском:
  - -*es* после шипящих (*s*, *x*, *z*, *ch*, *sh*, *zh*).
  - -*y* после согласных перед -*s* переходит в -*ie*.
- Удобней разбирать $witches = witche{+}s$, $enemies = enemie{+}s$.
- Искомое выражение $X = Ys$, где Y — выражение для основы.
- Основа — всё, что не кончается на *s*, *x*, *z*, *ch*, *sh*, *zh*, $\boldsymbol{C}y$.
- Хочется задать отрицание условия...
- Симулируется через перечисление.

Примеры регулярных выражений

- Корректная основа:

Примеры регулярных выражений

- Корректная основа:
  - Заканчивается на гласный, не равный y: $(C|V)^*(a|e|i|o|u)$.

Примеры регулярных выражений

- Корректная основа:
    - Заканчивается на гласный, не равный y: $(C|V)^*(a|e|i|o|u)$.
    - Заканчивается на гласный+y: $(C|V)^*Vy$.

Примеры регулярных выражений

- Корректная основа:
  - Заканчивается на гласный, не равный y: $(C|V)^*(a|e|i|o|u)$.
  - Заканчивается на гласный+y: $(C|V)^*Vy$.
  - Содержит гласный и заканчивается на согласный, но не на s, x, z, h ($\mathbf{C}'$ — полный список таких согласных): $(C|V)^*V(C|V)^*C'$

Примеры регулярных выражений

- Корректная основа:
  - Заканчивается на гласный, не равный y: $(C|V)^*(a|e|i|o|u)$.
  - Заканчивается на гласный+y: $(C|V)^*Vy$.
  - Содержит гласный и заканчивается на согласный, но не на s, x, z, h ($\mathbf{C}'$ — полный список таких согласных): $(C|V)^*V(C|V)^*C'$
  - Содержит гласный и заканчивается на h или $\mathbf{C}''h$, где $\mathbf{C}''$ обозначает любой согласный, кроме s, c, h: $(C|V)^*V((C|V)^*C'')?h$

Примеры регулярных выражений

- Корректная основа:
  - Заканчивается на гласный, не равный y: $(C|V)^*(a|e|i|o|u)$.
  - Заканчивается на гласный+y: $(C|V)^*Vy$.
  - Содержит гласный и заканчивается на согласный, но не на s, x, z, h ($\mathbf{C}'$ — полный список таких согласных): $(C|V)^*V(C|V)^*C'$
  - Содержит гласный и заканчивается на h или $\mathbf{C}''h$, где $\mathbf{C}''$ обозначает любой согласный, кроме s, c, h: $(C|V)^*V((C|V)^*C'')?h$
- Всё вместе: $(C|V)^*((a|e|i|o|u|Vy) \mid V(h|(C|V)^*(C'|C''h))s$.

Определение конечного автомата

Пусть Σ — конечный алфавит.

Определение конечного автомата

Конечный автомат: кортеж $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$, где

Определение конечного автомата

Пусть Σ — конечный алфавит.

Определение конечного автомата

Конечный автомат: кортеж $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$, где

- Q — конечное множество состояний

Определение конечного автомата

Пусть Σ — конечный алфавит.

Определение конечного автомата

Конечный автомат: кортеж $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$, где

- Q — конечное множество состояний
- $\Delta \subseteq Q \times (\Sigma \cup \{\varepsilon\}) \times Q$ — конечное множество переходов

Определение конечного автомата

Пусть Σ — конечный алфавит.

Определение конечного автомата

Конечный автомат: кортеж $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$, где

- Q — конечное множество состояний
- $\Delta \subseteq Q \times (\Sigma \cup \{\varepsilon\}) \times Q$ — конечное множество переходов
- $q_0 \in Q$ — стартовое состояние
- $F \subseteq Q$ — завершающие состояния.

Определение конечного автомата

Пусть Σ — конечный алфавит.

Определение конечного автомата

Конечный автомат: кортеж $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$, где

- Q — конечное множество состояний
- $\Delta \subseteq Q \times (\Sigma \cup \{\varepsilon\}) \times Q$ — конечное множество переходов
- $q_0 \in Q$ — стартовое состояние
- $F \subseteq Q$ — завершающие состояния.

Неформально, конечный автомат — граф с выделенными стартовой и заверщающими вершинами, рёбра которого помечены символами алфавита или пустым словом.

Определение конечного автомата

Пусть Σ — конечный алфавит.

Определение конечного автомата

Конечный автомат: кортеж $M = \langle Q, \Sigma, \Delta, q_0, F \rangle$, где

- Q — конечное множество состояний
- $\Delta \subseteq Q \times (\Sigma \cup \{\varepsilon\}) \times Q$ — конечное множество переходов
- $q_0 \in Q$ — стартовое состояние
- $F \subseteq Q$ — завершающие состояния.

Неформально, конечный автомат — граф с выделенными стартовой и заверщающими вершинами, рёбра которого помечены символами алфавита или пустым словом.

$L(M)$ — метки путей из начального состояния в завершающие.

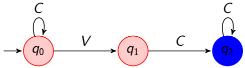
Язык автоматный — задаётся некоторым конечным автоматом.

Примеры конечных автоматов

- Закрытый слог

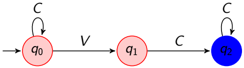
Примеры конечных автоматов

- Закрытый слог



Примеры конечных автоматов

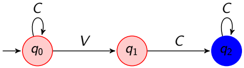
- Закрытый слог



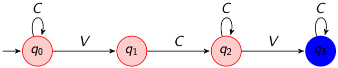
- Слово с 2 гласными, разделёнными хотя бы одним согласным:

Примеры конечных автоматов

- Закрытый слог

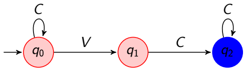


- Слово с 2 гласными, разделёнными хотя бы одним согласным:

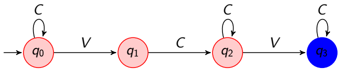


Примеры конечных автоматов

- Закрытый слог



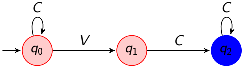
- Слово с 2 гласными, разделёнными хотя бы одним согласным:



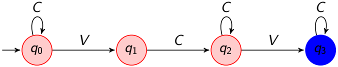
- Слогоделение ровно с одним открытым слогом:

Примеры конечных автоматов

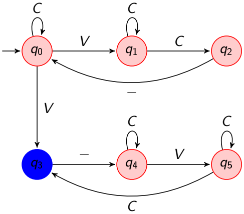
- Закрытый слог



- Слово с 2 гласными, разделёнными хотя бы одним согласным:



- Слогоделение ровно с одним открытым слогом:

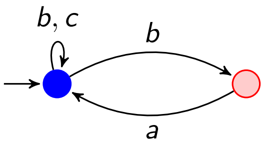


Конечные автоматы: примеры

- Каждой a непосредственно предшествует b, алфавит a, b, c.

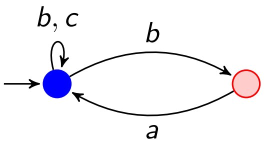
Конечные автоматы: примеры

- Каждой a непосредственно предшествует b, алфавит a, b, c.



Конечные автоматы: примеры

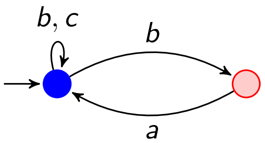
- Каждой a непосредственно предшествует b, алфавит a, b, c.



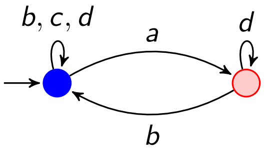
- Справа от каждой a есть парная ей b, между парными буквами нет a, c, алфавит a, b, c, d.

Конечные автоматы: примеры

- Каждой a непосредственно предшествует b, алфавит a, b, c.

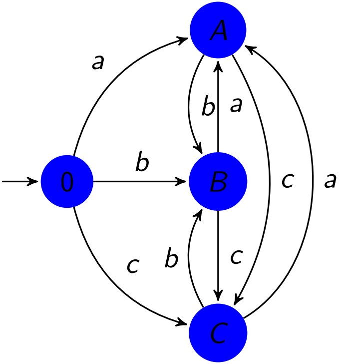


- Справа от каждой a есть парная ей b, между парными буквами нет a, c, алфавит a, b, c, d.



Конечные автоматы: примеры

Нет повторяющихся букв, алфавит a, b, c. Состояния соответствуют буквам:



Конечные автоматы: примеры

- Формы множественного числа представимы в виде $\mathrm{stem} + \mathrm{s}$, где

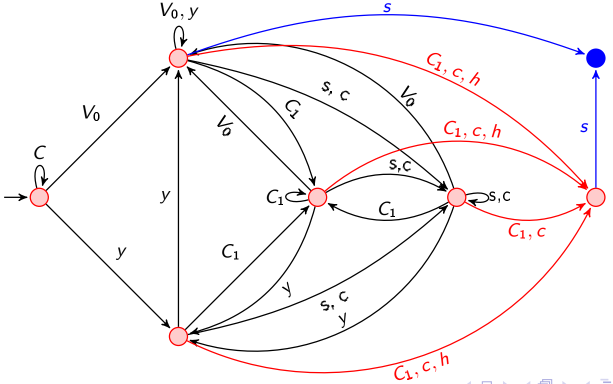
Конечные автоматы: примеры

- Формы множественного числа представимы в виде $\mathrm{stem} + \mathsf{s}$, где
- stem обязательно содержит гласную и не кончается на:
  - *-s*, *-x*, *-z*, *-sh*, *-ch*, *-zh* (шипящие).
  - **C***y*.

Конечные автоматы: примеры

- Формы множественного числа представимы в виде $\mathrm{stem} + s$, где
- stem обязательно содержит гласную и не кончается на:
  - *-s, -x, -z, -sh, -ch, -zh* (шипящие).
  - **C***y*.
- Автомат для основ

  $(C_0 = C - \{s, x, z, c, h\}, C_1 = C_0 \cup \{s, x, z\})$:



Автоматы с однобуквенными переходами

Теорема

Каждый автоматный язык распознаётся автоматом с однобуквенными переходами.

Автоматы с однобуквенными переходами

Теорема

Каждый автоматный язык распознаётся автоматом с однобуквенными переходами.

Схема доказательства

- Сделать завершающими все состояния, из которых достижимо по ε (возможно, за несколько шагов) другое завершающее.
- Добавить все рёбра вида $\langle q_1, a\rangle \rightarrow q_2$,для которых существуют состояние q_3, такое что есть ребро $(\langle q_3, a\rangle \rightarrow q_2) \in \Delta$ и ε-путь из q_1 в q_3.
- Удалить ε-рёбра.

Детерминированные конечные автоматы

Детерминированные конечные автоматы

Определение

Автомат с однобуквенными переходами — детерминированный, если ни из какого состояния не выходит двух рёбер, помеченных одинаковыми буквами.

Теорема

Каждый автоматный язык распознаётся детерминированным автоматом.

Детерминированные конечные автоматы

Определение

Автомат с однобуквенными переходами — детерминированный, если ни из какого состояния не выходит двух рёбер, помеченных одинаковыми буквами.

Теорема

Каждый автоматный язык распознаётся детерминированным автоматом.

Схема доказательства

- Новые состояния — множества старых состояний.

Детерминированные конечные автоматы

Определение

Автомат с однобуквенными переходами — детерминированный, если ни из какого состояния не выходит двух рёбер, помеченных одинаковыми буквами.

Теорема

Каждый автоматный язык распознаётся детерминированным автоматом.

Схема доказательства

- Новые состояния — множества старых состояний.
- Ребро, помеченное a, ведёт из Q_1 в Q_2, если Q_2 содержит в точности состояния,достижимые из Q_1 по a.

Детерминированные конечные автоматы

Определение

Автомат с однобуквенными переходами — детерминированный, если ни из какого состояния не выходит двух рёбер, помеченных одинаковыми буквами.

Теорема

Каждый автоматный язык распознаётся детерминированным автоматом.

Схема доказательства

- Новые состояния — множества старых состояний.
- Ребро, помеченное a, ведёт из Q_1 в Q_2, если Q_2 содержит в точности состояния,достижимые из Q_1 по a.
- Стартовое множество состояний $Q_0 = \{q_0\}$.

Детерминированные конечные автоматы

Определение

Автомат с однобуквенными переходами — детерминированный, если ни из какого состояния не выходит двух рёбер, помеченных одинаковыми буквами.

Теорема

Каждый автоматный язык распознаётся детерминированным автоматом.

Схема доказательства

- Новые состояния — множества старых состояний.
- Ребро, помеченное a, ведёт из Q_1 в Q_2, если Q_2 содержит в точности состояния,достижимые из Q_1 по a.
- Стартовое множество состояний $Q_0 = \{q_0\}$.
- Завершающие состояния: множества, содержащие хотя бы одно завершающее.