

## Вопрос 1

Выполнен

Баллов: 1,00 из 1,00

Выберите регулярное выражение, которое приведёт к указанной токенизации.

Входное предложение:

"Пожалуйста, отправляйте свои заявки на адрес электронной почты inbox@example.com или по телефону 123456."

Ожидаемая токенизация:

['Пожалуйста', ',', 'отправляйте', 'свои', 'заявки', 'на', 'адрес', 'электронной', 'почты', 'inbox@example.com', 'или', 'по', 'телефону', '123456', '.']

Используйте синтаксис регулярных выражений Python версии >3.5 (<https://docs.python.org/3/library/re.html>).

Токенизация будет выполняться с помощью следующего фрагмента кода:

```
import re
txt = "Пожалуйста, отправляйте свои заявки на адрес электронной почты inbox@example.com или по телефону 123456."
tokenize_regex = re.compile(<your answer>, re.I)
tokens = tokenize_regex.findall(txt)
print(tokens)
```

Выберите один ответ:

- ☐ a. `r'\S+'`
- ☐ b. `r'\w+'`
- ☐ c. `r'\w+|\S'`
- ☐ d. `r'\w+'`
- ☐ e. `r'((?:\w+@\w+\.\w+)|\S|\w+)'`
- ☐ f. `r' '`
- ☐ g. `r'(\w+|(?:\w+@\w+\.\w+)|\S)'`
- ☒ h. `r'((?:\w+@\w+\.\w+)|\w+|\S)'`

Ваш ответ верный.

**Вопрос 2**

Выполнен

Баллов: 1,00 из 1,00

С помощью каких методов обычно решают задачу извлечения именованных сущностей?

Выберите один или несколько вариантов ответа.

Выберите один или несколько ответов:

- ☒ a. Искусственные нейросети
- ☒ b. Регулярные выражения и системы правил (rule-based)
- ☒ c. Вероятностные модели: Скрытые марковские модели (HMM, Hidden Markov Models) и Условные случайные поля (CRF, Conditional Random Field)
- ☒ d. Поиск по словарю

Ваш ответ верный.

**Вопрос Инфо**

### Вопрос 3

Выполнен

Баллов: 1,00 из 1,00

Выберите все задачи лингвистического анализа, которые можно решать в два этапа:

1. выделение всех кандидатов (это могут быть слова, токены, сущности состоящие из нескольких слов)
2. обучение классификатора, принимающего **пару** кандидатов, и предсказывающего класс **отношения между ними**.

Попробуйте предположить и выбрать **все** варианты, где такая схема могла бы применяться, даже если в лекциях непосредственно такой вариант не упоминался.

Выберите один или несколько ответов:

- ☒ a. Синтаксический анализ (построение дерева, описывающего структуру фраз, грамматическое подчинение и согласование слов)
- ☒ b. Разрешение анафорических связей (англ. anaphora resolution: построение графа, в котором рёбра связывают референта и ссылающиеся на него местоимения или словосочетания; референт - некоторый объект, который может именоваться в тексте разными способами)
- ☒ c. Извлечение отношений между сущностями внутри предложения (построение графа, описывающего отношения между упоминаемыми в предложении объектами)
- ☐ d. Извлечение именованных сущностей (англ. Named Entity Recognition: классификация токенов, классы соответствуют видам сущностей)
- ☒ e. Семантический анализ (построение графа, описывающего высокоуровневую ситуацию - кто какое действие с чем выполняет, когда, почему и т.п.; в вершинах такого графа обычно находятся отдельные слова или фразы)
- ☐ f. Морфологический анализ (предсказание морфологических характеристик отдельных токенов)
- ☐ g. POS-теггинг (снятие частеречной неоднозначности, т.е. выбор наиболее правдоподобного сочетания частей речи для токенов в предложении)

Ваш ответ **верный**.

[◀ 1.3 Особенности обработки естественных языков](#)

Перейти на...

[1.5 В общих чертах: Извлечение признаков ▶](#)