**Билет 1**

1. Describe the communicational system between Dave and HAL. Describe distinct categories of the knowledge of language. Опишите систему взаимодействия Dave – HAL. Опишите выделенные категории знаний языка.
2. Describe different tagsets for English. What two algorithms are used for automatically assigning part-of-speech tagging. Describe stochastic part of speech tagging algorithm in details. Опишите раздличные наборы тегов частей речи английского языка. Опишите два основных алгоритма для автоматического определения части речи. Детально опишите стохастический алгоритм.
3. Write a regular expression to find words **String** and **string** in the following sentence. Напишите регулярное выражение для поиска слов **String** и **string** в тексте.

I did not see what String he chose from the list of strings.

**Билет 2**

1. What are regular expressions? Describe their meaning and goal. What symbols are commonly use when we need to choose between several characters or digits? Что такое регулярные выражения? Опишите, что они собой представляют и их цель. Какие символы обычно используются, когда нам необходимо выбрать буквы и цифры из определенного диапазона?
2. Describe different tagsets for English. What two algorithms are used for automatically assigning part-of-speech tagging. Describe a rules-based part of speech tagging algorithm in details. Опишите раздличные наборы тегов частей речи английского языка. Опишите два основных алгоритма для автоматического определения части речи. Детально опишите алгоритм, основанный на знании правил.
3. Correct the following substitution or transposition mistakes to make right words. Исправьте ошибки замены и перестановки в следующих словах:

sutisfaction, bfttle, anddlesc, papricots, ruots, fields.

**Билет 3**

1. What are regular expressions? Describe their meaning and goal. What symbol is used when we need to choose any symbol or digit in a word? Что такое регулярные выражения? Опишите, что они собой представляют и их цель. Какой символ используется, когда нам необходимо выбрать любой символ или цифру в слове?
2. Why do we use smoothing in N-gram model? Write formulas for finding unsmoothed frequency $C$ and probability $P(w\_{n}|w\_{n-1})$ for bigrams. Write formulas for finding smoothed $C^{\*}$ and $P(w\_{n}|w\_{n-1})^{\*}$ for bigrams using **Add-one method**. Почему мы используем сглаживание в модели N-gram? Напишите формулы для нахождения несглаженных частоты $C$ и вероятности $P(w\_{n}|w\_{n-1})$ для биграмм. Напишите формулы для нахождения сглаженных частоты $C^{\*}$ и вероятности $P(w\_{n}|w\_{n-1})^{\*}$ для биграмм, используя метод сглаживания **Добавление единицы**.
3. Correct the following insertion or deletion mistakes to make right words. Исправьте ошибки вставки и удаления в следующих словах:

conatulations, eorganizations, tbles, trouobles, feindings, kep, stuctur.

**Билет 4**

1. What are morphological rules and morphological parsing? What are main morphemes of words? Что такое морфологические правила и морфологический разбор? Какие основные морфемы слов вы знаете?
2. Why do we use smoothing in N-gram model? Write formulas for finding unsmoothed frequency $C$ and probability $P(w\_{n}|w\_{n-1})$ for bigrams. Write formulas for finding smoothed $C^{\*}$ and $P(w\_{n}|w\_{n-1})^{\*}$ for bigrams using **Add-one method**. Почему мы используем сглаживание в модели N-gram? Напишите формулы для нахождения несглаженных частоты $C$ и вероятности $P(w\_{n}|w\_{n-1})$ для биграмм. Напишите формулы для нахождения сглаженных частоты $C^{\*}$ и вероятности $P(w\_{n}|w\_{n-1})^{\*}$ для биграмм, используя метод сглаживания **Добавление единицы**.
3. Write a regular expression to find these words in the text: cdcdcdcd, cccccccc, dddddddd. Напишите регулярное выражения на поиска заданных слов в тексте: cdcdcdcd, cccccccc, dddddddd

These codes were used in the encryption processes: cdcdcdcd, cccccccc, dddddddd.

**Билет 5**

1. What is morphological parsing? What is inflectional morphology? Describe its features. Что такое морфологический разбор? Что такое флективная морфология. Опишите ее особенности.
2. Why do we use smoothing in N-gram model? Write formulas for finding unsmoothed frequency $C$ and probability $P(C)$ for unigrams. Write formulas for finding smoothed $C^{\*}$ and $P(C)^{\*}$ for unigrams using **Add-one method**. Почему мы используем сглаживание в модели N-gram? Напишите формулы для нахождения несглаженных частоты $C$ и вероятности $P(C)$ для униграмм. Напишите формулы для нахождения сглаженных частоты $C^{\*}$ и вероятности $P(C)^{\*}$ для униграмм, используя метод сглаживания **Добавление единицы**.
3. Write a regular expression to find these words in the text: begin, began, begun. Напишите регулярное выражения на поиска заданных слов в тексте: begin, began, begun.

I begin to do this exercise. They began to work on this project. He has begun to conduct his research in this specific scientific field.

**Билет 6**

1. What is morphological parsing? What is derivational morphology? Describe its features? Что такое морфологический разбор? Что такое дерифационная морфология? Опишите ее особенности.
2. What are N-grams? Write a formula for counting a probability of a given sentence in a text. Что такое N-граммы? Напишите формулу подсчета вероятности появления определенной строки в тексте. Вы можете использовать биграммы.
3. Write a regular expression to find all uppercase letters in the text. Напишите регулярное выражение для поиска всех слов в верхнем регистре.

With a view 100 times bigger than that of NASA’s Hubble Space Telescope, WFIRST will aid researchers in their efforts to unravel the secrets of dark energy and dark matter, and explore the evolution of the cosmos.

**Билет 7**

1. What is the lexicon and morphotactics? Describe a finite-state automaton for English nominal inflection. Что такое лексикон и морфотактика? Опишите автомат конечных состояний для инфлекционной морфологии английского языка.
2. How is Bayesian formula used for describing the correction of errors in words? Estimate the probability of each correction $P(c)$ describing the formula $P\left(c\right)= \frac{C\left(c\right)+0.5}{N+0.5V}$. Как используется формула Байеса для описания исправления ошибок в словах? Оцените вероятность $P(c)$, описывая каждый компонент формулы $P\left(c\right)= \frac{C\left(c\right)+0.5}{N+0.5V}$
3. Write a regular expression to find words **String** and **string** in the following sentence. Напишите регулярное выражение для поиска слов **String** и **string** в тексте.

I did not see what String he chose from the list of strings.

**Билет 8**

1. What is the lexicon and morphotactics? Describe a finite-state automaton for English derivational morphology. Что такое лексикон и морфотактика? Опишите автомат конечных состояний для деривационной морфологии английского языка.
2. How is Bayesian formula used for describing the correction of errors in words? Write a general formula Of Bayes and its specified version. Как используется формула Байеса для описания исправления ошибок в словах? Напишите общую формулу Байеса и ее частный случай.
3. Compute the probability of a sentence using a table with probabilities of bigrams. Посчитайте вероятность предложения, используя таблицу вероятностей биграмм.

We have lost the mighty king who protected our lands.

|  |  |  |
| --- | --- | --- |
| <s> We 0.3We have 0.09have lost 0.04lost the 0.045 | the mighty 0.048mighty king 0.03king who 0.024 | who protected 0.032protected our 0.052our lands 0.078 |

**Билет 9**

1. Describe two-level morphology for finite-state transducers. What are four ways of thinking about transducers? Опишите двухуровневую морфологию преобразователей конечных состояний. Какие четыре вида преобразователей существуют?
2. What spelling error patterns do you know? Describe them and give examples for each category. Какие ошибки написания слов вы знаете? Опишите их и привидите примеры для каждой категории.
3. Count down the unsmoothed and add-one-smoothed probabilities of finding the following bigram in the text. Посчитать несглаженную и сглаженную методом добавления единицы вероятности нахождения следующей биграммы с тексте

$P(have|I)$; биграмма $(I have)$ найдена в тексте 30 раз, слово I найдено в тексте 420 раз, размер словаря V составляет 300.

**Билет 10**

1. Describe three-level morphology for finite-state transducers. Write orthographical rules like E insertion, E deletion and so on. Опишите трехуровневую морфологию преобразователей конечных состояний. Напишите орфографичекие правила по типу Вставка Е, удаление Е и другие.
2. What is optical character recognition? What is the handwriting recognition? Describe isolated and context-dependent word errors detection and correction. Что такое оптическое распознавание символов? Что такое распознавание ручного письма? Опишите распознавание ошибок в словах и их исправление, рассматривая слова отдельно и в контексте.
3. Do the morphological parsing and build the finite-state automata for the derivational morphology.

Words: specification, realizable, localize, computational, freedom, partially

**Билет 11**

1. Describe the communicational system between Dave and HAL. Describe distinct categories of the knowledge of language. Опишите систему взаимодействия Dave – HAL. Опишите выделенные категории знаний языка.
2. Why do we use smoothing in N-gram model? Write formulas for finding unsmoothed frequency $C$ and probability $P(C)$ for unigrams. Write formulas for finding smoothed $C^{\*}$ and $P(C)^{\*}$ for unigrams using **Add-one method**. Почему мы используем сглаживание в модели N-gram? Напишите формулы для нахождения несглаженных частоты $C$ и вероятности $P(C)$ для униграмм. Напишите формулы для нахождения сглаженных частоты $C^{\*}$ и вероятности $P(C)^{\*}$ для униграмм, используя метод сглаживания **Добавление единицы**.
3. Write a regular expression to find all uppercase letters in the text. Напишите регулярное выражение для поиска всех слов в верхнем регистре.

With a view 100 times bigger than that of NASA’s Hubble Space Telescope, WFIRST will aid researchers in their efforts to unravel the secrets of dark energy and dark matter, and explore the evolution of the cosmos.

**Билет 12**

1. What are regular expressions? Describe their meaning and goal. What symbols are commonly use when we need to choose between several characters or digits? Что такое регулярные выражения? Опишите, что они собой представляют и их цель. Какие символы обычно используются, когда нам необходимо выбрать буквы и цифры из определенного диапазона?
2. How is Bayesian formula used for describing the correction of errors in words? Estimate the probability of each correction $P(c)$ describing the formula $P\left(c\right)= \frac{C\left(c\right)+0.5}{N+0.5V}$. Как используется формула Байеса для описания исправления ошибок в словах? Оцените вероятность $P(c)$, описывая каждый компонент формулы $P\left(c\right)= \frac{C\left(c\right)+0.5}{N+0.5V}$
3. Compute the probability of a sentence using a table with probabilities of bigrams. Посчитайте вероятность предложения, используя таблицу вероятностей биграмм.

We have lost the mighty king who protected our lands.

|  |  |  |
| --- | --- | --- |
| <s> We 0.3We have 0.09have lost 0.04lost the 0.045 | the mighty 0.048mighty king 0.03king who 0.024 | who protected 0.032protected our 0.052our lands 0.078 |

**Билет 13**

1. What are regular expressions? Describe their meaning and goal. What symbol is used when we need to choose any symbol or digit in a word? Что такое регулярные выражения? Опишите, что они собой представляют и их цель. Какой символ используется, когда нам необходимо выбрать любой символ или цифру в слове?
2. Why do we use smoothing in N-gram model? Write formulas for finding unsmoothed frequency $C$ and probability $P(w\_{n}|w\_{n-1})$ for bigrams. Write formulas for finding smoothed $C^{\*}$ and $P(w\_{n}|w\_{n-1})^{\*}$ for bigrams using **Add-one method**. Почему мы используем сглаживание в модели N-gram? Напишите формулы для нахождения несглаженных частоты $C$ и вероятности $P(w\_{n}|w\_{n-1})$ для биграмм. Напишите формулы для нахождения сглаженных частоты $C^{\*}$ и вероятности $P(w\_{n}|w\_{n-1})^{\*}$ для биграмм, используя метод сглаживания **Добавление единицы**
3. Do the morphological parsing and build the finite-state automata for the inflectional morphology.

Words: created, brought, demonstrates, apples, teeth, boundaries

**Билет 14**

1. What is morphological parsing? What is inflectional morphology? Describe its features. Что такое морфологический разбор? Что такое флективная морфология. Опишите ее особенности.
2. Describe different tagsets for English. What two algorithms are used for automatically assigning part-of-speech tagging. Describe a rules-based part of speech tagging algorithm in details. Опишите раздличные наборы тегов частей речи английского языка. Опишите два основных алгоритма для автоматического определения части речи. Детально опишите алгоритм, основанный на знании правил.
3. Write a regular expression to find these words in the text: cdcdcdcd, cccccccc, dddddddd. Напишите регулярное выражения на поиска заданных слов в тексте: cdcdcdcd, cccccccc, dddddddd