**Модели и методы обработки естественных языков (вопросы экзамена)**

**1 блок**

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. 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?
3. 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?
4. 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?
5. What is morphological parsing? What is inflectional morphology? Describe its features. Что такое морфологический разбор? Что такое флективная морфология. Опишите ее особенности.
6. What is morphological parsing? What is derivational morphology? Describe its features? Что такое морфологический разбор? Что такое дерифационная морфология? Опишите ее особенности.
7. 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.
8. 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.
9. 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?
10. 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.
11. 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.
12. 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.
13. 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.
14. 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 $P\left(c\right)= \frac{C\left(c\right)+0.5}{N+0.5V}$
15. 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.
16. What are N-grams? Write a formula for finding a probability of a bigram in a text using a frequency C. Что такое N-граммы? Напишите формулу для нахождения вероятности появления определенной биграммы с тексте через частоту C. Опишите данную формулу.
17. 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**.
18. 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**.
19. 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 различные наборы тегов частей речи английского языка. Опишите два основных алгоритма для автоматического определения части речи. Детально опишите алгоритм, основанный на знании правил.
20. 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.

**2 блок**

1. Describe rules of context-free grammar. What parts of speech can be included in Noun phrases, verb phrases and preposition phrases? Give examples of such sentences. Опишите правила контекстно-свободной грамматики. Какие части речи включены в фразы типы Существительные, глаголы и предлоги?
2. Describe rules of context-free grammar. Give examples of declarative, imperative and interrogative sentences. Опишите правила контекстно-свободной грамматики. Напишите примеры утвердительных, побудительных и вопросительных предложений.
3. How is parsing used with context-free grammar? Describe top-down and bottom-up approaches of parsing. Give an example of parsing the whole sentence with a top-down approach. Как используется парсинг в контекстно-свободной грамматике? Опишите два подхода парсинга: сверху-вниз и снизу-вверх. Напишите пример разбора предложения с использованием подхода сверху-вниз.
4. How is parsing used with context-free grammar? Describe top-down and bottom-up approaches of parsing. Give an example of parsing the whole sentence with a bottom-up approach. Как используется парсинг в контекстно-свободной грамматике? Опишите два подхода парсинга: сверху-вниз и снизу-вверх. Напишите пример разбора предложения с использованием подхода снизу-вверх.
5. Describe one of the simplest ways to encode properties using feature structures. Опишите один из легких способов представления свойств с использованием функциональных структур.
6. Describe one of the simplest ways to encode properties using feature structures. Include agreement features in these structures. Опишите один из легких способов представления свойств с использованием функциональных структур. Включите функции соглашения в данные структуры.
7. Describe one of the simplest ways to encode properties using feature structures. Write block and graph representation of feature structures. Опишите один из легких способов представления свойств с использованием функциональных структур. Используйте два вида их представления: блочное и в виде графов.
8. Describe the way we can merge the information content of two or more structures – unifications. Give examples where this process succeeds or fails. Опишите способ слияния информации двух и более структур – унификации. Дайте примеры, где данный процесс завершается успешно или не выполняется в итоге.
9. What is the meaning representation? What differs it from the morphological parsing and part-of-speech tagging? Что такое представление значений? Что отличает его от морфологического парсинга и тегов частей речи?
10. What is the meaning representation? Describe the way where we use predicates for representing meanings. Give examples of predicates for simple sentences. Что такое представление значений? Опишите способ, где мы используем предикаты для представления значений. Приведите примеры предикатов для простых предложений.
11. What is the first-order predicate calculus? Describe main elements of this calculus. Give examples of predicates with constants as arguments. Что такое исчисление предикатов первого порядка? Опишите основные элементы данного исчисления. Приведите примеры предикатов с константами в качестве аргументов.
12. What is the first-order predicate calculus? Describe main elements of this calculus. Give examples of predicates with variables as arguments. Что такое исчисление предикатов первого порядка? Опишите основные элементы данного исчисления. Приведите примеры предикатов с переменными в качестве аргументов.
13. What is the first-order predicate calculus? Describe main elements of this calculus. What is the difference between predicates and functions. Give examples of both of them. Что такое исчисление предикатов первого порядка? Опишите основные элементы данного исчисления. Чем отличаются предикаты от функций? Приведите примеры их обоих (предикат, функция).
14. What is the first-order predicate calculus? Describe main elements of this calculus. What are logical connectives? Give examples of them. Что такое исчисление предикатов первого порядка? Опишите основные элементы данного исчисления.

Что такое логические соединители? Приведите их примеры?

1. What is the first-order predicate calculus? Describe main elements of this calculus. What two quantifiers do you know? Give examples where you use both of them. Что такое исчисление предикатов первого порядка? Опишите основные элементы данного исчисления. Какие два квантификаторы Вы знаете? Приведите примеры использования их обоих.
2. How is a list of states used to represent the meaning of a sentence (one predicate is used for all states)? Give an example of this list. Как используется список состояний для представления значения предложения (один предикат используется для всех состояний)? Приведите пример такого списка.
3. Describe the meaning representation using events. What do time and belief aspects mean in this representation? Опишите представления значений с использованием событий. Что означают аспекты времени и веры в данном представлении.
4. Describe the meaning representation using events. Draw time diagrams for different tenses in time representations. What do E, R, U symbols mean there? Опишите представления значений с использованием событий. Нарисуйте временную диаграмму представления значений. Что на ней означают символы E, R, U?
5. What is the word sense disambiguation process? What algorithms do you know? Describe Resnik’s selectional association-based word sense disambiguation algorithms. Что такое процесс снятия неоднозначности слова? Какие алгоритмы решения данной проблемы Вы знаете? Опишите выборный алгоритм снятия неоднозначности слова Резника.
6. What is the word sense disambiguation process? What algorithms do you know? Describe the feature-vector approach for the disambiguation approach Что такое процесс снятия неоднозначности слова? Какие алгоритмы решения данной проблемы Вы знаете? Опишите векторный алгоритм снятия неоднозначности слова.

**3 блок**

1. Do the morphological parsing and build the finite-state automata for the inflectional morphology.

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

1. Do the morphological parsing and build the finite-state automata for the derivational morphology.

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

1. Do the morphological parsing and build the finite-state automata for the inflectional morphology.

Words: opportunities, observes, operated, crowns, mice, deer

1. Do the morphological parsing and build the finite-state automata for the derivational morphology.

Words: recreational, kindness, glorify, enlarge, personally, drinkable

1. 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 |

1. Compute the probability of a sentence using a table with probabilities of bigrams. Посчитайте вероятность предложения, используя таблицу вероятностей биграмм.

We have finished the trip across the universe.

|  |  |
| --- | --- |
| <s> We 0.3We have 0.09have finished 0.05finished the 0.065 | the trip 0.047trip across 0.036across the 0.071the universe 0.039 |

1. 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.

1. 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

$P(books|many)$; биграмма $(many books)$ найдена в тексте 15 раз, слово $many$ найдено в тексте 50 раз, размер словаря V составляет 600.

1. Use top-down and bottom-up approach for the context-free parsing of the following sentences. Используйте два подхода сверху-вниз и снизу-вверх для парсинга следующих предложений: The minister put documents in his safe. Start the engine of the helicopter when we get in it.
2. Use top-down and bottom-up approach for the context-free parsing of the following sentences. Используйте два подхода сверху-вниз и снизу-вверх для парсинга следующих предложений: They knew how to reach the port on that side. Bring me a hamburger that I can eat in the afternoon.
3. Write feature structures for these words. If the same structures are applicable for several words use them. Создайте функциональные структуры для следующих слов. Если одна и та же структура применима для нескольких слов, укажите это. **Words**: shown, implement, observed, done, went, simulate.
4. Write feature structures for these words. If the same structures are applicable for several words use them. Создайте функциональные структуры для следующих слов. Если одна и та же структура применима для нескольких слов, укажите это.

**Words**: mice, apples, monument, you, he, they, John, Grant

1. Use unification to combine feature structures of these words if possible. Write a structure for every word. Используйте унификации для объединения функциональных структур данных слов. При этом напишите структуру для каждого слова.
* Cream, water, flower
* Done, shown, rotated
* Children, rose, foxes
* Go, doing, watched
1. Use unification to combine feature structures of these words if possible. Write a structure for every word. Используйте унификации для объединения функциональных структур данных слов. При этом напишите структуру для каждого слова.
* icebergs, ships, goat
* women, geese, oxen
* writing, listening, rotating
* executed, demonstrate, ridden
1. Characterize these sentences as predicates. Запишите данные предложения в виде предикатов.
* Find me a right solution of this problem.
* Find the way out of this maze.
* Bring me a book at 12 p.m. tomorrow.
* Helicopter above the building
* Plants and flowers in the garden
1. Characterize these sentences as predicates. Запишите данные предложения в виде предикатов.
* A block is on the table
* There is one man who knows all the ways
* Create a new mechanism
* Seeds and roots are on the farm
* This man ordered some food
1. Express the following sentences in the first-order predicate calculus. Выразите данные предложения с помощью исчисления предикатов первого порядка.
* This box contains a huge number of books.
* The TCRI building is located near a subway station
* There is only one laboratory that contains test tubes
* Every hunter knows the location of pheasants
* A military base that is located near the nuclear test site
1. Express the following sentences in the first-order predicate calculus. Выразите данные предложения с помощью исчисления предикатов первого порядка.
* Books are located on the shelf.
* The central gallery is near the center of the city.
* There is only one factory that produces ovens.
* Every character wanted to get the prize.
* All players were tired, so they lost the game.
1. Write events for the following sentences in the form of representing the time. Напишите события для данных предложений в форме представления времени.
* A minister has arrived in our city
* Many people will have a meeting tomorrow at 4 p.m.
* My sister had finished the work by the time I came home
1. Write events for the following sentences in the form of representing the time. Напишите события для данных предложений в форме представления времени.
* My friend will come to the office tomorrow
* I had done this exercise before my mother arrived
* You are running to the stadium