**Методы семантического анализа (вопросы экзамена)**

**1 блок**

1. Write about regular expressions. Describe their meaning and goal. Write what symbols are commonly used when we need to choose between several characters or digits. Напишите о регулярных выражениях. Опишите, что они собой представляют и их цель. Напишите, какие символы обычно используются, когда нам необходимо выбрать буквы и цифры из определенного диапазона.
2. Write about regular expressions. Describe their meaning and goal. Write what symbol is used when we need to choose any symbol or digit in a word. Напишите о регулярных выражения. Опишите, что они собой представляют и их цель. Напишите, какой символ используется, когда нам необходимо выбрать любой символ или цифру в слове.
3. Write about morphological rules and morphological parsing. Describe main morphemes of words. Напишите, что такое морфологические правила и морфологический разбор. Опишите основные морфемы слов, которые вы знаете.
4. Write about morphological parsing. Describe inflectional morphology and its features. Напишите, что такое морфологический разбор. Опишите флективную морфологию и ее особенности.
5. Write about morphological parsing. Describe derivational morphology and its features. Напишите про морфологический разбор. Опишите деривационную морфологию и ее особенности.
6. Write about the lexicon and morphotactics. Describe a finite-state automaton for English nominal inflection. Напишите про лексикон и морфотактику. Опишите автомат конечных состояний для инфлекционной морфологии английского языка.
7. Write about the lexicon and morphotactics. Describe a finite-state automaton for English derivational morphology. Напишите про лексикон и морфотактику. Опишите автомат конечных состояний для деривационной морфологии английского языка.
8. Describe two-level morphology for finite-state transducers. Write what are four ways of thinking about transducers. Опишите двухуровневую морфологию преобразователей конечных состояний. Напишите, какие четыре вида преобразователей существуют.
9. 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.
10. Write what is optical character recognition and the handwriting recognition. Describe isolated and context-dependent word errors detection and correction. Напишите, что такое оптическое распознавание символов и распознавание ручного письма. Опишите распознавание ошибок в словах и их исправление, рассматривая слова отдельно и в контексте.
11. Write how Bayesian formula is used for describing the correction of errors in words. Describe a general formula of Bayes and its specified version. Напишите, как используется формула Байеса для описания исправления ошибок в словах. Опишите общую формулу Байеса и ее частный случай.
12. Write how Bayesian formula is used for describing the correction of errors in words. Estimate the probability of each correction https://univer.kaznu.kz/Content/test/i/51026_irotp1x0db1lw0.jpg describing the formula https://univer.kaznu.kz/Content/test/i/51029_18nprq2izte6bo.jpg. Напишите, как используется формула Байеса для описания исправления ошибок в словах. Оцените вероятность https://univer.kaznu.kz/Content/test/i/51026_irotp1x0db1lw0.jpg, описывая каждый компонент формулы https://univer.kaznu.kz/Content/test/i/51029_18nprq2izte6bo.jpg.
13. Write what are N-grams. Describe a formula for counting a probability of a given sentence in a text. Напишите, что такое N-граммы. Опишите формулу подсчета вероятности появления определенной строки в тексте. Вы можете использовать биграммы.
14. Write about the reason we use smoothing in N-gram model. Describe formulas for finding unsmoothed frequency https://univer.kaznu.kz/Content/test/i/51036_9uqr6qxvvrfx0p.jpgand probability https://univer.kaznu.kz/Content/test/i/51037_i0d8vycj5qgnjo.jpg for unigrams. Write formulas for finding smoothed https://univer.kaznu.kz/Content/test/i/51040_o0aj78jnvdlys4.jpg and https://univer.kaznu.kz/Content/test/i/51043_2q69woxopkpzlg.jpg for unigrams using **Add-one method**.  Напишите, почему мы используем сглаживание в модели N-gram. Напишите формулы для нахождения несглаженных частоты https://univer.kaznu.kz/Content/test/i/51036_9uqr6qxvvrfx0p.jpg   и вероятности https://univer.kaznu.kz/Content/test/i/51037_i0d8vycj5qgnjo.jpg для униграмм. Напишите формулы для нахождения сглаженных частоты https://univer.kaznu.kz/Content/test/i/51040_o0aj78jnvdlys4.jpg  и вероятности https://univer.kaznu.kz/Content/test/i/51043_2q69woxopkpzlg.jpg  для униграмм, используя метод сглаживания **Добавление единицы**.
15. Write about the reason we use smoothing in N-gram model. Describe formulas for finding unsmoothed frequency https://univer.kaznu.kz/Content/test/i/51036_9uqr6qxvvrfx0p.jpg  and probability https://univer.kaznu.kz/Content/test/i/51056_smn0tqtk95r4j6.jpg for bigrams. Write formulas for finding smoothed https://univer.kaznu.kz/Content/test/i/51040_o0aj78jnvdlys4.jpg and https://univer.kaznu.kz/Content/test/i/51057_pynt9meqh0lyw1.jpg for bigrams using **Add-one method**. Напишите причину, по которой мы используем сглаживание в модели N-gram. Опишите формулы для нахождения несглаженных частоты https://univer.kaznu.kz/Content/test/i/51036_9uqr6qxvvrfx0p.jpgи вероятности https://univer.kaznu.kz/Content/test/i/51056_smn0tqtk95r4j6.jpg  для биграмм. Напишите формулы для нахождения сглаженных частоты https://univer.kaznu.kz/Content/test/i/51040_o0aj78jnvdlys4.jpgи вероятности https://univer.kaznu.kz/Content/test/i/51057_pynt9meqh0lyw1.jpgдля биграмм, используя метод сглаживания **Добавление единицы**.
16. 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.
17. 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.
18. 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?
19. 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.

**2 блок**

1. Describe the natural language processing with the use of NLTK library in Python. Опишите обработку естественных языков с использованием библиотеки NLTK на языке программирования Python.
2. Describe the term corpus and main corpuses available in NLTK library. Опишите значение термина «корпус» и основные корпуса, доступные в библиотеке NLTK.
3. Describe the processing of raw text. Focus on tokenization. Опишите процесс обработки неочищенного текста. Уделите внимание токенизации.
4. Describe the processing of raw text. Write a code of tokenization in Python. Опишите процесс обработки неочищенного текста. Напишите код токенизации на Питоне.
5. Write the following operations with text in Python: joining and splitting strings, and making it lowercase and uppercase. Напишите следующие операции с текстом на Питоне: соединение и разделение строк, приведение текста к верхнему и к нижнему регистру.
6. Describe how regular expressions are written in Python. Опишите процесс составления регулярных выражений на Питоне.
7. Describe all main steps of text normalization. Give examples in Python. Опишите все основные шаги нормализации текста. Приведите примеры на Питоне.
8. Describe stemming and lemmatization of words. Write about their meaning and give examples in Python. Опишите процесс стемминга и лемматизации слов. Укажите их значение и приведите примеры на Питоне.
9. Describe the process of removing stop words from the text and its purpose. Give examples in Python. Опишите процесс удаления стоп слов из текста и его назначение. Приведите примеры на Питоне.
10. Describe the vectorization process and TF-IDF metric. Опишите процесс векторизации текстов и использования метрики TF-IDF.
11. Describe sentiment analysis of texts and write what specific features it has. Опишите анализ тональности текстов и укажите его специфические особенности.
12. Describe the process of text categorization (multiclass classification). Опишите процесс категоризации текстов (многоклассовая классификация).
13. Describe implementation of Bayes theorem for text classification. Опишите применение теоремы Байеса для классификации текстов.
14. Describe implementation of SVM algorithm for text classification. Опишите применение машины опорных векторов для классификации текстов.
15. 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.
16. 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.
17. Describe one of the simplest ways to encode properties using feature structures. Опишите один из легких способов представления свойств с использованием функциональных структур.
18. 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.
19. 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.
20. 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.

**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 0.3  We have 0.09  have lost 0.04  lost the 0.045 | the mighty 0.048  mighty king 0.03  king who 0.024 | who protected 0.032  protected our 0.052  our 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 0.3  We have 0.09  have finished 0.05  finished the 0.065 | the trip 0.047  trip across 0.036  across the 0.071  the 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

; биграмма 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

; биграмма найдена в тексте 15 раз, слово 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 a program in Python that tokenizes the following text « Twenty-five millennia ago, the world of [Argus](https://wowwiki.fandom.com/wiki/Argus) was the home of the wondrous eredar civilization».
4. Write a program in Python that tokenizes the following text « Eredar society was shattered when these three leaders were contacted by Sargeras the Destroyer».
5. Write a program in Python to do stemming with the following list of words « Powers, wonderful, relentlessly, vanished, accepting».
6. Write a program in Python to do stemming with the following list of words «Civilizations, commanded, clans, ruling, beautiful».
7. Write a program in Python to do lemmatization with the following list of words «Troops, found, exterminated, cities, stubbornly».
8. Write a program in Python to do lemmatization with the following list of words «plains, overtaken, astonishing, generalization, approved».
9. Write a program in Python to delete stop words from the text «Instead the magics of the Sunwell simply react against him, and he is pulled back through it and into the Twisting Nether.»
10. Write a program in Python to delete stop words from the text «It is his very large necklace that rips off and lands on the surface of the Sunwell after he sinks in that is looted by players.»
11. Write a program to delete punctuations and make a text lowercase and uppercase « However, when used properly, an exclamation point can set a light, flirtatious tone… can convey excitement… and can even demonstrate interest in the person.»
12. Write a program to delete punctuations and make a text lowercase and uppercase « No doubt about it, he or she wants to stay informed on your plans and will make a clear effort to meet up. The exclamation point here shows the person writing is excited, engaged and interested.»