

Задачи и упражнения к межфакультетскому курсу «Введение в компьютерную лингвистику»

Алексей Андреевич Сорокин, Мати Рейнович Пентус

Весна 2023

1. Как с помощью большой языковой модели, обученной предсказывать вероятность следующего слова при условии контекста, решить задачи: 1) анализа тональности, 2) проверки логического следования, 3) ответа на общий (да/нет) вопрос?
2. Сколько обучаемых коэффициентов у бинарного линейного классификатора, если на вход подаются векторы длины 16 и модель обучается на 700 примерах?
3. Сколько слоёв нужно нейронной сети, чтобы бинарный классификатор отличил множество из двух векторов $\{[0, 0], [3, 3]\}$ от множества из двух векторов $\{[0, 3], [3, 0]\}$
4. Какую функцию активации обычно используют в свёрточных нейронных сетях между слоями?
5. Какая архитектура нейронных сетей подходит для задачи распознавания именованных сущностей?
6. Какие недостатки энграммных моделей мешают применять их в задачах порождения последовательностей?
7. Какую функцию потерь обычно используют в задачах многоклассовой текстовой классификации?
8. Какие из следующих задач включены в бенчмарк GLUE: 1) вопросы на «да»/«нет» по короткому отрывку из статьи в Википедии, 2) определение, когда истинность одного предложения следует из другого предложения, 3) предсказывание замаскированных слов и фраз из списка вариантов в отрывках текстов?
9. В чём заключается дообучение (тонкая настройка) больших языковых моделей?
10. Найти самое короткое регулярное выражение, эквивалентное данному регулярному выражению.
`[aou] [b-np-tv-z]*[aou] [a-z]*`
11. Найти самое короткое регулярное выражение, эквивалентное данному регулярному выражению.
`\w\s(?:\w\w?\s)*\w?`
12. Какие из следующих регулярных выражений эквивалентны?
`(?:a|n|d+a)*[da]*`
`(?:a|n|d+a)*d*`
`(?:a|n|da)*d*`
`n*(?:a|d|an+)*`
13. Какие из следующих регулярных выражений эквивалентны?
`(?:\w\d)*\w`
`\w(?:\d\w)*`
`(?:\w|\d)*`
`\w*`

14. Найти регулярное выражение, принимающее текст на русском языке тогда и только тогда, когда в этом тексте хотя бы в одном месте оглушается звонкий согласный перед глухим согласным или концом слова.
15. Задайте регулярным выражением множество непустых слов в алфавите $\{a, i, k, l, m, n, p, u\}$, в которых не более 3 согласных, которые вдобавок находятся на чётных позициях.
16. Задайте регулярным выражением множество непустых слов в алфавите $\{a, e, i, j, k, l, m, n, o, p, s, t, u, w\}$, в которых не встречаются одновременно гласные заднего ряда и гласные переднего ряда. В этом языке буквы a, o, u обозначают гласные заднего ряда, а буквы e, i обозначают гласные переднего ряда.
17. Задайте регулярным выражением множество непустых слов в алфавите $\{a, e, i, j, k, l, m, n, o, p, s, t, u, w\}$, в которых нет закрытых слогов и не встречаются сочетания ti, ji, wo, wu .
18. Найти конечный автомат, эквивалентный данному регулярному выражению.
 $[ab]^+(ab)^+$
19. Найти конечный автомат, эквивалентный данному регулярному выражению.
 $(ab?a^?|bb)^*$
20. Задайте конечным автоматом множество слов в алфавите $\{a, k, m, n, o, p, s, t, u\}$, в которых не встречаются сочетания mt, mk , пр..
21. Найти расстояние Левенштейна между словами «платье» и «пальто». Найти взвешенное расстояние, если замены гласный—гласный и согласный—согласный стоят 1, замены разнотипных символов 5, удаление — 3, а перестановка соседних символов — 2.
22. Найти с помощью префиксного бора слова, отстоящие от слова «столе» на одну ошибку в словаре из слов «сталь», «стиль», «соль», «тальк», «толь», «стол».
23. Всегда ли дополнение регулярного языка является автоматным?
24. Всегда ли дополнение неавтоматного языка является неавтоматным?
25. Всегда ли объединение двух неавтоматных языков является неавтоматным?
26. Всегда ли пересечение автоматного языка и неавтоматного языка является автоматным?
27. Существуют ли такие автоматные языки L и M , что язык $L \setminus M$ неавтоматный?
28. Существуют ли такие бесконечные языки L и M , что язык $L \setminus M$ неавтоматный?
29. Всегда ли объединение неавтоматного языка и конечного языка является неавтоматным?