

## §2 Второй отборочный этап

Второй отборочный этап проводится в командном формате в сети интернет, работы оцениваются автоматически средствами системы онлайн-тестирования. Продолжительность второго отборочного этапа — 3 месяца с учетом того, что задачи открываются участникам в онлайн-системе последовательно, на каждую задачу отводится 2 недели. Задачи носят междисциплинарный характер и в более простой форме воссоздают инженерную задачу заключительного этапа. Решение каждой задачи дает определенное количество баллов. Баллы зачисляются в полном объеме за правильное решение задачи в зависимости от значимости задачи для финального этапа.

### Задача 2.1

*Условие:*

Вениамин вложил деньги в банк с необычной системой начисления бонусов. Каждый месяц приходит некоторая сумма денег. Вениамин может или взять эту сумму или отказаться от нее навсегда, но тогда бонус следующем месяце увеличится в 1.75 раза. Бонус всегда целый и округляется вниз (т.е. 375.99 округляется до 375). Нельзя отказываться от денег дважды подряд, т.е. если Вениамин отказался от денег, то в следующем месяце обязан взять бонус, но через месяц может опять отказаться ради умножения суммы в будущем.

На вход подается последовательность натуральных чисел, которые означают суммы денег, которые полагаются Вениамину каждый месяц от 1го до N месяца, числа от 1 до 500000.

В качестве ответа необходимо выдать число равное максимальной сумме денег, которую Вениамин может заработать оптимально, забирая бонусы и отказываясь от них ради удвоения. Обращаем внимание на ограничение по времени работы программы!

Возможен некорректный ввод: отрицательные числа, буквы, в этом случае программа должна выдавать в качестве ответа 0

Sample Input:

```
234 294 337 372 418 478 341 252 394 136 299 206 280 105 327 268 366
125 385 140 219 337 227 358 442 141 488 443 366 309 474 357 412 438
229 426 306 392 272 279 438 345 221 117 482 198 143 456 305 144 130
189 109 470 354 120 419 115 345 351 141 455 158 145 264 128 417 162
220 305 105 287 139 125 427 460 197 349 153 247 249 349 138 444 266
310 310 475 134 225 123 352 249 472 457 319 181 308 173 149 276 177
393 350 219 181 289 125 446 492 462 170 135 347 151 300 269 423 257
309 333 230 132 411 497 233 104 418 278 419 273 449 263 408 115 288
146 380 439 489 162 384 333 342 187 333 124 427 215 439 293 342 429
315 205 106 383 252 198 184 327 217 481 492 106 397 222 117 352 101
287 100 466 189 238 121 162 395 116 483 242 455 249 373 145 223 175
416 266 471 414 486 307 269 388 318 448 222 216 155 205 126 134 118
403 172 319 435 155 442 259 209 462 206 162 290 202 198 499 145 384
345 487 169 378 343 192 378 188 467 205 194 328 357 230 327 424 435
107 199 417 246 126 494 451 392 390 445 230 135 395 229 145 437 327
108 395 236 400 238 294 232 196 374 418 255 420 254 348 260 134 342
294 241 199 204 336 135 118 268 169 237 357 214 304 183 333 478 185
325 301 492 314 147 238 167 217 214 281 266 388 365 349 440 402 165
479 430 223 272 491 209 228 490 190 435 269 281 248 234 443 188 201
370 223 498 488 433 278 103 280 315 292 416 416 307 132 140 145 231
257 208 370 242 118 408 274 417 283 396 189 313 391 296 362 373 210
135 230 276 221 280 288 200 361 150 494 243 336 215 166 192 437 492
185 121 279 117 122 158 343
```

Sample Output:  
122218

## Задача 2.2

*Условие:*

Условия задачи по классификации текстов:

Исходный датасет содержит:

- csv-файл, в котором в первой колонке содержится предложение с сохранением пунктуации, во второй колонке - название произведения, в третьей колонке - номер автора. [https://drive.google.com/open?id=1IESSRC-3GqL8Wlf\\_YRCF9sfliiDxKvY3Q1J9IFduBGg](https://drive.google.com/open?id=1IESSRC-3GqL8Wlf_YRCF9sfliiDxKvY3Q1J9IFduBGg)
- csv-файл с номерами авторов: в первой колонке номер автора, во второй колонке - фамилия, в третьей - имя, в четвертой - отчество (если есть). <https://drive.google.com/open?id=1TdeuKTItvV8n1Q8i3uWWlua14RCR8WR7oeHbkv73kJE>

Вам необходимо определить тональность текста: оценить каждое предложение по шкале: "позитивное", "негативное", "нейтральное".

Результат представить в виде csv-файла, в котором в первой колонке содержится предложение с сохранением пунктуации, во второй колонке - слово "позитивное", "негативное" или "нейтральное".

Для тестирования вашего решения задачи (описание в предыдущем шаге урока) можно использовать датасет, в котором представлены уже размеченные предложения.

Скачать датасет можно по ссылке:

<https://docs.google.com/spreadsheets/d/1B2QwVXhy9v7na05hGN5o4C5jmmrts43qbrRBtK0yz9Q/edit?usp=sharing>

Для загрузки вашего решения задачи необходимо:

- Скачать файл из строчки со словом "sentence " и 700 записей предложений (всего 701 строчка).

Пример содержания файла на вход (с 3-мя предложениями) data.csv:

```
"sentence"
```

```
-- А почерк превосходный."
```

```
-- Что, что, что?"
```

```
-- Врешь!"
```

- Создать новый файл с ответами, используя написанный вами алгоритм, и в течение 5 минут загрузить его для проверки. В загружаемом файле в каждой строке должна быть указана окраска соответствующего предложения из входных данных. Ответ в кавычках без отступа с начала строки, в первой строке слово label. Всего 701 строчка.

Пример файла-ответа (на три предложения) answer.csv:

```
"label"
```

```
"positive"
```

```
"neutral"
```

```
"negative"
```

*Критерии оценки:*

Количество попыток не ограничено и нет никакого штрафа за количество попыток, однако каждый раз генерируется новый входной файл с новым набором предложений. Количество баллов от 0 до 10. Оно определяется по последнему загруженному ответу, даже если ранее были ответы на большее количество баллов.

Ввиду особенностей платформы вам будет отображено что задача решилась неправильно (если ваш ответ получает менее 10 баллов), однако точное количество баллов вы сможете посмотреть в комментарии к ответу.

Эти баллы будут учтены при определении финалистов второго тура.

### **Задача 2.3**

*Условие:*

В этой задаче мы будем анализировать открытые данные на материале текстов из русской Википедии.

На вход задаче будет подаваться разделенные двойным пропуском строки заглавные абзацы статей из википедии, которые относятся к одной или нескольким категориям из приведенного ниже списка (в конце описания задачи). В качестве результата необходимо определить к каким категориям из приведенного ниже списка относится статья.

*Входные данные:* набор строчек. Каждая новая строчка - новая статья

*Выходные данные:* файл, в котором в каждой строке набор чисел, разделенных пробелом, обозначающих, к каким категориям относится статья

*Пример:*

*Вход:*

Пьётро Аарон (итал. Pietro Aaron), он же Пьетро Аронне (итал. Pietro Aronne), или Пьеро Арон (итал. Piero Aron; около 1480 года, лоренция, Флорентийская республика — после 1545 года, Венеция, Венецианская республика или Флоренция, Флорентийское герцогство) — итальянский теоретик музыки.[1]

Абда́лка или Абда́льская (укр. Абдалка, крымско-тат. Abdal, Абдал), ранее называлась Боурча[2][3] — небольшая река, протекающая в Симферополе. Один из основных притоков реки Малый Салгир.

Тихон Иванович Абабков (15 августа 1908, деревня Коренец, Российская империя — 13 сентября 1984, Москва, РСФСР, СССР) — советский государственный и партийный деятель.

*Выход:*

1 11

6

1 13

*Список Категорий:*

1. Персоналии по алфавиту (339 727 объектов)
2. Населённые пункты по алфавиту (211 477 объектов)
3. Спортсмены по алфавиту (65 050 объектов)
4. Водные объекты по алфавиту (36 460 объектов)
5. Учёные по алфавиту (35 781 объект)
6. Реки по алфавиту (31 069 объектов)
7. Животные по алфавиту (29 011 объектов)
8. Футболисты по алфавиту (28 894 объекта)
9. Кавалеры ордена Ленина (25 337 объектов)
10. Фильмы по алфавиту (25 305 объектов)
11. Музыканты по алфавиту (21 288 объектов)
12. Писатели по алфавиту (18 616 объектов)
13. Умершие в Москве (16 393 объекта)
14. Альбомы по алфавиту (13 577 объектов)

15. UPD: Статьи из этой категории не будет в результирующей выборке Статьи проекта Биология (13 279 объектов)
16. Герои Советского Союза (12 800 объектов)
17. Родившиеся в Москве (11 374 объекта)
18. Актёры по алфавиту (10 828 объектов)
19. Растения по алфавиту (10 185 объектов)
20. Компании по алфавиту (10 056 объектов)
21. Члены КПСС (9701 объект)
22. Актрисы по алфавиту (9657 объектов)
23. Улицы по алфавиту (9400 объектов)
24. Музыкальные коллективы по алфавиту (7855 объектов)
25. Политики по алфавиту (7090 объектов)
26. Умершие в Санкт-Петербурге (7062 объекта)
27. Награждённые медалью «За победу над Германией в Великой Отечественной войне 1941—1945 гг.» (7003 объекта)
28. Родившиеся в Санкт-Петербурге (6833 объекта)
29. Галактики по алфавиту (6635 объектов)
30. Компьютерные игры по алфавиту (5746 объектов)
31. Синглы по алфавиту (5470 объектов)
32. Виды вне опасности (5345 объектов)
33. Участники Первой мировой войны (5193 объекта)
34. Футбольные клубы по алфавиту (4922 объекта)
35. Умершие в Париже (4833 объекта)
36. Офицеры СССР (4646 объектов)
37. Актёры XXI века (4279 объектов)

На этом шаге осуществляется попытка учета решений. Неограниченное количество попыток. Вам загружается файл с текстами, необходимо в ответ написать номера категорий. Абзацы все занимают одну строку и ответы тоже. Пустых строк нет.

*Пример:*

*Вход:*

Пьётро Аарон (итал. Pietro Aaron), он же Пьетро Аронне (итал. Pietro Aronne), или Пьеро Арон (итал. Piero Aron; около 1480 года, Флоренция, Флорентийская республика — после 1545 года, Венеция, Венецианская республика или Флоренция, Флорентийское герцогство) — итальянский теоретик музыки.[1]

Абдалка или Абдальская (укр. Абдалка, крымско-тат. Abdal, Абдал), ранее называлась Боурча[2][3] — небольшая река, протекающая в Симферополе. Один из основных притоков реки Малый Салгир.

Тихон Иванович Абабков (15 августа 1908, деревня Коренец, Российская империя — 13 сентября 1984, Москва, РСФСР, СССР) — советский государственный и партийный деятель.

*Выход:*

1 11

6

1 13

*Критерии оценки:*

При проверке учитываются только те категории, которые указаны в самой статье понятия.

Количество баллов за задачу считается:  $10 * ((\text{Количество правильных ответов})^2) / ((\text{общее количество ответов}) * (\text{количество ответов у правильного решения}))$