

Лабораторная работа №3

Исследование генетического алгоритма для решения задач оптимизации

1 Цель работы

Ознакомление с методом генетического алгоритма для решения задач оптимизации.

Реализация генетического алгоритма.

2 Теоретическая справка

2.1 Генетический алгоритм

Генетические алгоритмы – это метод решения оптимизационных задач, основанный на биологических принципах естественного отбора и эволюции. Генетический алгоритм повторяет определенное количество раз процедуру модификации популяции (набора отдельных решений), добиваясь тем самым получения новых наборов решений (новых популяций). При этом на каждом шаге из популяции выбираются «родительские особи», то есть решения, совместная модификация которых (скрещивание) и приводит к формированию новой особи в следующем поколении. Генетический алгоритм использует три вида правил, на основе которых формируется новое поколение: правила отбора, скрещивания и мутации. Мутация позволяет путем внесения изменений в новое поколение избежать попадания в локальные минимумы оптимизируемой функции.

Данные алгоритмы основаны на принципах естественного отбора Ч. Дарвина и предложены относительно недавно – в 1975 году Джоном Холландом. В них используются как аналог механизма генетического наследования, так и аналог естественного отбора. При этом сохраняется биологическая терминология в упрощенном виде и основные понятия линейной алгебры.

Генетические алгоритмы оптимизации являются алгоритмами случайно-направленного поиска и применяются в основном там, где сложно или невозможно сформулировать задачу в виде, пригодном для более быстрых алгоритмов локальной оптимизации, либо если стоит задача оптимизации не дифференцируемой функции или задача многоэкстремальной глобальной оптимизации.

Основной идеей ГА является организация «борьбы за существование» и естественный отбор среди пробных решений задачи. Поскольку ГА используют биологические аналогии, то применяющаяся терминология напоминает биологическую. Как известно, эволюционная теория утверждает, что жизнь на нашей планете возникла вначале лишь в простейших ее формах – в виде одноклеточных организмов. Эти формы постепенно усложнялись, приспосабливаясь к окружающей среде и порождая новые виды, и только через многие миллионы лет появились первые животные и люди. Можно сказать, что каждый биологический вид с течением времени улучшает свои качества так, чтобы наиболее эффективно справляться с важнейшими задачами выживания, самозащиты, размножения и т. д. Таким путем возникла защитная окраска у многих рыб и насекомых, панцирь у черепахи, яд у скорпиона и многие другие полезные приспособления.

Основными параметрами ГА являются:

- вероятность мутации;
- точность получения результата;
- количество итераций алгоритма или количество поколений;
- размер популяции.

Генетический алгоритм работает согласно следующей схеме:

1) Прежде всего, в данном алгоритме для организации начала счета создается произвольное исходное семейство.

2) Далее алгоритм производит некую последовательность новых семейств или поколений. На каждом отдельном шаге алгоритм использует определенные индивидуумы из текущего поколения, для того чтобы создать последующее поколение. При формировании нового поколения в алгоритме проводятся следующие действия:

- Отмечается каждый член текущего семейства посредством вычисления соответствующего значения пригодности – fitness функция;
- Проводится масштабирование полученного ряда значений функции пригодности, что позволяет построить диапазон значений более удобный для последующего использования;
- Выбираются родительские значения на основе значений их пригодности;
- Часть индивидуумов из родительского поколения имеет более меньшие значения функции пригодности и которые в далее выбираются как элитные значения. Эти элитные значения передаются далее уже в последующее поколение;
- Дочерние значения образуются или путем неких случайных изменений отдельного одного родителя - мутация - или путем комбинации

векторных компонентов некой пары родителей – кроссовер;

- Замена текущего семейства на дочернее с целью формирования последующего поколения.

3) Останов алгоритма производится тогда, когда выполняется какой-нибудь критерий останова.

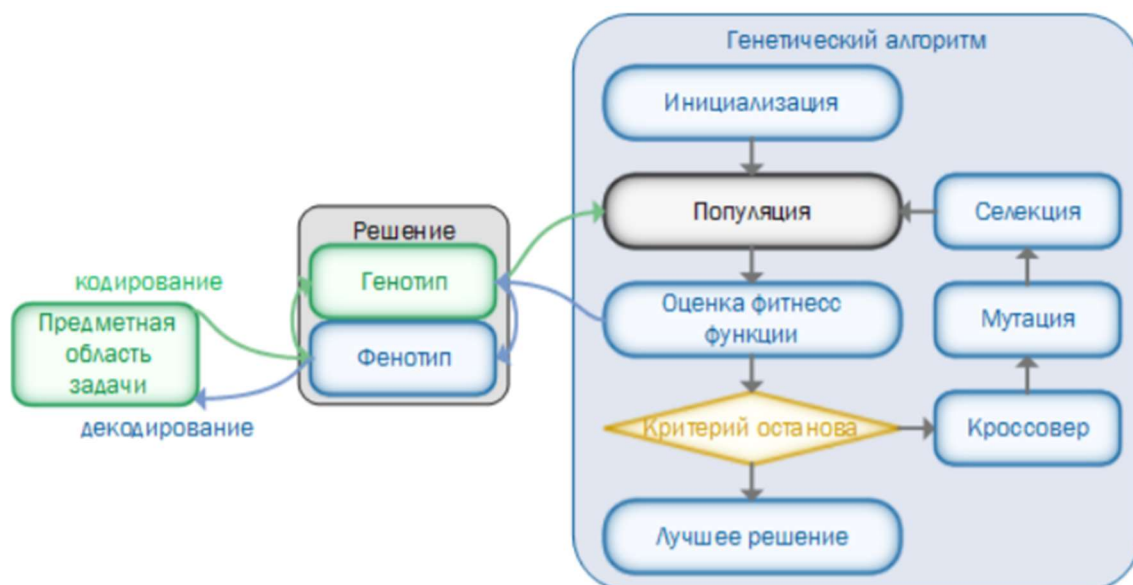


Рис. 1. Схема работы генетического алгоритма

2.1 Оператор селекции

Селекция (отбор) необходима, чтобы выбрать более приспособленных особей для скрещивания. Существует множество вариантов селекции, опишем наиболее известные из них.

Рулеточная селекция. В данном варианте селекции вероятность i -й особи принять участие в скрещивании p_i пропорциональна значению ее приспособленности f_i . Процесс отбора особей для скрещивания напоминает игру в «рулетку». Рулеточный круг делится на сектора, причем площадь i -го сектора пропорциональна значению p_i . После этого n раз «вращается» рулетка, где n – размер популяции, и по сектору, на котором останавливается рулетка, определяется особь, выбранная для скрещивания.

Селекция усечением. При отборе усечением после вычисления значений приспособленности для скрещивания выбираются l_n лучших особей, где l – «порог отсечения», $0 < l < 1$, n – размер популяции. Чем меньше значение l , тем сильнее давление селекции, т.е. меньше шансы на выживание у плохо приспособленных особей. Как правило, выбирают l в интервале от 0,3 до 0,7.

Турнирный отбор. В случае использования турнирного отбора для скрещивания, как и при рулеточной селекции, отбираются n особей. Для этого из популяции случайно выбираются t особей, и самая приспособленная из них допускается к скрещиванию. Говорят, что формируется турнир из t особей, t – размер турнира. Эта операция повторяется n раз. Чем больше значение t , тем больше давление селекции. Вариант турнирного отбора, когда $t = 2$, называют бинарным турниром. Типичные значения размера турнира $t = 2, 3, 4, 5$.

2.2 Оператор скрещивания (кроссовер)

Отобранные в результате селекции особи (называемые родительскими) скрещиваются и дают потомство. Хромосомы потомков формируются в процессе обмена генетической информацией (с применением оператора кроссовера) между родительскими особями. Созданные таким образом потомки составляют популяцию следующего поколения. Будем рассматривать случай, когда из множества родительских особей случайным образом выбираются 2 особи и скрещиваются с вероятностью P_C , в результате чего создаются 2 потомка. Этот процесс повторяется до тех пор, пока не будет создано n потомков. Вероятность скрещивания P_C является одним из ключевых параметров генетического алгоритма и в большинстве случаев ее значение находится в диапазоне от 0,6 до 1. Процесс скрещивания на псевдоязыке выглядит следующим образом (предполагается, что размер подпопуляции родительских особей равен размеру популяции, RANDOM – случайное число из диапазона $[0; 1]$):

```

k = 0;
ПОКА (k < РАЗМЕР_ПОПУЛЯЦИИ) {
    i = RANDOM *
    РАЗМЕР_ПОПУЛЯЦИИ; j
    = RANDOM *
    РАЗМЕР_ПОПУЛЯЦИИ;
    ЕСЛИ ( $P_C > \text{RANDOM}$ ) {
        СКРЕЩИВАНИЕ (РОДИТЕЛЬ[i], РОДИТЕЛЬ[j],
                     ПОТОМОК[k], ПОТОМОК[k+1]);
        k = k+2;
    }
}

```

1-точечный кроссовер. Работает аналогично операции перекреста для хромосом при скрещивании биологических организмов. Для этого выбирается произвольная точка разрыва и для создания потомков производится обмен частями родительских хромосом. Иллюстративный пример работы 1-точечного кроссовера представлен на рисунке 2.

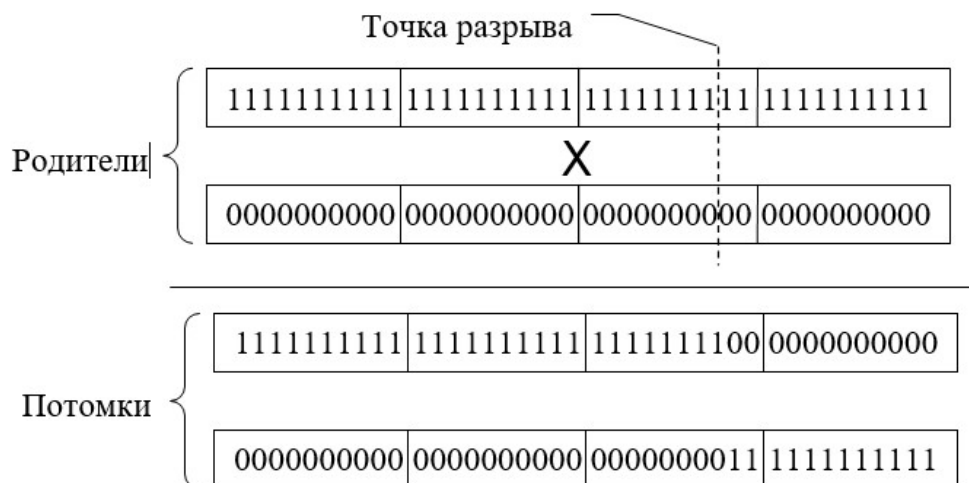


Рисунок - 2. Пример работы 1-точечного кроссовера

Для оператора *2-точечного кроссовера* выбираются 2 случайные точки разрыва, после чего для создания потомков родительские хромосомы обмениваются участками, лежащими между точками разрыва.

При использовании *однородного оператора кроссовера* разряды родительских хромосом наследуются независимо друг от друга. Для этого определяют вероятность p_0 , что i -й разряд хромосомы 1-го родителя попадет к первому потомку, а 2-го родителя – ко второму потомку. Вероятность противоположного события равна $(1 - p_0)$. В большинстве случаев вероятность обоих событий одинакова, т.е. $p_0 = 0,5$.

Операторы кроссовера характеризуются способностью к разрушению (disruption) родительских хромосом. Кроссовер для целочисленного кодирования считается более разрушительным, если в результате его применения расстояние по Хэммингу между получившимися хромосомами потомков и хромосомами родителей велико. Другими словами, способность целочисленного кроссовера к разрушению зависит от того, насколько сильно он «перемешивает» (рекомбинирует) содержимое родительских хромосом. Так, 1-точечный кроссовер считается слаборазрушающим, а однородный кроссовер в большинстве случаев является сильноразрушающим оператором.

Одновременно со способностью к разрушению говорят также о способности к созданию (creation, construction) кроссовером новых особей. Тем самым подчеркивается, что, разрушая хромосомы родительских особей, кроссовер может создать совершенно новые хромосомы, не встречавшиеся ранее в процессе эволюционного поиска.

Как уже упоминалось выше, в результате скрещивания создаются потомки, которые формируют популяцию следующего поколения. Обновленная таким образом популяция не обязательно должна включать

одних только особей-потомков. Если доля обновляемых особей равна T , то в новое поколение попадает Tn потомков, n – размер популяции, а $(1 - T)n$ особей в новой популяции являются наиболее приспособленными родительскими особями (так называемые элитные особи). Параметр T называют разрыв поколений (generation gap). Использование элитных особей позволяет увеличить скорость сходимости генетического алгоритма.

2.3 Оператор мутации

Оператор мутации используется для внесения случайных изменений в хромосомы особей. Это позволяет «выбираться» из локальных экстремумов и, тем самым, эффективнее исследовать пространство поиска. Так же, как и для оператора кроссовера, существует вероятность применения мутации P_M .

Оператор битовой мутации. В случае целочисленного кодирования мутация изменяет отдельные разряды в хромосоме. Для этого каждый разряд инвертируется с вероятностью P_M . Ниже приведен пример мутации на псевдоязыке:

```
для КАЖДОЙ k ОСОВИ в ПОПУЛЯЦИИ {
  для КАЖДОГО i РАЗРЯДА в ХРОМОСОМЕ k {
    ЕСЛИ ( $P_M > \text{RANDOM}$ ) {
      БИТОВАЯ_МУТАЦИЯ (ОСОБЬ[k], i);
    }
  }
}
```

В силу того, что применение мутации разыгрывается столько раз, сколько разрядов содержится в хромосоме, значение P_M выбирают небольшим, чтобы сильно не разрушать найденные хорошие хромосомы. Один из типичных вариантов $P_M = L^{-1}$, где L – длина хромосомы в битах, в этом случае каждая хромосома мутирует в среднем один раз.

2.4 Рекомендации по реализации ГА

При использовании генетического алгоритма для решения задачи оптимизации необходимо:

1. Определить количество и тип переменных задачи, которые необходимо закодировать в хромосоме.
2. Определить критерий оценки особей, задав функцию приспособленности (целевую функцию).
3. Выбрать способ кодирования и его параметры.
4. Определение параметров ГА (размер популяции, тип селекции, генетические операторы и их вероятности, величина разрыва поколений).

Результат работы генетического алгоритма сильно зависит от того, каким образом настроены его параметры. Различные параметры влияют на разные аспекты эволюционного поиска, среди которых можно выделить два наиболее общих:

1. Исследование пространства поиска (exploration).
2. Использование найденных «хороших» решений (exploitation).

Первый аспект отвечает за способности ГА к эффективному поиску решения и характеризует способности алгоритма избегать локальных экстремумов. Второй аспект важен для постепенного улучшения имеющихся результатов от поколения к поколению на основе уже найденных «промежуточных» решений. Пренебрежение исследовательскими способностями приводит к существенному увеличению времени работы ГА и ухудшению результатов из-за «застревания» алгоритма в локальных экстремумах. В итоге становится возможной преждевременная сходимость генетического алгоритма (также говорят о вырождении популяции), когда решение еще не найдено, но в популяции практически все особи становятся одинаковыми и долгое время (порядка нескольких десятков и сотен поколений) не наблюдается улучшения приспособленности.

Игнорирование найденных решений может привести к тому, что работа ГА будет напоминать случайный поиск, что также отрицательно сказывается на эффективности поиска и качестве получаемых решений.

Основная цель в настройке параметров ГА и, одновременно, необходимое условие для стабильного получения хороших результатов работы алгоритма – это достижение баланса между исследованием пространства поиска и использованием найденных решений.

Неправильная настройка параметров может стать причиной различных проблем в работе ГА. Краткий список таких проблем и возможные пути их исправления приведены в таблице 1.

Таблица 1. Проблемы в работе ГА и возможные пути их исправления

Проблема	Возможные способы исправления
1. Плохая приспособленность решений	<ol style="list-style-type: none"> 1. Увеличение числа поколений эволюционного поиска. 2. Увеличение численности популяции. 3. Изменение критерия оценки особей. 4. Исправление способа формирования родительских пар для скрещивания. 5. Исправление стратегии скрещивания и формирования нового поколения.
2. Преждевременная сходимость	<ol style="list-style-type: none"> 1. Изменение стратегии выбора родительских пар для скрещивания. 2. Отслеживание появления в популяции идентичных особей и их удаление. 3. Использование сильно разрушающего оператора кроссовера. 4. Увеличение вероятности мутации.
3. Низкая «стабильность» эволюции популяции (значительные скачки значений приспособленности от поколения к поколению).	<ol style="list-style-type: none"> 1. Применение «элитизма» (уменьшение разрыва поколений). 2. Уменьшение вероятности мутации. 3. Использование кроссовера со слабой разрушающей способностью.
4. Преобладание удовлетворительных результатов над хорошими.	<ol style="list-style-type: none"> 1. Изменение стратегии выбора родительских пар для скрещивания. 2. Изменение операторов скрещивания и/или мутации. 3. Распараллеливание поиска. Инициализация нескольких независимых популяций, которые развиваются независимо и, время от времени, обмениваются особями.

2.5 Канонический генетический алгоритм

Канонический генетический алгоритм разработан Джоном Холландом и описан в его книге «Адаптация в естественных и искусственных системах». Канонического ГА имеет следующие характеристики:

- целочисленное кодирование;
- все хромосомы в популяции имеют одинаковую длину;
- постоянный размер популяции;
- рулеточная селекция;
- одноточечный кроссовер;
- битовая мутация;
- новое поколение формируется только из особей-потомков.

3 Порядок выполнения работы

1. Изучить теоретическую справку по работе генетического алгоритма.
2. Реализовать на языке Python3 генетический алгоритм, решающий задачу угадывания слова и предложений из классического литературного произведения в соответствии с вариантом задания в Таблица 1.
3. Подготовить отчет о выполненной работе.

4 Методические указания к выполнению работы

К пункту 2:

Задача разделена на 3 основные части:

- Угадывание четырехбуквенного русскоязычного слова;
- Угадывание русскоязычной фразы;
- Угадывание фразы, содержащей как слова на русском языке, так и слова на иностранном языке.

Входными данными для алгоритма являются наборы алфавитов соответственно части задачи:

Для первой части задания:

[абвгдеёжзийклмнопрстуфхцчшщъыьэюя] и
[абвгдеёжзийклмнопрстуфхцчшщъыьэюяАБВГДЕЁЖЗИЙКЛМНОПРСТ
УФХЦЧШЩЪЫЬЭЮЯ]

Для первой второй задания:

[абвгдеёжзийклмнопрстуфхцчшщъыьэюяАБВГДЕЁЖЗИЙКЛМНОПРСТ
УФХЦЧШЩЪЫЬЭЮЯ-.,!::;?«»—]

Для третьей части задания:

[абвгдеёжзийклмнопрстуфхцчшщъыьэюяАБВГДЕЁЖЗИЙКЛМНОПРСТ
УФХЦЧШЩЪЫЬЭЮЯabcdefghijklmnopqrstuvwxyzABCDEF
GHIJKLMNOPQRSTUVWXYZ]

Промежуточными данными для работы алгоритма является количество правильно угаданных символов на каждой итерации. Алгоритм прекращает работу после полного угадывания слова, фразы или предложения.

Выходными данными алгоритма является затраченное на угадывание время.

Для первой части задания необходимо угадать слово методом полного перебора и проанализировать полученные результаты решения задачи данными методами.

Также необходимо придумать и реализовать способ улучшения времени выполнения генетического алгоритма без сокращения исходного набора символов для **хотя бы для одной** части задания. Улучшение не должно быть подогнано под вариант вашего задания и должно показывать уменьшение времени работы алгоритма на любом другом примере.

Таблица 1 - Варианты заданий

№ Варианта	Часть 1	Часть 2	Часть 3
1	борщ/гРоб	Как только адъютант сказал это, старый усатый офицер с счастливым лицом и блестящими глазами, подняв кверху саблю, прокричал: «Виват!»	Исполняешь на high левле
2	диск/Учет	Ну, хоть некоторое время... Меня же возьмите в друзья, в компаньоны, и уж уверяю, что затеем отличное предприятие. Слушайте, я вам в подробности это всё растолкую — весь проект!	яркий make Up
3	чушь/ценаА	А, почтеннейший! Вот и вы... в наших краях... — начал Порфирий, протянув ему обе руки. — Ну, садитесь-ка, батюшка!	Сходил на интересный blockbuster
4	цикл/ЮНЕц	— Успеем-с, успеем-с!.. А вы курите? Есть у вас? Вот-с, папиросочка-с... — продолжал он, подавая гостю папироску. — Знаете, я принимаю вас здесь, а ведь квартира-то моя вот тут же, за перегородкой... казенная-с, а я теперь на вольной на время.	Выразил искренний respect

5	хлор/флаг	— Да что это! Да где это я стою! — проговорила она в глубоком недоумении, как будто еще не придя в себя, — да как вы, вы, такой... могли на это решиться?.. Да что это!	осели на Brighton Beach
6	миля/ОРЁл	Бедные, кроткие, с глазами кроткими... Милые!.. Зачем они не плачут? Зачем они не стонут?.. Они всё отдают... глядят кротко и тихо... Соня, Соня! Тихая Соня!..	Это все made in Чайна
7	серб/ПАРК	Глаза его горели лихорадочным огнем. Он почти начинал бредить; беспокойная улыбка бродила на его губах. Сквозь возбужденное состояние духа уже проглядывало страшное бессилие.	Выглядишь просто super
8	роба/блЮЗ	«Хорошо ли? Натурально ли? Не преувеличил ли? — трепетал про себя Раскольников. — Зачем сказал: «женщины»?»	Познакомились на концерте Little Big
9	нога/БЛоК	В бреду, дескать!.. Ха-ха-ха! Он про весь вечер вчерашний знает! Про приезд матери не знал!.. А ведьма и число прописала карандашом!..	Встретились в Crocus City холле
10	леди/Торт	Видишь, я тогда всё себя спрашивал: зачем я так глуп, что если другие глупы и коли я знаю уж наверно, что они глупы, то сам не хочу быть умнее?	Fitness хорошо помогает расслабиться
11	кофе/риНГ	«Ну, вот и люди, — подумал он радостно, увидав несколько человек, бежавших к нему. — Они мне помогут!»	American boy уеду с тобой
12	игла/Рейс	Все живут точно в стеклянных банках... Лучше уж я сам пойду к этому... юноше... хотя, бог его знает,	Мое имя Таре а фамилия на бабках

		может быть, ему шестьдесят лет?..	
13	мама/ПАПА	Пусть я был смешон в Ваших глазах и в глазах Вашего брата, Николая Николаевича. Уходя, я в восторге говорю: «Да святится имя Твое».	съел Black Star бургер
14	тень/жила	— О, это-то совсем пустое дело! — возразил пренебрежительно Николай Николаевич. — Нам известны инициалы этого Пе Пе Же... Как его, Вера?	Вечером будет chill и драйв
15	вояж/ТаТу	— Что вы! Что вы?! Помилуйте!.. Большевистский фортель... Какой-то негодяй осмелился!.. Я сейчас же прикажу вновь выбелить стену. Черт знает что!.. Простите... такое нелепое происшествие...	Стакан водки с Cinzano
16	Желе/пиЛа	«Какие все вы сейчас довольные, радостные, счастливые, — все: и купцы, и биржевые маклеры, и чиновники разных рангов, и помещики, и люди голубой крови! А что с вами было три-четыре дня назад?	прошел GTA Vice City
17	зной/Пень	— ...И на их опыте, — очень горьком, к слову сказать, — должны мы учиться. Казаков Первого и Четвертого полков, — хотя, впрочем, какие они теперь казаки? — в будущем придется вешать через одного, а то и просто свалить всех...	Вчера Приобрёл пятый Iphone
18	мята/КОжа	В Петрограде вам делать нечего. Никаких бунтов там нет. Знаете вы, для чего вас туда посылают? Чтобы свергнуть Временное правительство... Вот! Кто вас ведет? — царский	Видел у парня на Avito

		генерал Корнилов. Для чего ему надо спихнуть Керенского?	
19	стоп/КОРЬ	«Мне доверили такое большое дело — и вот я его поганю собственными руками... Слова не свяжу... Да что же это со мной? Другой на моем месте сказал бы и убедил в тысячу раз лучше... О, черт, какая же я бездарь!»	Неважно Visa или Matercard
20	тяга/ОкНО	«Пролетарии и отпрыски, эпизод из дневных и вседневных грабежей! Прогресс! Реформа! Справедливость!»	Учил java и C++ в университете
21	утка/лето	Ипполит едва слушал Евгения Павловича, которому если и говорил «ну» и «дальше», то, казалось, больше по старой, усвоенной привычке в разговорах, а не от внимания и любопытства.	Я похож на Gucci
22	овод/ГИРЯ	Но что «мокрая курица» — в этом сомнения нет никакого: «Спокойна до того, что и растолкать нельзя! Впрочем, и „мокрые курицы“ не спокойны, — фу! Сбилась я с ними совсем!».	Взяли Hennessy VSOP на троих
23	клещ/ЛапА	— Ну, доктор, решайте нашу судьбу, — сказала княгиня. — Говорите мне всё. — «Есть ли надежда?» — хотела она сказать, но губы ее задрожали, и она не могла выговорить этот вопрос. — Ну что, доктор?..	Забронировал номер на Booking
24	ночь/ноРа	Она подумала: «Любит? Разве он может любить? Если б он не слышал, что бывает любовь, он никогда и не употреблял бы этого слова. Он и не знает, что	Пришел Cashback на карту

		такое любовь».	
25	лень/Игра	Он помнил, как он пред отъездом в Москву сказал раз своему скотнику Николаю, наивному мужику, с которым он любил поговорить: «Что, Николай! хочу жениться», — и как Николай поспешно отвечал, как о деле, в котором не может быть никакого сомнения: «И давно пора, Константин Дмитрич».	Поищи на Youtube Music
26	щука/ЛоЗА	«Эк какую баню задал! смотри ты какой!» Тут много было посулено Ноздреву всяких нелегких и сильных желаний; попались даже и нехорошие слова. Что ж делать? Русский человек, да еще и в сердцах.	Вышел новый процессор Intel
27	пляж/Бита	— Что ж барин? у себя, что ли? — Здесь хозяин, — сказал ключник. — Где же? — повторил Чичиков. — Что, батюшка, слепы-то, что ли? — сказал ключник. — Эхва! А вить хозяин-то я!	Необходимо развернуть Docker Container
28	узор/ГНоМ	Старик поднял глаза и произнес с расстановкою: — Здесь нет дел по крепостям. — А где же? — Это в крепостной экспедиции. — А где же крепостная экспедиция? — Это у Ивана Антоновича.	Новостное агентство NASA сообщило
29	хряк/перО	Поедет на дрожках, даст порядок, а между тем и словцо промолвит тому-другому: «Что, Михеич! нужно бы нам с тобою доиграть когда-нибудь в горку». — «Да, Алексей Иванович, — отвечал тот, снимая шапку, — нужно бы».	Включается Dark Mode на телефоне
30	плен/шаРЫ	«Павел Иванович! Ах Боже мой, Павел Иванович!	Ищу

		Любезный Павел Иванович! Почтеннейший Павел Иванович! Душа моя Павел Иванович! Вот вы где, Павел Иванович! Вот он, наш Павел Иванович! Позвольте прижать вас, Павел Иванович!	Наставника по React
--	--	--	------------------------

К пункту 3:

В отчете для каждой части задания продемонстрировать по 20 выводов алгоритма, снятых через равные промежутки итераций, а также итоговое время работы алгоритма, и его улучшенное значение, в случае если производилась доработка алгоритма.

По каждому этапу работы сделать промежуточные выводы.