

SPb HSE, 1 курс, весна 2021/22

Конспект лекций по алгоритмам

Собрано 28 августа 2022 г. в 12:19

Содержание

0. Графы и базовый поиск в глубину	1
0.1. Определения	1
0.2. Хранение графа	2
0.2.1. Мультисписок	3
0.3. Поиск в глубину	4
0.4. Классификация рёбер	4
0.5. Топологическая сортировка	5
1. Поиск в глубину (часть 1)	5
1.1. Компоненты сильной связности	6
1.2. Эйлеровость	7
2. Поиск в глубину (часть 2)	9
2.1. Раскраски	9
2.2. Рёберная двусвязность	9
2.3. Вершинная двусвязность	11
2.4. 2-SAT	12
2.4.1. Решение 2-SAT за $\mathcal{O}(nm)$	13
2.4.2. Решение 2-SAT за $\mathcal{O}(n + m)$	13
3. Введение в теорию сложности	15
3.1. Основные классы	15
3.2. NP (non-deterministic polynomial)	16
3.3. NP-hard, NP-complete	17
3.4. Сведения, новые NP-полные задачи	18
3.5. Задачи поиска	19
3.6. Гипотезы	20
3.7. Дерево сведений	20
4. (*) Дополнительная сложность	20
4.1. (*) Алгоритм Левина	21
4.2. (*) Расщепление	21
4.3. (*) Оценка с использованием весов	22
5. Рандомизированные алгоритмы	24
5.1. Случайные числа в C++ Немного про rand()	24
5.2. Определения: RP, coRP, ZPP	24
5.3. Примеры	25
5.4. Проверка на простоту	25

5.5. $ZPP = RP \cap coRP$	26
5.6. Двусторонняя ошибка, класс BPP	27
5.7. Как ещё можно использовать случайные числа?	27
5.8. Парадокс дней рождений. Факторизация: метод Полларда	28
5.9. 3-SAT и random walk	29
5.10. Лемма Шварца-Зишеля	30
5.11. Random shuffle	31
5.12. Дерево игры [stanford.edu]	31
5.13. (*) Квадратный корень по модулю	32
6. Кратчайшие пути	33
6.1. Short description	33
6.2. bfs	34
6.3. Модификации bfs	34
6.3.1. 1-k-bfs	34
6.3.2. 0-1-bfs	35
6.4. Дейкстра	35
6.5. A* (A-звездочка)	36
6.6. Флойд	37
6.6.1. Восстановление пути	37
6.6.2. Поиск отрицательного цикла	37
6.7. (*) Кратчайшие пути в прикладных задачах	38
6.8. (*) Перебор и кратчайшие пути	38
6.9. (*) Beam search	40
7. Кратчайшие пути	41
7.1. Алгоритм Форд-Беллмана	42
7.2. Выделение отрицательного цикла	43
7.3. Модификации Форд-Беллмана	43
7.3.1. Форд-Беллман с break	43
7.3.2. Форд-Беллман с очередью	44
7.3.3. (*) Форд-Беллман с random shuffle	44
7.4. Потенциалы Джонсона	45
8. DSU, MST и Йен	47
8.1. DSU: Система Непересекающихся Множеств	47
8.1.1. Решения списками	47
8.1.2. Решения деревьями	47
8.1.3. Оценка $\mathcal{O}(\log^* n)$	49
8.2. (*) Оценка $\mathcal{O}(\alpha^{-1}(n))$	49
8.2.1. (*) Интуиция и $\log^{**} n$	49
8.2.2. (*) Введение обратной функции Аккермана	50
8.2.3. (*) Доказательство	50
8.3. MST: Минимальное Остовное Дерево	52
8.3.1. Алгоритм Краскала	52
8.3.2. Алгоритм Прима	52
8.3.3. Алгоритм Борувки	52

8.3.4. Сравнение алгоритмов	52
8.3.5. Лемма о разрезе и доказательства	53
8.4. (*) Алгоритм Йена	53
8.5. (*) Алгоритм Эпштейна для k -го пути	54
9. Жадность и приближённые алгоритмы	55
9.1. Хаффман	55
9.1.1. Хранение кодов	56
9.2. Компаратор и сортировки	56
9.2.1. Задача про 2 станка	57
9.2.2. Выбор максимума	58
9.3. Жадность для гамильтонова пути. Варнсдорф	59
10. Приближённые алгоритмы	60
10.1. Коммивояжёр	60
10.1.1. 2-ОПТ через MST	60
10.1.2. 1.5-ОПТ через MST (Кристофидес)	60
10.2. Set cover	60
10.3. Рюкзаки	61
10.4. Схемы приближений	61
10.5. Knapsack	61
10.6. Partition	62
10.6.1. Алгоритм Кармаркара-Карпа (LDM)	63
10.6.2. (*) Применение для $\langle P, M \rangle$ -антихеш-теста	64
10.7. Bin packing	64
10.8. Надстрока	66
10.9. Литература	66
11. Центроидная декомпозиция	67
11.1. Построение и минимум на пути дерева	67
11.2. Реализация	68
12. Модели вычислений	70
12.1. RAM	70
12.2. RAM-w	70
12.3. Машина Тьюринга	70
13. BST и AVL	71
13.1. BST, базовые операции	71
13.2. Немного кода	72
13.3. AVL (Адельсон-Вельский, Ландис'1962)	74
13.3.1. Добавление в AVL-дерево	74
13.4. Split/Merge	77
13.5. Персистентность	78
13.6. Дополнительные операции, групповые операции	78
13.7. Неявный ключ	79
13.8. Reverse на отрезке	80
13.9. Итоги	80

13.10. Персистентность: итоги	80
13.11. В-дерево (Bayer, McCreight'1972)	81
13.11.1. Поиск по В-дереву	81
13.11.2. Добавление в В-дерево	81
13.11.3. Удаление из В-дерева	81
13.11.4. Модификации	82
13.11.5. Split/Merge	82
13.12. Производные В-деревя	82
13.12.1. 2-3-дерево (Hopcroft'1970)	82
13.12.2. 2-3-4-дерево и RB-дерево (Bayer'1972)	82
13.12.3. AA-дерево (Arne Anderson'1993)	83
14. Декартово дерево	83
14.1. Treap (Seidel, Aragon'1989)	84
14.2. Операции над Treap	85
14.3. Сборка мусора (garbage collection)	85
14.4. Дополнение о персистентности	86
14.4.1. Offline	86
14.4.2. Персистентная очередь за $\mathcal{O}(1)$	86
14.4.3. Простой персистентный дек (pairing)	87
14.4.4. Treap и избавление от Y	89
14.5. (*) Частичная персистентность: fat nodes	89
15. Дерево отрезков	91
15.1. Общие слова	91
15.2. Дерево отрезков с операциями снизу	91
15.3. Дерево отрезков с операциями сверху	92
15.4. (*) Хаки для памяти к дереву отрезков сверху	94
15.5. Динамическое дерево отрезков и сжатие координат	94
15.6. Персистентное дерево отрезков, сравнение с BST	95
15.7. 2D-запросы, дерево сортированных массивов	95
15.8. Многомерные деревья	96
15.9. Сканирующая прямая	97
15.10. k -я порядковая статистика на отрезке	98
15.11. (*) Фенвик	99
15.12. (*) Fractional Cascading	100
15.12.1. (*) Для дерева отрезков	100
15.12.2. (*) Для k массивов	100
15.13. (*) КД-дерево	101
16. LCA & RMQ	101
16.1. RMQ & Sparse table	102
16.2. RMQ за $\mathcal{O}(1)$ одним махом	103
16.3. LCA & Двоичные подъёмы	103
16.4. RMQ ± 1 за $\langle n, 1 \rangle$	104
16.5. LCA \rightarrow RMQ ± 1 и эйлеров обход	104
16.6. RMQ \rightarrow LCA	105

16.7. (*) LCA в offline, алгоритм Тарьяна	106
16.8. LA (level ancestor)	106
16.9. (*) Euler-Tour-Tree	106
16.10. (*) LA, быстрые решения	107
16.10.1. (*) Вишкин за $\langle \mathcal{O}(n \log n), \mathcal{O}(1) \rangle$	107
16.10.2. (*) Ladder decomposition + четыре русских	107

Лекция #0: Графы и базовый поиск в глубину

Январь 2022

0.1. Определения

Def 0.1.1. *Граф G – пара множеств вершин и рёбер $\langle V, E \rangle$. E – множество пар вершин.*

- Вершины ещё иногда называют *узлами*.
- Если направление ребёр не имеет значение, граф *неориентированный* (неорграф).
- Если направление ребёр имеет значение, граф *ориентированный* (орграф).
- Если ребру дополнительно сопоставлен вес, то граф называют *взвешенным*.
- Рёбра в орграфе ещё называют *дугами* и у ребра вводят понятие *начало* и *конец*.
- Если E – мультимножество, то могут быть равные рёбра, их называют *кратными*.
- Иногда, чтобы подчеркнуть, что E – мультимножество, говорят *мультиграф*.
- Для ребра $e = (a, b)$, говорят, что e *инцидентно* вершине a .
- *Степень* вершины v в неорграфе $deg v$ – количество инцидентных ей рёбер.
- В орграфе определяют ещё *входящую* и *исходящую степени*: $deg v = deg_{in} v + deg_{out} v$.
- Два ребра с общей вершиной называют *смежными*.
- Две вершины, соединённых ребром тоже называют *смежными*.
- Вершину степени ноль называют *изолированной*.
- Вершину степени один называют *висячей* или *листом*.
- Ребро (v, v) называют *петлёй*.
- *Простым* будем называть граф без петель и кратных рёбер.

Def 0.1.2. *Путь – чередующаяся последовательность вершин и рёбер, в которой соседние элементы инцидентны, а крайние – вершины. В орграфе направление всех рёбер от i к $i+1$.*

- Путь *вершинно простой* или просто *простой*, если все вершины в нём различны.
- Путь *рёберно простой*, если все рёбра в нём различны.
- Пути можно рассматривать и в неорграфах и в орграфах. Если в графе нет кратных рёбер, обычно путь задают только последовательностью вершин.

Замечание 0.1.3. Иногда отдельно вводят понятие *маршрута*, *цепи*, *простой цепи*. Мы, чтобы не захламлять лексикон, ими пользоваться не будем.

- *Цикл* – путь с равными концами.
- Циклы тоже бывают вершинно и рёберно простыми.
- *Ациклический* граф – граф без циклов.
- *Дерево* – ациклический неорграф.

0.2. Хранение графа

Будем обозначать $|V| = n$, $|E| = m$. Иногда сами V и E будут обозначать размеры.

• Список рёбер

Можно просто хранить рёбра: `pair<int,int> edges[m];`

Чтобы в будущем удобно обрабатывать и взвешенные графы, и графы с потоком:

```
1 struct Edge { int from, to, weight; };
2 Edge edges[m];
```

• Матрица смежности

Можно для каждой пары вершин хранить наличие ребра, или количество рёбер, или вес...

`bool c[n][n];` для простого невзвешенного графа. n^2 бит памяти.

`int c[n][n];` для простого взвешенного графа или невзвешенного мультиграфа. $\mathcal{O}(n^2)$ памяти.

`vector<int> c[n][n];` для взвешенного мультиграфа придётся хранить список всех весов всех рёбер между парой вершин.

`vector<vector<bool>> c(n, vector<bool>(n));` – чтобы первый способ правда весил n^2 бит.

Константа времени работы увеличится (нужно достать определённый бит 32-битного числа).

• Списки смежности

Можно для каждой вершины хранить список инцидентных ей рёбер: `vector<Edge> c[n];`

Чтобы списки смежности умели быстро удалять, заменяем `vector` на `set/unordered_set`.

• Сравнение способов хранения

Основных действий, которых нам нужно будет проделывать с графом не так много:

- `adjacent(v)` перебрать все инцидентные v рёбра.
- `get(a,b)` посмотреть на наличие/вес ребра между a и b .
- `all` просмотреть все рёбра графа
- `add(a,b)` добавить ребро в граф
- `del(a,b)` удалить ребро из графа

Ещё важно оценить дополнительную память.

	<code>adjacent</code>	<code>get</code>	<code>all</code>	<code>add</code>	<code>del</code>	memory
Список рёбер	$\mathcal{O}(E)$	$\mathcal{O}(E)$	$\mathcal{O}(E)$	$\mathcal{O}(1)$	$\mathcal{O}(E)$	$\mathcal{O}(E)$
Матрица смежности	$\mathcal{O}(V)$	$\mathcal{O}(1)$	$\mathcal{O}(V^2)$	$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\mathcal{O}(V^2)$
Списки смежности (vector)	$\mathcal{O}(deg)$	$\mathcal{O}(deg)$	$\mathcal{O}(V+E)$	$\mathcal{O}(1)$	$\mathcal{O}(deg)$	$\mathcal{O}(E)$
Списки смежности (hashTable)	$\mathcal{O}(deg)$	$\mathcal{O}(1)$	$\mathcal{O}(V+E)$	$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\mathcal{O}(E)$

Единственные плюсы первого способа – не нужна допамять; в таком виде удобно хранить граф в файле (чтобы добавить одно ребро, допишем его в конец файла).

Если матрица смежности уж слишком велика, можно хранить хеш-таблицу $\langle a, b \rangle \rightarrow c[a, b]$.

В большинстве задач граф хранят **списками смежности** (иногда с `set` вместо `vector`).

Пример задачи, которую хорошо решает матрица смежности:

- даны граф и последовательность вершин в нём, проверить, что она – простой путь.

Пример задачи, которую хорошо решают списки смежности:

- пометить все вершины, смежные с v .

Пример задачи, где нужна сила обеих структур:

- даны две смежные вершины, найти третью, чтобы получился треугольник.

0.2.1. Мультисписок

Рёбра, смежные с v , лежат в односвязном списке $head[v]$, $next[head[v]]$, $next[next[head[v]]]$...
Все перечисленные элементы – номера рёбер.

По номеру ребра e можем хранить любую информацию про него, например, куда оно ведёт.

```

1 struct MultiList {
2     struct Edge { int next, to; };
3     vector<int> head; // для каждой вершины первое ребро
4     vector<Edge> es; // все рёбра графа
5     int e; // количество рёбер
6     MultiList(int n, int m) :
7         head(n, -1), // -1 = признак конца списка
8         es(m), // максимальное число рёбер в графе
9         e = 0 { } // изначально рёбер нет
10    void addEdge(int a, int b) { // одно ориентированное ребро
11        es[e] = {head[a], b}, head[a] = e++;
12    }
13    void adjacent(int v) { // все рёбра смежные с v
14        for (int e = head[v]; e != -1; e = es[e].next)
15            cout << es[e].next << " "; // а можем делать что-нибудь другое
16    }
17 };

```

По сути эти те же «списки смежности», но более аккуратно сохранённые.

Пусть $V = E = 10^6$, граф случайный. Оценим память.

На 64-битной машине `vector<vector<int>> g` будет в итоге весить $X = ?_1 = 3 \cdot 8 + ?_2$ мегабайта (сам по себе вектор – 3 указателя). Можно подумать, что $?_2 = E \cdot 4$, но нет, в векторе `size` \neq `capacity` \Rightarrow нужно проводить эксперимент. На двух экспериментах видим, что $?_2 \approx E \cdot 12$ и $E \cdot 16$. Итого $X \approx 36 - 38$ мегабайт.

А мультисписок $12 = 3 \cdot 4$ мегабайта. Итого разница в ≈ 3 раза. По скорости создания/удаления (выделить память, положить все элементы, освободить память) мультисписок будет в ≈ 10 раз быстрее.

0.3. Поиск в глубину

Поиск в глубину = depth-first-search = dfs

- **Задача:** пометить все вершины, достижимые из a .
- **Решение:** рекурсивно вызываемся от всех соседей a .

```

1 vector<vector<int>> g(n); // собственно наш граф
2 void dfs(int v) {
3     if (mark[v]) return; // от каждой вершины идём вглубь только один раз
4     mark[v] = 1; // пометили
5     for (int x : g[v])
6         dfs(x);
7 }
8 dfs(a);

```

Время работы $\mathcal{O}(E)$, так как по каждому ребру dfs проходит не более одного раза.

• Компоненты связности

Def 0.3.1. Вершины неор графа лежат в одной компоненте связности iff существует путь между ними. Иначе говоря, это классы эквивалентности отношения достижимости.

Чуть модифицируем dfs: mark[v]=cc. И будем запускать его от всех вершин:

```

1 for (int a = 0; a < n; a++)
2     if (!mark[a])
3         cc++, dfs(a);

```

Теперь каждая вершина покрашена в цвет своей компоненты. Время работы $\mathcal{O}(V+E)$ – по каждому ребру dfs проходит не более одного раза, и кроме того запускается от каждой вершины.

• Восстановление пути

Ищем путь в определённую вершину? на обратном ходу рекурсии можно восстановить путь!

```

1 bool dfs(int v) {
2     if (v == goal) { path.push_back(v); return 1; }
3     mark[v] = 1;
4     for (int x : g[v])
5         if (dfs(x)) { path.push_back(x); return 1; }
6     return 0;
7 }
8 dfs(start);

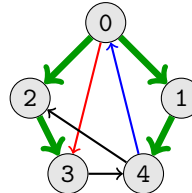
```

Вершины пути будут записаны в path в порядке от goal к start.

0.4. Классификация рёбер

После dfs(v) остаётся дерево с корнем в v . Отец вершины x – та, из которой мы пришли в x . Пусть все вершины достижимы из v . Рёбра разбились на следующие классы.

- **Древесные:** принадлежат дереву.
- **Прямые:** идут вниз по дереву.
- **Обратные:** идут вверх по дереву.
- **Перекрыстные:** идут между разными поддеревьями.



Рёбра можно классифицировать относительно любого корневого дерева, но именно относительно

но дерева, полученного `dfs` в неорграфе, нет перекрёстных рёбер.

Lm 0.4.1. *Относительно дерева `dfs` неорграфа нет перекрёстных рёбер*

Доказательство. Если есть перекрёстное ребро $a \rightarrow b$, есть и $b \rightarrow a$ (граф неориентированный). Пусть $t_{in}[a] < t_{in}[b]$. $a \rightarrow b$ перекрёстное $\Rightarrow t_{in}[b] > t_{out}[a]$. Противоречие с тем, что `dfs` пытался пройти по ребру $a \rightarrow b$. ■

Времена входа-выхода – стандартные полезные фишки `dfs`-а, считаются так:

```
1 bool dfs(int v) {
2     t_in[v] = T++;
3     ...
4     t_out[v] = T++;
5 }
```

0.5. Топологическая сортировка

Def 0.5.1. *Топологической сортировка орграфа – сопоставление вершинам номеров $ind[v]: \forall e(a \rightarrow b) ind[a] < ind[b]$.*

Lm 0.5.2. *Топологическая сортировка \exists iff граф ацикличесен.*

Доказательство. Если есть цикл, рассмотрим неравенства по циклу, получим противоречие. Если циклов нет, \exists вершина v нулевой входящей степени, сопоставим ей $ind[v] = 0$, удалим её из графа, граф остался ациклическим, по индукции пронумеруем оставшиеся вершины. ■

В процессе док-ва получили рекурсивный алгоритм топологической сортировки за $\mathcal{O}(V+E)$: поддерживаем входящие степени и очередь вершин нулевой входящей степени. Итерация:

```
1 v = q.pop(); for (int x : g[v]) if (--deg[x] == 0) q.push(x)
```

Алгоритм 0.5.3. *Рекурсивный топсорт*

`dfs` умеет сортировать вершины по времени входа и времени выхода.

```
1 bool dfs(int v) {
2     in_time_order.push_back(v);
3     ...
4     out_time_order.push_back(v);
5 }
```

Топологический порядок вершин записан в `reverse(out_time_order)`;

Доказательство: пусть есть ребро $a \rightarrow b$, тогда мы сперва выйдем из b и только затем из a .

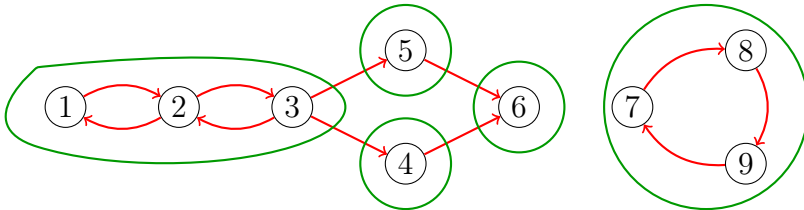
Лекция #1: Поиск в глубину (часть 1)

Январь 2022

1.1. Компоненты сильной связности

Def 1.1.1. В орграфе вершины a и b сильно связны iff существуют пути $a \rightsquigarrow b$ и $b \rightsquigarrow a$.

Def 1.1.2. Компонента сильной связности – класс эквивалентности отношения сильной связности (рефлексивно, симметрично, транзитивно).



Научимся выделять компоненты сильной связности.

- Алгоритм за $O(VE)$

Берём вершину v , запускаем dfs-ы из v по прямым и обратным рёбрам. Получили компоненту v .

- Алгоритм за $O(V+E)$

1. Выпишем вершины в порядке убывания времени выхода из dfs.

```

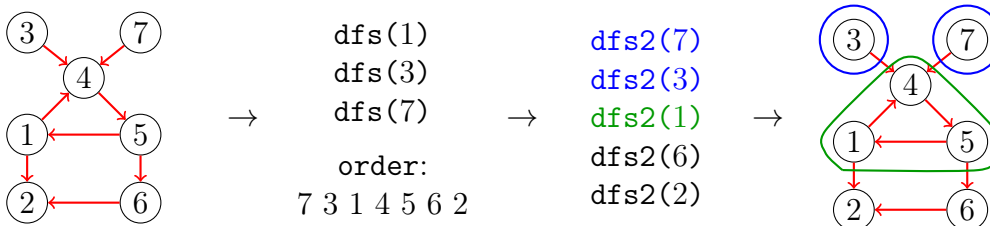
1 void dfs(int v) {
2     ...
3     order.push_front(v);
4 }
5 for (int v = 0; v < n; v++)
6     if (!used[v])
7         dfs(v); // dfs ходит по прямым рёбрам графа
    
```

Поскольку наш код делает то же, что и topsort (Algo 0.5.3), иногда говорят “запустим топсорт”.

2. Перебираем вершины в полученном порядке. Если вершина v ещё не покрашена, красим “не покрашенные вершины, достижимые из неё по обратным рёбрам”. Утверждается, что покрасим ровно компоненту сильной связности v .

```

1 int cc = 0; // количество компонент
2 for (int v : order)
3     if (color[v] == 0)
4         dfs2(v, ++cc); // dfs2 ходит по обратным рёбрам графа
    
```



Def 1.1.3. Конденсация орграфа $G = \langle V, E \rangle$ – новый граф, результат стягивания компонент сильной связности. Вершинами конденсации являются компоненты сильной связности G , а рёбра G : $(a, b) \in E$ порождают рёбра конденсации $(color[a], color[b])$.

Петли (случай $color[a] = color[b]$) в конденсацию обычно не добавляют.

Теорема 1.1.4. *Корректность алгоритма сильной связности за $O(V + E)$*

Доказательство. `dfs2(v)` точно обошёл всю компоненту сильной связности v . Почему он не обошёл лишнего? Пусть есть вершина x , достижимая по обратным рёбрам, но не достижимая по прямым. У неё есть компонента сильной связности $C(x)$. Утверждается, что до v мы уже перебрали хотя бы одну $a \in C(x) \Rightarrow C(x)$ уже была покрашена. См. лемму. ■

Lm 1.1.5. *Если есть путь $x \rightsquigarrow v$, но x и v в разных компонентах, $\exists a \in C(x): pos_a < pos_v$*

Доказательство. Достаточно доказать для соседних компонент в конденсации.

Пусть x' – первая вершина $\in C(x)$, до которой дошёл `dfs`.

Пусть v' – первая вершина $\in C(v)$, до которой дошёл `dfs`.

Если мы сперва посетили v' , то в x' из неё мы не дойдём – победа, в `order` вся $C(v)$ будет записана позже всей $C(x)$.

Если мы сперва посетили x' – мы не выйдем из неё, пока не посетим все вершины $C(x) \Rightarrow$ пройдем ребро (здесь воспользовались тем, что компоненты соседние), ведущее в $C(v) \Rightarrow$ обойдем всю $C(v) \Rightarrow$ из v выйдем до того, как выйдем из x .

Итого: x' – искомая вершина a . ■

1.2. Эйлеровость

Def 1.2.1. *Рёберно простой цикл/путь **эйлеров**, если содержит все рёбра графа.*

Def 1.2.2. *Граф **эйлеров**, если содержит эйлеров цикл.*

Теорема 1.2.3. *Связный неорграф эйлеров iff все степени вершин чётны.*

Доказательство. Возьмём \forall цикл, удалим его рёбра, в оставшихся компонентах связности по индукции есть эйлеровы циклы, соединим всё это счастье. Чтобы найти \forall цикл, возьмём \forall вершину v , пойдём жадно вперёд, удаляя пройденные рёбра: поскольку все степени чётны, остановиться мы можем только в v . ■

• Алгоритм построения

Будем делать всё, как в доказательстве теоремы. Сперва найдём \forall цикл.

```

1 vector<set<int>> g; // наивный способ хранения, удобно удалять рёбра
2 void dfs( int v ) {
3     if (!g[v].empty()) {
4         int x = *g[v].begin();
5         g[v].erase(x), g[x].erase(v);
6         dfs(x);
7         answer.push_back(edge(v, x));
8     }
9 }
```

Слово `dfs` уже намекает на рекурсию. Чтобы в оставшихся компонентах связности найти эйлеровы циклы и вставить в наш в нужные места, сделаем +1 рекурсивный вызов...

```

1 vector<set<int>> g; // наивный способ хранения, удобно удалять рёбра
2 void dfs( int v ) {
3     while (!g[v].empty()) { // единственное изменение кода
4         int x = *g[v].begin();
5         g[v].erase(x), g[x].erase(v);
6         dfs(x); // dfs(x) найдёт цикл одной из компонент и вставит ровно сюда
7         answer.push_front(edge(v, x));
8     }
9 }

```

Чтобы алгоритм работал за линию, нам бы от `set`-а избавиться. Будем удалять лениво:

```

1 struct Edge { int a, b, isDeleted; };
2 vector<Edge> edges;
3 vector<vector<int>> ids; // каждое ребро встретится два раза

```

В цикле `while` вынимаем все рёбра, пропускаем с пометкой `isDeleted`.

Идя из вершины v по ребру e , мы попадём в вершину $e.a + e.b - v$.

Lm 1.2.4. Слабосвязный оргграф эйлеров *iff* у всех вершин входящая степень равна исходящей.

Доказательство и алгоритм аналогичны. Реализация алгоритма даже проще: орребро не нужно удалять из вектора другой вершины \Rightarrow для хранения графа достаточно `vector<vector<Edge>`.

Лекция #2: Поиск в глубину (часть 2)

Январь 2022

2.1. Раскраски

Def 2.1.1. *Задача вершинной покраски – раскрасить вершины графа так, чтобы смежные были разных цветов.*

Что мы уже знаем?

Покрасить вершины в 2 цвета мы умеем за $\mathcal{O}(V + E)$ dfs-ом.

Покрасить в минимальное число цветов мы умеем динамикой за $\mathcal{O}(2.44^n)$. На самом деле красить уже в 3 цвета трудно (люди умеют только за экспоненту).

Как же красить большие графы?

Жадно за $\mathcal{O}(V + E)$! Получится не минимальное число цветов, но тоже неплохо.



- **Алгоритм**

Берём вершину минимальной степени, удаляем её из графа. Красим всё кроме неё, и в конце докрашиваем её в минимально возможный цвет.

Реализация за $\mathcal{O}(V + E)$: нужно уметь быстро выбирать вершину минимальной степени. Степени только уменьшаются причём ровно на 1 \Rightarrow поддерживаем `list[degree]` – для каждой степени список всех вершин такой степени и указатель на минимальную степень. Собственно удаление вершины: каждого из её соседей за $\mathcal{O}(1)$ переместить в соседний список.

Lm 2.1.2. *Если в каждый момент алгоритм $\min(\text{degree}_v) \leq x$, то наш алгоритм успешно покрасит граф в $\leq x+1$ цветов.*

Пример: любое дерево покрасится в 2 цвета.

Утверждение 2.1.3. Поскольку в любом планарном графе есть вершина степени ≤ 5 , наш алгоритм раскрасит его в ≤ 6 цветов.

2.2. Рёберная двусвязность

Def 2.2.1. *Говорят, что две вершины рёберно двусвязны iff между ними есть два рёберно не пересекающихся пути.*

Def 2.2.2. *Рёберная двусвязность – отношение эквивалентности на вершинах неорграфа. Классы эквивалентности – компоненты рёберной двусвязности.*

Теорема 2.2.3. *Рёберная двусвязность – отношение эквивалентности.*

Доказательство. Рефлексивность и симметричность очевидны.

Транзитивность: (u, v) и (v, w) рёберно двусвязны, доказываем двусвязность (u, w) .

Возьмём два пути $u \leftrightarrow v$, они образуют рёберно простой цикл C .

Пойдём двумя путями из w в v , остановимся, когда дойдём до цикла: мы на цикле в вершинах a и b , чтобы из них попасть в u – выберем нужное направление и пойдём по циклу C . ■

Def 2.2.4. Ребро (a, b) называется мостом, если после его удаления a и b не связны.

Замечание 2.2.5. После удаления из графа мостов, компоненты связности и рёберной двусвязности будут совпадать. Доказательство: $((a, b) - \text{не мост}) \Leftrightarrow (a \text{ и } b \text{ связны после удаления } (a, b))$.

Тривиальный алгоритм поиска мостов за $\mathcal{O}(VE)$: заметим, что любой мост лежит в остовном дереве \Rightarrow dfs-ом найдём любой остов, каждое ребро остова проверим за $\mathcal{O}(E)$.

• **Алгоритм поиска мостов за $\mathcal{O}(V + E)$**

Начало, как и в тривиальном: обойдём граф dfs-ом, посмотрим на остовное дерево, которое оставит нам dfs и на древесное ребро $e: v \rightarrow x$ в нём. Теперь придумаем, как за $\mathcal{O}(1)$ проверить, является ли e мостом. Не мост iff из поддерева x можно пройти наверх в обход e . Чтобы понимать, можно ли, стоя в x , пробраться наверх, насчитаем *динамику по дереву* — самая высокая вершина, которую можно достигнуть из x , если запрещено ходить по ребру $(v \leftrightarrow x)$.

Замечание 2.2.6. В канонической реализации вместо «самой высокой вершины» обычно ищут «минимальную по времени входа», разницы никакой, просто так принято.

Замечание 2.2.7. Важно, что мы используем именно dfs. Не любой обход графа подойдёт, например, bfs не подойдёт. См. доказательство.

Ниже мы вместе с мостами найдём сразу и компоненты.

Как только выходим из dfs и находим мост, на нём висит новая компонента, вершины которой мы уже обошли и заботливо сложили в стек, остаётся только не забыть достать.

Для полного понимания алгоритма прочитайте комментарии в коде.

```

1 void dfs(int v, int parent):
2     stack.push(v) // чтобы восстановить компоненты двусвязности
3     time[v] = ++T // время входа (а можно использовать высоту)
4     min_time[v] = ++T; // наша динамика по дереву
5     for (int x : adjacent[v])
6         if (x != parent) { // если бы в графе были кратные рёбра, проверка была бы сложнее
7             if (time[x] == 0) { // ещё не ходили в x  $\Rightarrow$  идём и делаем ребро древесным
8                 int stack_level = stack.size(); // уровень стека до рекурсивного вызова
9                 dfs(x, v);
10                if (min_time[x] > time[v]) { // мост!
11                    vector<int> component;
12                    while (stack.size() > stack_level)
13                        component.push(stack.pop())
14                }
15                relaxMin(min_time[v], min_time[x]);
16            } else { // обратное или прямое ребро (x или предок, или потомок v в дереве)
17                relaxMin(min_time[v], time[x]);
18            }
19        }

```

Строки 2, 8 и 11-13 нужны, чтобы восстановить компоненты в том же dfs. Если мосты находятся корректно, то в 11-13 мы откусываем компоненту, которая висит на ребре $v \rightarrow x$.

$\text{time}[v]$ — время входа в вершину v .

$\text{min_time}[v]$ — динамика по дереву, минимальное достижимое время входа из v по путям, которые спускаются по древесным рёбрам dfs, а потом делают один шаг по произвольному ребру.

Теорема 2.2.8. Алгоритм корректно находит все мосты

Доказательство. Нужно доказать, что мы нашли мосты и только их.

Если ребро $e: v \rightarrow x$ мост, то из x нельзя достичь $v \Rightarrow \text{min_time}[x] > \text{time}[v] \Rightarrow$ наш алгоритм корректно поймёт, что e — мост.

Если $e: v \rightarrow x$ не мост, то есть путь p из x в v в обход e . Поскольку мы используем именно **dfs**, кроме древесных и прямых рёбер есть только обратные и только они могут вести из поддерева x наружу, причём вести они могут только в вершины $z \in$ пути $\text{root} \rightsquigarrow v$. Осталось заметить, что $\text{min_time}[x] \leq \text{time}[z] \leq \text{time}[v] \Rightarrow$ наш алгоритм корректно поймёт, что e — не мост. ■

Замечание 2.2.9. Рёберная двусвязность — отношение эквивалентности \Rightarrow относительно него граф можно сконденсировать. Получится лес, вершинами которого являются компоненты рёберной двусвязности, а рёбрами — мосты.

2.3. Вершинная двусвязность

Def 2.3.1. Говорят, что два **ребра** вершинно двусвязны iff есть вершинно простой цикл, содержащий оба ребра.

Def 2.3.2. Вершинная двусвязность — отношение эквивалентности на **рёбрах** неорграфа. Классы эквивалентности — множества рёбер, компоненты вершинной двусвязности.

Теорема 2.3.3. Вершинная двусвязность — отношение эквивалентности.

Доказательство. Рефлексивность и симметричность очевидны.

Транзитивность: (e_1, e_2) и (e_2, e_3) вершинно двусвязны, доказываем двусвязность (e_1, e_3) .

Возьмём простой цикл C через e_1 и e_2 . Пойдём двумя путями из концов e_3 в концы e_2 , остановимся, когда дойдём до цикла: мы на цикле в различных вершинах a и b , чтобы из них попасть в концы e_1 , выберем нужное направление и пойдём по циклу C . ■

Замечание 2.3.4. Если ребро мост, оно одиноко, в его компоненте двусвязности лишь одно ребро.

Def 2.3.5. Точка сочленения — вершина, при удалении которой увеличивается число компонент связности.

Утверждение 2.3.6. Точки сочленения — ровно те вершины, у которых есть смежные рёбра из разных компонент вершинной двусвязности.

Тривиальный алгоритм поиска точек сочленения за $\mathcal{O}(VE)$: проверим каждую вершину v за $\mathcal{O}(E)$ **dfs**-ом, проверяющим граф на связность, запретив ему ходить через v . Затем, когда уже известны точки сочленения, компоненты вершинной двусвязности можно найти за $\mathcal{O}(V + E)$ почти обычным **dfs**-ом для поиска компонент связности — ходим рекурсивно из ребра в вершину, из вершины в ребро: ребро \rightarrow вершина \rightarrow ребро \rightarrow вершина $\rightarrow \dots$, запрещаем проходить через точки сочленения.

• Алгоритм поиска точек сочленения и компонент двусвязности за $\mathcal{O}(V + E)$

Почти идентичен поиску мостов. Идея та же: если мы из точки сочленения v пошли вглубь по ребру $e: v \rightarrow x$, подняться выше, не проходя через v , мы не сможем \Rightarrow

$$\text{min_time}[x] \geq \text{time}[v]$$

Это неравенство и будет критерием того, что мы нашли новую компоненту. Если неравенство окажется строгим, ребро ещё и будет мостом. По аналогии с алгоритмом поиска мостов, рёбра новой компоненты обнаружатся на вершине стека, в который мы бережно складываем все пройденные рёбра. Для полного понимания алгоритма прочитайте все комментарии в коде ниже.


```

1 void dfs(int v, int parent):
2     time[v] = min_time[v] = ++T;
3     int count = 0; // сколько компонент уже цепляется снизу за вершину v
4     for (int x : adjacent[v])
5         if (x != parent):
6             int stack_level = stack.size();
7             if (ещё не ходили по ребру {x,v}) // неорграф => по каждому ребру ходим дважды!
8                 stack.push({v,x})
9             if (time[x] == 0) {
10                dfs(x, v);
11                if (min_time[x] >= time[v]) { // новая компонента
12                    vector<Edge> component;
13                    while (stack.size() > stack_level)
14                        component.push(stack.pop())
15                    // Вершина v является точкой сочленения, если на ней смыкаются хотя бы
16                    // две компоненты. Одна точно есть, мы её только что нашли.
17                    // Где может быть вторая? Если v — не корень, выше. Если count ≥ 2, ниже.
18                    if (parent != -1 || ++count >= 2)
19                        points.push_back(v); // v — точка сочленения
20                }
21                relaxMin(min_time[v], min_time[x]);
22            } else {
23                relaxMin(min_time[v], time[x]);
24            }

```

Теорема 2.3.7. Алгоритм корректно находит все компоненты и точки сочленения

Доказательство. Похоже на историю с мостами. Далее мы предполагаем, что у читателя в кеше доказательство про мосты [Thm 2.2.8](#). Пусть мы шли в `dfs` $p \rightarrow v \rightarrow x$. Рёбра $p \rightarrow v$ и $v \rightarrow x$ лежат в одной компоненте iff есть путь $x \rightarrow$ предок v в дереве `dfs` $\Leftrightarrow \min_time[x] < time[v]$. ■

2.4. 2-SAT

Задача **2-SAT**: нам дана **КНФ** формула φ , в каждом клозе не более двух литералов. Нужно найти выполняющий набор x_i , или сказать, что φ противоречива. Пример:

$$\varphi = (x_1 \vee x_2) \wedge (x_1 \vee \bar{x}_3) \wedge (\bar{x}_2 \vee \bar{x}_3) \wedge (x_3 \vee x_3) \longrightarrow (x_1 = 1, x_2 = 0, x_3 = 1)$$

Примеры задач, которые можно решить, используя 2-SAT:

• Расположение геометрических объектов без наложений

Если для каждого объекта есть только два варианта расположения, то задача расположения объектов без пересечений сводится к 2-SAT. Пример: есть множество отрезков на плоскости, каждому нужна подпись с одной из двух сторон отрезка.

• 2-List-Coloring (будет разобрана на практике)

Дан неорграф граф, для каждой вершины v есть список $l[v]$ из двух цветов.

Покрасить вершины так, чтобы соседние были покрашены в разные цвета.

Сведение: x_v – номер выбранного цвета в $l[v]$, каждое ребро (a, b) для каждого цвета $c \in l[a] \cap l[b]$ задаёт кюз $\neg(x_a = pos_{a,c} \wedge x_b = pos_{b,c}) \Leftrightarrow (x_a = \overline{pos_{a,c}} \vee x_b = \overline{pos_{b,c}})$

• Кластеризация на два кластера (2-sat можно заменить на раскраску в 2 цвета)

Даны объекты, и матрица расстояний d_{ij} (непохожести объектов). Нужно разбить объекты на два множества: \max из диаметров $\rightarrow \min$. Решение: бинпоиск по ответу, внутри 2-SAT.

2.4.1. Решение 2-SAT за $\mathcal{O}(nm)$

Попробуем жадно сказать $x_1 = 0$. Если был клоз вида $(x_1 = 1 \vee x_i = e)$, то поскольку $x_1 = 0$ мы точно можем сделать вывод $x_i = e$. Осталось увидеть «граф следствий», в котором $2n$ вершин вида $x_i = 0$, $x_i = 1$ и рёбра – выводы, которые мы можем сделать. Сделаем dfs из вершины $x_1 = 0$. Если по ходу такого dfs-а мы посетим две вершины с противоположным смыслом: $x_i = 0$, $x_i = 1$, это противоречие. Если пришли к противоречию $\Rightarrow x_1 = 1$, запустим dfs уже из неё. Если всё равно противоречие \Rightarrow решений нет. Если не пришли к противоречию \Rightarrow все сделанные нами выводы корректны, мнение мы уже не изменим.

Время работы: dfs пускаем не более $2n$ раз, каждый отработает за $\mathcal{O}(m)$.

2.4.2. Решение 2-SAT за $\mathcal{O}(n + m)$

Клоз $(a \vee b) \Leftrightarrow$ двум импликациям $(\neg a \Rightarrow b)$ и $(\neg b \Rightarrow a)$.

Построим граф следствий. Вершины графа – литералы x_i и $\neg x_i$.

Для каждого клоза $(a \vee b)$ проведём рёбра $\neg a \rightarrow b$ и $\neg b \rightarrow a$.

Если n – число неизвестных, m – число клозов, мы получили $2n$ вершин и $2m$ рёбер.

Lm 2.4.1. Решение 2-SAT – для каждого i в графе следствий выбрать ровно одну из двух переменных x_i , $\neg x_i$ так, чтобы не было рёбер из выбранных в невыбранные.

Lm 2.4.2. Путь в графе следствий $x_i \rightsquigarrow \neg x_i$, означает что в любом корректном решении $x_i = 0$.

Следствие 2.4.3. $\exists i: x_i$ и $\neg x_i$ в одной компоненте сильной связности \Rightarrow формула противоречива

Теорема 2.4.4. $\exists i: x_i$ и $\neg x_i$ в одной компоненте сильной связности \Leftrightarrow формула противоречива

Доказательство. Пусть $\forall i x_i$ и $\neg x_i$ в разных компонентах. Алгоритм расстановки значений x_i :

1. Топологически отсортируем конденсацию графа.
2. $k[x_i]$ – номер компоненты в топологическом порядке. \forall ребра графа $a \rightarrow b$ верно $k[a] \leq k[b]$.
3. Для каждой переменной i выберем x_i iff $k[x_i] > k[\neg x_i]$.

Корректность решения:



Если в решение есть противоречие, то по **Lm 2.4.1** есть выбранный y , невыбранный z и ребро $y \rightarrow z$. Рёбра добавлялись парами \Rightarrow есть и ребро $\neg z \rightarrow \neg y$. Из \exists -я этих рёбер делаем вывод: $k[z] \geq k[y] \wedge k[\neg y] \geq k[\neg z]$. Из выбранности y , не выбранности z вывод: $k[y] > k[\neg y] \wedge k[\neg z] > k[z]$.

Итого мы получили решение а $\mathcal{O}(V + E)$: построить граф + найти ксс + вывести ответ. Заметьте, сами компоненты хранить не нужно, мы строим только массив $k[]$.

• Упрощение реализации

Предполагая, что ксс мы ищем через два dfs-а, можно модифицировать второй из них:

```
1 void paintComponent(int v) {
2     used[v] = 1;
3     if (x[v/2] == -1) // пусть вершины  $x_i, \neg x_i$  идут парами; -1 значит неопределено
4         x[v/2] = 1 - v%2; // для  $x_{v/2}$  выберем не  $v$ , так как компоненту  $v$  перебираем раньше
5     for (int x : graph[v])
6         if (!used[x])
7             paintComponent(x);
8 }
```

Лекция #3: Введение в теорию сложности

4–11 февраля 2022

3.1. Основные классы

• Алгоритмически не разрешимые задачи

Почему вообще существуют неразрешимые задачи?

Рассмотрим задачи, на входе которых натуральное число, на выходе — `true` или `false`. Любую такую задачу можно задать подмножеством натуральных чисел, на которых ответ `true`. Например, задача проверки на простоту задаётся множеством \mathbb{P} .

Задач по крайней мере столько, сколько множеств натуральных чисел — $|\mathbb{N}^{\mathbb{N}}|$. Алгоритмов счётное число (то есть $|\mathbb{N}|$), ведь их всех можно пронумеровать: сначала выпишем все однобуквенные, затем все двухбуквенные, и так далее.

$|\mathbb{N}^{\mathbb{N}}| > |\mathbb{N}|$ — тут можно либо сослаться на общую теорему Кантора ($|2^A| > |A|$), либо вспомнить школьные доказательства того, что $|\mathbb{N}^{\mathbb{N}}| = |\mathbb{R}|$ и $|\mathbb{R}| > |\mathbb{N}|$. Так что на самом деле «почти все» задачи неразрешимы.

Посмотрим на конкретные примеры таких задач.

Примеров таких **много**¹, **much enough**², **see also**³.

Можно выделить такую общую тему, как “предсказание будущего сложного процесса”.

Например, “в игре «жизнь» достижимо ли из состояния A состояние B ?”. Или не в игре...

Наиболее каноническим примером является «проблема останова» (halting problem):

Дана программа, остановится ли она когда-нибудь на данном входе?

Теорема 3.1.1. *Halting problem алгоритмически не разрешима.*

Доказательство. От противного. Пусть есть алгоритм `terminates(code, x)`, всегда останавливающийся, и возвращающий `true` iff `code(x)` останавливается. Рассмотрим программу:

```
1 def invert(code):
2   if terminates(code, code): while (true)
```

Если `invert(invert)` останавливается, то должен зависнуть, и наоборот. Противоречие. ■

Ещё есть теорема Успенского-Райса, которая говорит, что любое нетривиальное свойство программ неразрешимо (то есть нет алгоритма, которые бы его проверял). Поясним, что это значит. Будем говорить, что программы \mathcal{A} и \mathcal{B} эквивалентны ($\mathcal{A} \sim \mathcal{B}$), если для каждого входа x либо они обе зависают, либо обе останавливаются и печатают один и тот же ответ (время работы и память при этом могут отличаться).

Свойство программы — это такой предикат $P(\text{code})$, который для любых эквивалентных программ \mathcal{A} и \mathcal{B} выдаёт одно и то же: $\mathcal{A} \sim \mathcal{B} \Rightarrow P(\mathcal{A}) = P(\mathcal{B})$. «Нетривиальное» означает, что хотя бы одна программа ему удовлетворяет, но не все программы. Пример свойства: программа выдаёт только простые числа.

• Decision/search problem

Если в задаче ответ — `true/false`, то это decision problem (задача распознавания).

Иначе это search problem (задача поиска). Примеры:

1. Decision. Проверить, есть ли x в массиве a .

2. Search. Найти позицию x в массиве a .
3. Decision. Проверить, есть ли путь из a в b в графе G .
4. Search. Найти сам путь.
5. Decision. Проверить, есть ли в графе клика размера хотя бы k .
6. Search. Найти максимальный размер клики (или саму клику).

Decision problem f можно задавать, как язык (множество входов) $L = \{x: f(x) = \text{true}\}$.

• DTime, P, EXP (классы для decision задач)

Def 3.1.2. $\text{DTIME}(f(n))$ – множество задач распознавания, для которых $\exists C > 0$ и детерминированный алгоритм, работающий на всех входах не более чем $C \cdot f(n)$, где n – длина входа.

Def 3.1.3. $P = \bigcup_{k>0} \text{DTIME}(n^k)$. Т.е. задачи, имеющие полиномиальное решение.

Def 3.1.4. $\text{EXP} = \bigcup_{k>0} \text{DTIME}(2^{n^k})$. Т.е. задачи, имеющие экспоненциальное решение.

Теорема 3.1.5. (Об иерархии по времени)

$\text{DTIME}(f(n)) \subsetneq \text{DTIME}(f(n) \log^2 f(n))$ для конструктивных по времени f (см. ниже).

Доказательство. Задача, которую нельзя решить за $f(n)$: завершится ли данная программа за $f(n) \log f(n)$ шагов. Подробнее можно прочесть на [wiki](#). Там же дана более сильная формулировка теоремы, требующая большей аккуратности при доказательстве. ■

Теорема выше сформулирована только для конструктивных по времени функций. f конструктивна по времени, если есть алгоритм, который по строке из n единичек вычисляет $f(n)$ за время $\mathcal{O}(f(n))$. Все естественные не слишком маленькие функции ($f(n) \geq n$ – иначе мы не успеем прочитать вход) типа многочленов с неотрицательными коэффициентами и экспонент являются конструктивными по времени.

Следствие 3.1.6. $P \neq \text{EXP}$

Доказательство. $P \subseteq \text{DTIME}(2^n) \subsetneq \text{DTIME}(2^{2^n}) \subseteq \text{EXP}$ ■

3.2. NP (non-deterministic polynomial)

Def 3.2.1. $\text{NP} = \{L: \exists M, \text{ работающий за полином от } |x|, \forall x ((\exists y M(x, y) = 1) \Leftrightarrow (x \in L))\}$

Неформально. NP – класс языков $L: \forall x \in L$, если нам дадут подсказку $y(x)$, мы за полином сможем убедиться, что $x \in L$. Напомним, “язык” \equiv “decision задача”.

Ещё более неформально “ NP – класс задач, к которым ответ можно проверить за полином”.

Подсказку y так же называют свидетелем того, что x лежит в L .

• Примеры NP-задач

1. $\text{HAMPATH} = \{G \mid G \text{ – неорграф, в котором есть гамильтонов путь}\}$. Подсказка y – путь. M получает вход $x = G$, подсказку y , проверяет, что y прост, $|y| = n$ и $\forall (e \in y) e \in G$.
2. $k\text{-CLIQUE}$ – проверить наличие в графе клики размером k . Подсказка y – клика.
3. IS-SORTED – отсортирован ли массив? Она даже лежит в P .

Def 3.2.2. $\text{coNP} = \{L: \bar{L} \in \text{NP}\}$

Def 3.2.3. $\text{coNP} = \{L: \exists M, \text{ работающий за полином от } |x|, \forall x ((\forall y M(x, y) = 0) \Leftrightarrow (x \in L))\}$

Def 3.2.4. $\text{coNP} = \{L: \exists M, \text{ работающий за полином от } |x|, \forall x ((\exists y M(x, y) = 1) \Leftrightarrow (x \notin L))\}$

Неформально: дополнение языка лежит в NP или “ \exists свидетель того, что $x \notin L$ ”.

Пример coNP задачи

PRIME – является ли число простым. Подсказкой является делитель.

На самом деле PRIME \in P, но этого мы пока не умеем понимать.

Замечание 3.2.5. $P \subseteq \text{NP}$ (можно взять пустую подсказку).

Замечание 3.2.6. Вопрос $P = \text{NP}$ или $P \neq \text{NP}$ остаётся открытым ([wiki](#)). Предполагают, что \neq .

Def 3.2.7. BH = BOUNDED-HALTING: вход $x = \langle \underbrace{11\dots 1}_k, M, x \rangle$, проверить, \exists ли такой y : $M(x, y)$ остановится за k шагов и вернёт true.

BH \in NP. Подсказка – такой y . Алгоритм – моделирование k шагов M за $\mathcal{O}(\text{poly}(k))$.

Важно, что если бы число k было записано, используя $\log_2 k$ бит, моделирование работало бы за экспоненту от длины входа, и нельзя было бы сказать “задача лежит в NP”.

3.3. NP-hard, NP-complete

Def 3.3.1. \exists полиномиальное сведение (по Карпу) задачи A к задаче B : $(A \leq_P B) \Leftrightarrow \exists$ алгоритм f , работающий за полином, $(x \in A) \Leftrightarrow (f(x) \in B)$

Замечание 3.3.2. f работает за полином $\Rightarrow |f(x)|$ полиномиально ограничена $|x|$

Def 3.3.3. \exists сведение по Куку задачи A к задаче B : $(A \leq_C B) \Leftrightarrow \exists M$, решающий A , работающий за полином, которому разрешено обращаться за $\mathcal{O}(1)$ к решению B .

Ещё говорят “задача A сводится к задаче B ”.

В обоих сведениях мы решаем задачу A , используя уже готовое решение задачи B .

Другими словами доказываем, что « A не сложнее B ». Различие в том, что в первом случае решением B можно воспользоваться только один раз (и инвертировать ответ нельзя), во втором случае полином раз.

Def 3.3.4. NP-hard = NPh = $\{L: \forall A \in \text{NP} \Rightarrow A \leq_P L\}$

NP-трудные задачи – класс задач, которые не проще любой задачи из класса NP.

Def 3.3.5. NP-complete = NPC = NPh \cap NP

NP-полные задачи – самые сложные задачи в классе NP.

Если мы решим хотя бы одну из NPC за полином, то решим все из NP за полином.

Хорошая новость: все NP-полные по определению сводятся друг к другу за полином.

Замечание 3.3.6. Когда хотите выразить мысль, что задача трудная в смысле решения за полином (например, поиск гамильтонова пути), **неверно** говорить “это NP задача” (любая из P тоже в NP) и странно говорить “задача NP-полна” (в этом случае вы имеете в виду сразу, что и трудная, и в NP). Логично сказать “задача NP-трудна”.

Lm 3.3.7. $A \leq_P B, B \in P \Rightarrow A \in P$

Доказательство. Сведение f работает за n^s , B решается за $n^t \Rightarrow A$ решается за n^{st} . ■

Lm 3.3.8. $A \leq_P B, A \in NPh \Rightarrow B \in NPh$

Доказательство. $\forall L \in NP (\exists f: L$ сводится к A функцией $f(x)) \wedge (A \leq_P B$ функцией $g(x)) \Rightarrow L$ сводится к B функцией $g(f(x))$ (за полином). ■

• NP-полные задачи существуют!

Приведём простую и очень важную теорему. На экзамене доказательство можно сформулировать в одно предложение, здесь же оно для понимания расписано максимально подробно.

Теорема 3.3.9. $ВН = BOUNDED-HALTING \in NP_c$

Доказательство. $NP_c = NP \cap NPh$. Мы уже показали [Def 3.2.7](#), что ВН лежит в NP.

Теперь покажем, что $ВН \in NPh$. Для этого нужно взять $L \in NP$ и свести его к ВН.

Пусть $L \in NP \Rightarrow \exists$ полиномиальный M , проверяющий подсказки для L .

Полиномиальный $\Leftrightarrow \exists P(n)$, ограничивающий время работы M . $(x \in L) \Leftrightarrow \exists y M(x, y) = 1$.

Программа M всегда отработывает за $P(|x|)$, если запустить её с будильником $P(|x|)$, она не поменяется. Рассмотрим $f(x) = \langle \underbrace{11 \dots 1}_{P(|x|)}, M, x \rangle$.

Получили полиномиальное сведение: $(x \in L) \Leftrightarrow (\exists y M(x, y) = 1) \Leftrightarrow (f(x) \in ВН)$.

Заметьте, зная L , мы не умеем предъявить ни M , ни f , мы лишь знаем, что $\exists M, f$. ■

3.4. Сведения, новые NP-полные задачи

Началось всё с того, что в 1972-м Карп опубликовал список из 21 полной задачи, и дерево сведений. Кстати, в [его работе \(pdf\)](#) все сведения крайне лаконичны. Итак, приступим:

Чтобы доказать, что $B \in NPh$, нужно взять любую $A \in NPh$ и свести A к B полиномиально. Пока такая задача A у нас одна – ВН. На самом деле их очень **много**.

Чтобы доказать, что $B \in NP_c$, нужно ещё не забыть проверить, что $B \in NP$.

Во всех теоремах ниже эта проверка очевидна, мы проведём её только в доказательстве первой.

• $ВН \rightarrow CIRCUIТ-SAT \rightarrow SAT \rightarrow 3-SAT \rightarrow k-INDЕPЕNDЕNT \rightarrow k-CLIQUE$

Def 3.4.1. CIRCUIТ-SAT. Дана схема, состоящая из входов, выходов, гейтов AND, OR, NOT. Проверить, существует ли набор значений на входах, дающий true на выходе.

Теорема 3.4.2. $CIRCUIТ-SAT \in NP_c$

Доказательство. Подсказка – набор значений на входах $\Rightarrow CIRCUIТ-SAT \in NP$.

Сводим ВН к CIRCUIТ-SAT \Rightarrow нам даны программа M , время выполнения t , вход x .

За время t программа обратится не более чем к t ячейкам памяти.

Обозначим за $s_{i,j}$ состояние true/false j -й ячейки памяти в момент времени i .

$s_{0,j}$ – вход, $s_{t,output}$ – выход, $\forall i \in [1, t)$ $s_{i,j}$ зависит от $\mathcal{O}(1)$ переменных $(i-1)$ -го слоя.

Сейчас значение s_{ij} – произвольная булева формула f_{ij} от $\mathcal{O}(1)$ переменных из слоя s_{i-1} .

Перепишем f_{ij} в КНФ-форме, чтобы получить гейты вида AND, OR, NOT. Получили $\mathcal{O}(tn)$ булевых гейтов \Rightarrow по (M, t, x) за полином построили вход к CIRCUIТ-SAT. ■

Теорема 3.4.3. $SAT \in NPC$

Доказательство. В разборе практики смотрите сведение из CIRCUIT-SAT. ■

Теорема 3.4.4. $3-SAT \in NPC$

Доказательство. Пусть есть кюз $(x_1 \vee x_2 \vee \dots \vee x_n), n \geq 4$.

Введём новую переменную w и заменим его на $(x_1 \vee x_2 \vee w) \wedge (x_3 \vee \dots \vee x_n \vee \bar{w})$. ■

Теорема 3.4.5. $k-INDPENDENT \in NPC$

Доказательство. Наша формула – m кюзов $(l_{i1} \vee l_{i2} \vee l_{i3})$, где l_{ij} – литералы.

Построим граф из ровно $3m$ вершин – l_{ij} . $\forall i$ добавим треугольник (l_{i1}, l_{i2}, l_{i3}) (итого $3m$ рёбер).

В любое независимое множество входит максимум одна вершина из каждого треугольника.

$\forall k = 1..n$ соединим все вершины $l_{ij} = x_k$ со всеми вершинами $l_{ij} = \bar{x}_k$.

Теперь $\forall k = 1..n$ в независимое множество нельзя одновременно включить x_k и \bar{x}_k .

Итог: \exists независимое размера $m \Leftrightarrow$ у 3-SAT было решение. ■

Теорема 3.4.6. $k-CLIQUE \in NPC$

Доказательство. Есть простое двустороннее сведение $k-CLIQUE \leftrightarrow k-INDPENDENT$.

c_{ij} – есть ли ребро между i и j вершинами. Создадим новый граф: $c'_{ij} = \bar{c}_{ij} \wedge (i \neq j)$. ■

3.5. Задачи поиска

Def 3.5.1. \overline{NP} , \overline{NPC} , \overline{NPh} – аналогичные классы для задач поиска подсказки.

- Сведение задач минимизации, максимизации к decision задачам

Пусть мы умеем проверять, есть ли в графе клика размера k .

Чтобы найти размер максимальной клики, достаточно применить бинарный поиск по ответу.

Это общая техника, применимая для максимизации/минимизации численной характеристики.

- Сведение search задач к decision задачам

Последовательно фиксируются биты (части) подсказки y .

Пример: выполняющий набор для SAT.

Пусть $M(\varphi)$ проверяет выполнимость формулы φ .

$\varphi[x_i=e]$ – формула, полученная из φ подстановкой значения e в x_i . Индукция:

while $n > 0$:

 if $M(\varphi[x_n=0]) = 1$ then $r_n = 0$ else $r_n = 1$

$\varphi \leftarrow \varphi[x_n=r_n]$; $n--$

Пример: $k-INDPENDENT$.

Взять или выкинуть первую вершину?

if $(G \setminus \{1\}) \in k-INDPENDENT$ then $G \setminus = \{1\}$

else $k--$, $ans \cup = \{1\}$, $G \setminus = \{N(1)\}$ (выкинуть соседей 1-й вершины)

- Решение NP-полных задач

Пусть вам дана NP-полная задача. С одной стороны плохо – для неё нет быстрого решения.

С другой стороны её можно свести к SAT, для которого несколько десятилетий успешно оптимизируются специальные SAT-solvers. Например, вы уже можете решать $k-CLIQUE$, построив

вход к задаче SAT и скормить его python3 пакету **pycosat**.

А ещё можно принять участие в **соревновании**.

3.6. Гипотезы

Гипотеза 3.6.1. $P \neq NP$

Гипотеза 3.6.2. ETH (exponential time hypothesis): \nexists решения за $2^{o(n)}$ для 3-SAT

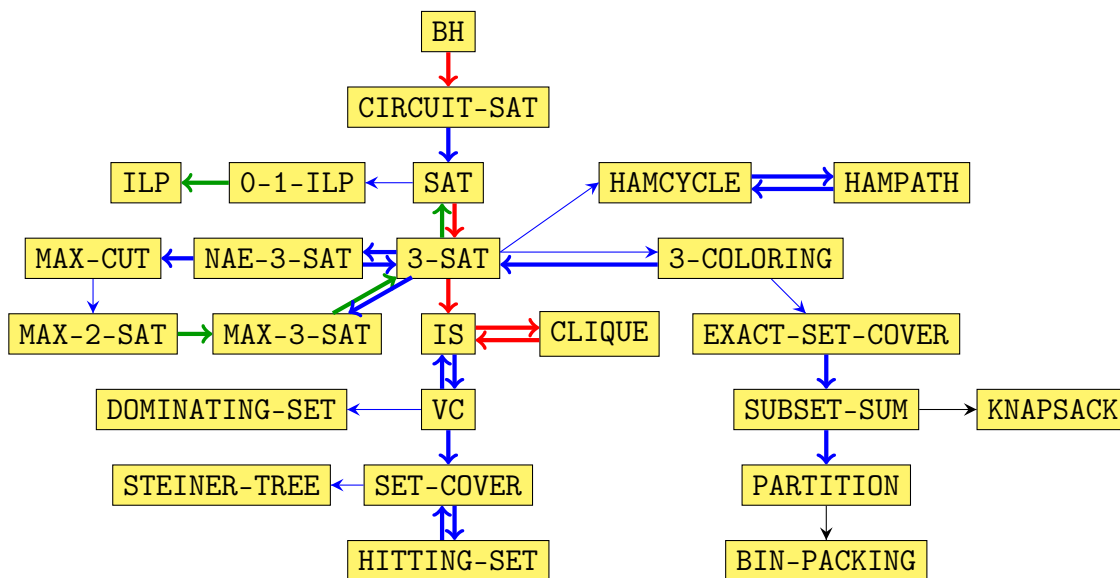
Гипотеза 3.6.3. SETH (strong ETH): $\forall \epsilon \exists k: \nexists$ решения за $(2 - \epsilon)^n$ для k -SAT

Lm 3.6.4. SETH \Rightarrow ETH $\Rightarrow P \neq NP$

3.7. Дерево сведений

Легенда:

- **Красное** – было на лекции.
- **Синее** – было в задачах практики/дз, нужно знать на экзамене.
- **Бледносинее** – было в задачах практики/дз, не будет на экзамене.
- **Зелёное** – очевидное вложение.
- Чёрное – будет доказано и использовано в будущем.



Лекция #4: (*) Дополнительная сложность

29 января 2021

4.1. (*) Алгоритм Левина

Есть универсальный алгоритм, который для любой задачи из \overline{NP} работает за время близкое к оптимальному. Рассмотрим любую задачу поиска $L \in \overline{NP}$.

Пусть $M(x, y)$ – полиномиальный чекер для L . Рассмотрим любой корректный алгоритм A , решающий L . $\exists S_A$ – программа на питоне, имплементирующая A . S_A – строка, то есть, число в 256-ичной системе счисления. Рассмотрим соответствующее ей число N_A .

N_A – константа. Пусть $t(|x|)$ – время работы A .

Будем перебирать все программы и время работы.

```

1 solve(x):
2     for (T = 1;; T *= 2) // время работы T
3         for (P = 1; 2^P <= T; P++) // программа P
4             y = call(P, T, x) // запускаем T шагов программы P на входе x
5             if M(x, y)
6                 return y

```

Если решений нет, наш алгоритм успешно зависнет. Далее обозначим $|x| = n$.

Если решение \exists , как только $T \geq \max(2^{N_A}, t(n)) = F = \Theta(t(n))$, мы найдём решение.

Оценим время работы.

Одна итерация внешнего `for` работает за $\mathcal{O}(M \cdot T \log T)$. Суммарное время:

$\sum_T (M \cdot T \log T) = \Theta(M \cdot F \log F) = \Theta(M \cdot t(n) \log t(n))$ (помним, что 2^{N_A} – просто константа).

Итого, мы получили алгоритм, который для любой \overline{NP} задачи при существовании решения находит его за время $\Theta(M \cdot t \log t)$, где t – время работы оптимального алгоритма.

Например, для гамильтонова пути это будет $\Theta(V \cdot t \log t)$.

Полученный алгоритм называется *алгоритмом Левина*.

Следствие 4.1.1. Если $P = NP$, то алгоритм Левина решает все \overline{NP} задачи за полином.

4.2. (*) Расщепление

Мы будем решать задачу поиска максимального независимого множества (IS). Но, конечно, все наши идеи подходят также для максимальной клики и минимального вершинного покрытия.

Итак, простейший алгоритм расщепления:

берём вершину v и или берём её в независимое множества (первая ветка рекурсии), или не берём (вторая ветка). Берём \Rightarrow удалим v и соседей. Не берём \Rightarrow удалим v .

Время работы $T(n) = T(n - \deg[v] - 1) + T(n - 1) \Rightarrow$ выгодно взять $v: \deg[v] = \max$.

Осталось заметить

1. Если \exists вершина степени 1 её выгодно жадно взять в IS.
2. Если \exists вершина v степени 2. Если мы берём ровно одного из её соседей, можно заменить его на $v \Rightarrow$ есть только два разумных варианта – или взять v , или её обоих соседей \Rightarrow модифицируем наш граф: удалим v , сдвинем её соседей a, b в одну вершину ab . Вариант «брать v » перешёл в «не брать ab », вариант «брать a, b » перешёл в «брать ab ».

Ответ уменьшился ровно на 1, а число вершин на 2. При оценке времени нам важно будет, что как только появляется вершина степени 2, сразу $n \rightarrow n - 2$.

Остался случай $\forall v \deg[v] \geq 3$. Уже получили асимптотику $T(n) = T(n-1) + T(n-4) \leq 1.3803^n$. В дальнейшем анализе рекурренты $T(n-a_1) + T(n-a_2) + \dots$ будем обозначать $[a_1, a_2, \dots]$.

Замечание 4.2.1. На самом деле все асимптотики α^n нужно записывать как $\mathcal{O}^*(\alpha^n)$, так как на каждом шаге присутствует ещё полиномиальное время на выбор нужного случая и, возможно, модификацию графа. Сейчас мы сосредоточены на минимизации α , про полином не думаем.

• Разбор случаев $\deg[v] \geq 3$

3а. $\forall v \deg[v] = 3 \Rightarrow$ после расщепления сразу попадаем в случай (2): $\exists v \deg[v] = 2$, итого $T(n) \leq T(n-1-2) + T(n-1-3-2) = [3, 6] \leq 1.174^n$. На самом деле появится сразу много вершин степени 2, оценку можно сильно улучшить, нам хватит такой.

3б. $\exists v \deg[v] = 3 \Rightarrow$ по непрерывности $\exists u \deg[u] \geq 4 \wedge u$ и u есть сосед степени 3. Расщепляемся по u : $T(n) \leq T(n-1-2) + T(n-1-\deg[u]) = [3, 5] \leq 1.194^n$. Из случаев (3б) и (3а) хуже (3б) \Rightarrow если появляются вершины степени 3, можно считать, что это именно (3б).

4а. $\forall v \deg[v] = 4 \Rightarrow$ после расщепления сразу попадаем в случай (3б): $\exists v \deg[v] = 3$, итого $T_{4a}(n) \leq T_{3b}(n-1) + T_{3b}(n-5) = [1+3, 1+5, 5+3, 5+5] = [4, 6, 8, 10] \leq 1.239^n$.

4б. $\exists v \deg[v] = 4 \Rightarrow$ по непрерывности $\exists u \deg[u] \geq 5 \wedge u$ и u есть сосед степени 4. Расщепляемся по u : $T_{4b}(n) \leq T_{3b}(n-1) + T(n-6) = [4, 6, 6] \leq 1.2335^n$.

5а. $\forall v \deg[v] = 5 \Rightarrow$ после расщепления сразу попадаем в случай (4б): $\exists v \deg[v] = 4$, итого $T(n) \leq T_{4b}(n-1) + T_{4b}(n-6) = [5, 7, 7, 10, 12, 12] \leq 1.2487^n$.

Поскольку $T_{4a} \geq T_{4b}$, важно, что именно 4б, это точно так, если $n > 1 + 5 + 25$, то есть, не $\mathcal{O}(1)$.

5б. $\exists v \deg[v] = 5 \Rightarrow$ по непрерывности $\exists u \deg[u] \geq 6 \wedge u$ и u есть сосед степени 5. Расщепляемся по u : $T(n) \leq T_{4b}(n-1) + T(n-7) = [5, 7, 7, 7] \leq 1.2413^n$.

6а. $\forall v \deg[v] = 6 \Rightarrow$ после расщепления сразу попадаем в случай (5б): $\exists v \deg[v] = 5$, итого $T(n) \leq T_{5b}(n-1) + T_{5b}(n-7) = [6, 8, 8, 8, 12, 14, 14, 14] \leq 1.24499^n$.

7. $\exists v \deg[v] \geq 7 \Rightarrow T(n) \leq T(n-1) + T(n-8) = [1, 8] \leq 1.23206^n$

Итого время работы = $\max\{T_{3a}, T_{3b}, T_{4a}, T_{4b}, T_{5a}, T_{5b}, T_{6a}, T_7\} = T_{5a} = [5, 7, 7, 10, 12, 12] \leq 1.2487^n$
P.S. Заметим, что $[1, 7] \approx 1.2554^n \Rightarrow$ все части анализа необходимы.

• Алгоритм

Анализ сложен. А алгоритм прост. Выбираем ребро (a, b) : $\deg[a] = \min, \deg[b] = \max$.

В первую очередь минимизируем $\deg[a]$, при равных максимизируем $\deg[b]$.

Если $\deg[a] \leq 2$, применяем жадность, иначе расщепляемся по b .

Наша оценка 1.2487^n не точна, её можно улучшать и улучшать. Например, на самом деле $T_{5a} \approx \max(T_{4b}, T_{4a}, T_{3b}, T_{4a})$, так как в обеих ветках рекурсии мы очень не скоро получим что-то кроме этих четырёх случаев. Но получим, в (2) степени увеличиваются, $\deg_{ab} = \deg_a + \deg_b - 2$.

4.3. (*) Оценка с использованием весов

Повторим анализ. Сперва более простую версию.

1. Если \exists вершина степени 1 её выгодно жадно взять в IS.

2. $\forall v \deg[v] = 2 \Rightarrow$ пути и циклы, задача решается жадно.

3. $\exists v \deg[v] \geq 3 \Rightarrow T(n) = T(n-1) + T(n-\deg-1) \leq T(n) = T(n-1) + T(n-4) \leq 1.3803^n$

Новая идея: сопоставим вершинам разной степени разные веса.

$w_1 = 0, w_2 = 0.5, w_3 = 1, w_4 = 1, \dots$ (веса больших степеней тоже 1).

Для вершин степени 1 мы по-прежнему планируем запускать жадность.

$$a) k = \sum_v w_{deg[v]}[v] \leq n$$

Заметим: b) $k = 0 \Rightarrow$ граф пуст, ответ известен

$$c) w_i \leq w_{i+1} \Rightarrow \text{после любых расщеплений } k \searrow$$

Заменяем число вершин n на суммарный вес вершин k . $k \leq n \Rightarrow T(k) \leq T(n)$.

3. $\max_v deg[v] = 3 \Rightarrow$ расщепляемся по 3-ке. Степень её соседей уменьшится, их вес уменьшится на 0.5 и в случае $3 \rightarrow 2$, и в случае $2 \rightarrow 1$. В ветке с удалением соседей то же произойдёт с их соседями. Итого $T(k) \leq T(k-1-3 \cdot 0.5) + T(k-1-5 \cdot 0.5) = [2.5, 3.5] \leq 1.263^k$.

4. $\max_v deg[v] = 4 \Rightarrow$ расщепляемся по 4-ке. Разберём 2 случая.

Все соседи 3-ки: $T(k) \leq T(k-1-4 \cdot 0.5) + T(k-1-4 \cdot 0.5) = [3, 3] \leq 1.25993^k$.

Все соседи 4-ки: $T(k) \leq T(k-1) + T(k-1-4) = [1, 5] \leq 1.3248^k$.

Случаев больше, их разбор здесь не представлен, худшим из них будет «все 4-ки».

• Применяем идею весов по полной.

1. Если \exists вершина степени 1 её выгодно жадно взять в IS.

2а. Если \exists две смежных вершины a, b степени 2, заменим $u - a - b - v$ на $u - v$.

2б. Если \exists вершина степени 2, удалим её, стянем её соседей.

Если их степени были 3, 3, новая вершина имеет степень $3+3-2 = 4$.

Подберём теперь веса. Самые простые: $w_1 = w_2 = 0, w_3 = 0.8, w_4 = 1, w_5 = 1, \dots$

Заметим, что попадая в состояние $\exists v deg[v] \leq 2$, мы сразу уменьшим k хотя бы на $2w_3 - w_4 = 0.6$.

3а. $\forall v deg[v] = 3 \Rightarrow T(k) \leq T_2(k-4 \cdot w_3) + T_2(k-6 \cdot w_3) = [3.2+0.6, 4.8+0.6] = [3.8, 5.4] \leq 1.165^k$.

3б. $\exists v deg[v] = 3$ и максимальный сосед тройки это 4. Расщепляемся по соседу. Случай.

Соседи 3333: ...

Соседи 3444: ...

3с. $\exists v deg[v] = 3$ и максимальный сосед ≥ 5 . Расщепляемся по соседу. Случай.

Соседи 33333: ...

Соседи 34444: ...

Соседи 35555: ...

4. ...

Как вы понимаете, разбор случаев весьма трудоёмкий. Может быть, его можно как-то запрограммировать? =)

Ссылки: [\[wiki\]](#) [\[Robson'2001, 1.1888ⁿ\]](#)

Кстати, $1.1888^n \approx 2^{n/4}$, то есть, для $n \approx 100$ алгоритм точно не плох.

Лекция #5: Рандомизированные алгоритмы

11–17 февраля 2022

5.1. Случайные числа в C++ Немного про rand()

Начиная с C++11 для генерации случайных чисел принято использовать **вихрь Мерсенна**:

```

1 mt19937 gen;
2 for (int i = 0; i < n; i++) // n целых чисел из [0, M)
3   a[i] = gen() % M; // gen() даст 32-битное целое случайное число.

```

Более подробный пример про случайные числа в C++ можно посмотреть [здесь](#)

Обратите внимание, что функциями `rand`, `srand`, `random_shuffle` уже давно не пользуются. Одна из причин: `rand()` возвращает псевдослучайное число от 0 до `RAND_MAX`, под Windows `RAND_MAX = 2^{15} - 1 = 32767`, что слишком мало, а в Linux `RAND_MAX = 2^{31} - 1`, что убивает потенциальную кроссплатформенность.

5.2. Определения: RP, coRP, ZPP

Рандомизированными называют алгоритмы, использующие случайные биты.

Первый тип алгоритмов: решающие decision задачи, работающие всегда за полином, ошибающиеся в одну сторону. Строго это можно записать так:

Def 5.2.1. $RP = \{L : \exists M \in PTime \left\{ \begin{array}{l} x \notin L \Rightarrow M(x, y) = 0 \\ x \in L \Rightarrow \Pr_y[M(x, y) = 1] \geq \frac{1}{2} \end{array} \right\}$

x – вход, y – подсказка из случайных бит. Расшифровка: RP = randomized polynomial time. То есть, если $x \notin L$, M не ошибается, иначе работает корректно с вероятностью хотя бы $\frac{1}{2}$. Заметим, если для какого-то y алгоритм M вернул 1, то это точно правильный ответ, $x \in L$.

Def 5.2.2. $coRP = \{L : \exists M \in PTime \left\{ \begin{array}{l} x \in L \Rightarrow M(x, y) = 1 \\ x \notin L \Rightarrow \Pr_y[M(x, y) = 0] \geq \frac{1}{2} \end{array} \right\}$

Заметим, если для какого-то y алгоритм M вернул 0, то это точно правильный ответ, $x \notin L$.

• Сравнение классов NP, RP

Если ответ 0, оба алгоритма для \forall подсказки выдадут 0. Если ответ 1, для NP-алгоритма \exists хотя бы одна подсказка, а RP-алгоритм должен корректно работать хотя бы на половине подсказок.

• Понижение ошибки

Конечно, алгоритм, ошибающийся с вероятностью $\frac{1}{2}$ никому не нужен.

Lm 5.2.3. Пусть M – RP-алгоритм, ошибающийся с вероятностью p .

Запустим его k раз, если хотя бы раз вернул 1, вернём 1. Получили алгоритм с ошибкой p^k .

Например, если повторить 100 раз, получится вероятность ошибки $2^{-100} \approx 0$.

Если есть алгоритм, *корректно* работающий с близкой к нулю вероятностью p , ошибающийся с вероятностью $1 - p$, то повторив его $\frac{1}{p}$ раз, получим вероятность ошибки $(1 - p)^{1/p} \leq e^{-1}$.

• ZPP (zero-error probabilistic polynomial time)

ZPP – класс задач, для которых есть никогда не ошибающийся вероятностный алгоритм с полиномиальным матожиданием временем работы.

Def 5.2.4. $ZPP = \{L: \exists P(n), M \text{ такие, что } E_y[Time(M(x, y))] \leq P(|x|)\}$

Про алгоритмы и для RP, и для ZPP говорят “вероятностные/рандомизированные”.

Чтобы подчеркнуть принадлежность классу, уточняют ошибку.

RP-алгоритм обладает односторонней ошибкой. ZPP-алгоритм работает без ошибки.

Мы определили основные классы только для задач распознавания, только для полиномиального времени. Можно определить аналогичные для задач поиска и для любого времени.

Важно помнить, что в RP, coRP, ZPP алгоритмы работают на *всех* тестах.

Например, вероятность ошибки в RP для любого теста не более $\frac{1}{2}$.

5.3. Примеры

Все наши примеры про поиск и не за «полином», а гораздо быстрее.

• Часто встречающееся число

Поиск числа, которое встречается в массиве больше половины раз.

У решения этой задачи есть две версии.

ZPP-версия: «мы уверены, что такой элемент есть, пробуем случайные, пока не попадём».

RP-версия: «хотим определить, есть ли такой элемент в массиве, делаем одну пробу».

• Поиск квадратичного невычета

Для любого простого p ровно $\frac{p-1}{2}$ ненулевых остатков обладают свойством $a^{\frac{p-1}{2}} \equiv -1 \pmod{p}$. Такие a называются «квадратичными невычетами».

Задача: дано p , найти невычет.

Алгоритм: пока не найдём, берём случайное a , проверяем.

Время одной проверки — возведение в степень ($\log p$ для $p \leq 2^w$).

Сколько в среднем понадобится проверок? $E = 1 + \frac{1}{2}E \Rightarrow E = 2$.

Итого ZPP-версия работает за $\mathcal{O}(\log p)$.

5.4. Проверка на простоту

• Тест Ферма

Малая теорема Ферма: $\forall p \in \mathbb{P} \forall a \in [1..p-1] \ a^{p-1} \equiv 1 \pmod{p}$.

Чтобы проверить простоту p , можно взять случайное a и проверить $a^{p-1} \equiv 1 \pmod{p}$

Lm 5.4.1. $\exists a: a^{p-1} \not\equiv 1 \pmod{p} \Rightarrow$ таких a хотя бы $\frac{p-1}{2}$.

Доказательство. Пусть b такое, что $b^{p-1} \equiv 1 \pmod{p} \Rightarrow (ab)^{p-1} = a^{p-1} b^{p-1} \not\equiv 1 \pmod{p}$. ■

Мы показали, что если тест вообще работает, то с вероятностью хотя бы $\frac{1}{2}$.

К сожалению, есть составные числа, для которых тест вообще не работает — числа Кармайкла.

Утверждение 5.4.2. У числа Кармайкла a есть простой делитель не более $\sqrt[3]{a}$.

Получаем следующий всегда корректный вероятностный алгоритм:

проверим возможные делители от 2 до $\sqrt[3]{a}$, далее тест Ферма.

• Тест Миллера-Рабина

```

1 bool isPrime(int p):
2     // p-1=2^st, t -- нечётно
3     if (p < 2) return 0; // 1 и 0 не простые
4     p → (s, t)
5     a = randInt(2, p-1) // [2, p-1]
6     g = pow(a, t, p) // возвели в степень по модулю
7     for (int i = 0; i < s; i++) {
8         if (g == 1) return 1;
9         if (g % p != -1 && g * g % p == 1) return 0;
10        g = g * g % p;
11    }
12    return g == 1; // тест Ферма

```

В строке 8 мы проверяем, что нет корней из 1 кроме 1 и -1 .

Время работы $\mathcal{O}(\log p)$ операций «умножение по модулю».

Утверждение 5.4.3. $\forall a$ вероятность ошибки не более $\frac{1}{4}$.

• Изученные алгоритмы

(*) k -я порядковая статистика учит нас:

“вместо того, чтобы выбирать медиану, достаточно взять случайный элемент”.

(*) 3-LIST-COLORING: вычеркнуть из каждого клона случайный элемент, получить 2-SAT.

Вероятность успеха этого алгоритма $\frac{2^n}{3^n}$, чтобы получить вероятность $\frac{1}{2}$, нужно было бы повторить его $1.5^n \ln 2$ раз, поэтому к RP-алгоритмам он не относится.

Ещё про него полезно понимать “работает на 2^n подсказках из 3^n возможных”.

• (*) Проверка $AB = C$

Даны три матрицы $n \times n$, нужно за $\mathcal{O}(n^2)$ проверить равенство $AB = C$. Все вычисления в \mathbb{F}_2 .

Вероятностный алгоритм генерирует случайный вектор x и проверяет $A(Bx) = Cx$.

Lm 5.4.4. Вероятность ошибки не более $\frac{1}{2}$.

Доказательство. Пусть $AB \neq C \Rightarrow D = AB - C \neq 0$.

Посчитаем $\Pr_x[A(Bx) = Cx] = \Pr_x[Dx = 0]$. $\exists i, j: D_{i,j} \neq 0$. Зафиксируем произвольный x .

$(Dx)_i = D_{i,0}x_0 + \dots + D_{i,j}x_j + \dots + D_{i,n-1}x_{n-1}$. $D_{i,j} \neq 0 \Rightarrow$ меняя значение x_j , меняем значение

$(Dx)_i \Rightarrow \Pr_x[(Dx)_i = 0] = \frac{1}{2} \Rightarrow \Pr_x[Dx = 0] \leq \frac{1}{2}$. ■

5.5. ZPP = RP \cap coRP

Lm 5.5.1. Неравенство Маркова: $x \geq 0 \Rightarrow \forall a \Pr[x > aE(x)] < \frac{1}{a}$

Доказательство. Пусть $\Pr[x > aE(x)] \geq \frac{1}{a}$, тогда матожидание должно быть больше чем

произведение $aE(x)$ и $\frac{1}{a}$ (так как «есть хотя бы $\frac{1}{a}$ x -ов со значением $aE(x)$). Формально: $\Rightarrow E(x) > (aE(x)) \cdot \frac{1}{a} + 0 \cdot (1 - \frac{1}{a}) = E(x)$!? ■

Следствие 5.5.2. $x \geq 0 \Rightarrow \forall a \Pr[x \leq aE(x)] \geq \frac{1}{a}$

Теорема 5.5.3. ZPP = RP \cap coRP

Доказательство. Пусть $L \in \text{ZPP} \Rightarrow \exists$ алгоритм M , работающий в среднем за $P(n)$ $\xRightarrow{\text{Cons 5.5.2}}$

$\Pr[Time(M) \leq 2P(n)] \geq \frac{1}{2} \Rightarrow$ запустим M с будильником на $2P(n)$ операций. Если алгоритм завершился до будильника, он даст верный ответ, иначе вернём 0 \Rightarrow RP или 1 \Rightarrow coRP.

Пусть $L \in RP \cap coRP \Rightarrow \exists M_1(RP), M_2(coRP)$, работающие за полином, ошибающиеся в разные стороны, которые можно запускать по очереди ($M_1 + M_2$). Пусть $x \notin L \Rightarrow M_1(x) = 0$, $\Pr[M_2(x) = 0] \geq \frac{1}{2}$. Сколько раз нужно в среднем запустить $M_1 + M_2$? $E \geq 1 + \frac{1}{2}E \geq 2$ (всегда запустим 1 раз, затем с вероятностью $\frac{1}{2}$ получим $M_2(x) = 1$ и повторим процесс с начала). ■

Lm 5.5.4. $P \subseteq coRP \cap RP = ZPP \subseteq RP \subseteq NP$

При этом строгие ли вложения неизвестно (про все три вложения).

Зато известна масса задач, для которых есть простое RP-решение, но неизвестно P-решение.

5.6. Двусторонняя ошибка, класс BPP

На практике такие алгоритмы нам вряд ли встретятся, но всё же они есть:

Def 5.6.1. $BPP = \{L: \exists M \in PTime \left\{ \begin{array}{l} x \notin L \Rightarrow \Pr_y[M(x, y) = 0] \geq \alpha \\ x \in L \Rightarrow \Pr_y[M(x, y) = 1] \geq \alpha \end{array} \right\}, \text{ где } \alpha = \frac{2}{3}\}$.

Другими словами алгоритм всегда даёт корректный ответ с вероятностью хотя бы $\frac{2}{3}$.

Для BPP тоже можно понижать ошибку: запустим n раз, выберем ответ, который чаще встречается (majority). На самом деле α в определении можно брать сколь угодно близким к $\frac{1}{2}$.

Lm 5.6.2. *О понижении ошибки.* Пусть $\beta > 0, \alpha = \frac{1}{2} + \varepsilon, \varepsilon > 0 \Rightarrow \exists n = poly(\frac{1}{\varepsilon})$ такое, что повторив алгоритм n раз, и вернув majority, мы получим вероятность ошибки не более β .

(*) *Доказательство.* Пусть $x \in L$, мы повторили алгоритм $n = 2k$ раз.

Если получили 1 хотя бы $k+1$ раз, вернём 1, иначе вернём 0.

Вероятность ошибки $Err = \sum_{i=0}^{i=k} p_i$, где p_i – вероятность того, что алгоритм ровно i раз вернул 1.

Выпишем в явном виде формулу для p_i , при этом удобно отсчитывать i от середины: $i = k - j$.

$$p_i = p_{k-j} = \binom{n}{i} \left(\frac{1}{2} + \varepsilon\right)^{k-j} \left(\frac{1}{2} - \varepsilon\right)^{k+j} = \binom{n}{i} \left(\frac{1}{4} - \varepsilon^2\right)^k \left(\frac{1/2-\varepsilon}{1/2+\varepsilon}\right)^j \leq \binom{n}{i} \left(\frac{1}{4} - \varepsilon^2\right)^k$$

$$Err = \sum_{i=0}^{i=k} p_i \leq \sum_{i=0}^{i=k} \binom{n}{i} \left(\frac{1}{4} - \varepsilon^2\right)^k = \frac{1}{2} 2^n \left(\frac{1}{4} - \varepsilon^2\right)^k = \frac{1}{2} (1 - 4\varepsilon^2)^k$$

Хотим, чтобы ошибка была не больше β , для этого возьмём $k = (1/4\varepsilon^2) \cdot \ln \frac{1}{\beta} = poly(\frac{1}{\varepsilon})$. ■

5.7. Как ещё можно использовать случайные числа?

• Идеальное кодирование

Алиса хочет передать Бобу некое целое число x от 0 до $m - 1$. У них есть общий ключ r от 0 до $m - 1$, сгенерированный равномерным случайным распределением. Тогда Алиса передаст по открытому каналу $y = (x + r) \bmod m$. Знание **одного** такого числа не даст злоумышленнику ровно никакой информации. Боб восстановит $x = (y - r) \bmod m$.

• Вычисления без разглашения

В некой компании сотрудники стесняются говорить друг другу, какая у кого зарплата.

Зарплату i -го обозначим x_i . Сотрудники очень хотят посчитать среднюю зарплату \Leftrightarrow посчитать сумму $x_1 + \dots + x_n$. Для этого они предполагают, что сумма меньше $m = 10^{18}$, и пользуются следующим алгоритмом:

0. Первый сотрудник генерирует случайное число $r \in [0, m)$.
1. Первый сотрудник передаёт второму число $(r + x_1) \bmod m$.
2. Второй сотрудник передаёт третьему число $(r + x_1 + x_2) \bmod m$.
3. ...
- n. Последний сотрудник передаёт первому $(r + x_1 + x_2 + \dots + x_n) \bmod m$.
- *. Первый, как единственный знающий r , вычитает его и говорит всем ответ.

• **Приближения**

На практике будут разобраны $\frac{1}{2}$ -ОПТ приближения для MAX-3-SAT и VERTEX-COVER.

5.8. Парадокс дней рождений. Факторизация: метод Полларда

«В классе 27 человек \Rightarrow с большой вероятностью у каких-то двух день рождения в один день»
 Пусть p_k – вероятность, что среди k случайных чисел от 1 до n все различны. Оценим p_k **снизу**.

Lm 5.8.1. $1 - p_k \leq \frac{k(k-1)}{2n}$

Доказательство. $f = \#\{(i, j) : x_i = x_j\}$, $1 - p_k = \Pr[f > 0] \leq E_k[f] = \frac{k(k-1)}{2} \cdot \Pr[x_i = x_j] = \frac{k(k-1)}{2} \frac{1}{n}$ ■

Следствие 5.8.2. При $k = o(\sqrt{n})$ получаем $1 - p_k = o(1) \Rightarrow$ с вероятностью ≈ 1 все различны.
 Теперь оценим p_k при $k = \sqrt{n}$.

Lm 5.8.3. $k = \sqrt{n} \Rightarrow 0.4 \leq p_k \leq 0.8$

Доказательство. $p_k = \frac{n}{n} \cdot \frac{n-1}{n} \cdot \dots \cdot \frac{n-\sqrt{n}/2}{n} \cdot \frac{n-\sqrt{n}/2-1}{n} \cdot \dots \cdot \frac{n-\sqrt{n}}{n} \Rightarrow \left(\frac{n-\sqrt{n}/2}{n}\right)^{\sqrt{n}/2} \left(\frac{n-\sqrt{n}}{n}\right)^{\sqrt{n}/2} \leq$

$p_k \leq (1)^{\sqrt{n}/2} \left(\frac{n-\sqrt{n}/2}{n}\right)^{\sqrt{n}/2} \Rightarrow e^{-1/4} e^{-1/2} \leq p_k \leq e^{-1/4} \Rightarrow 0.4723 \leq p_k \leq 0.7788$ ■

$\forall k$ можно оценить p_k гораздо точнее:

$$p_k = \frac{n}{n} \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdot \dots \cdot \frac{n-k+1}{n} = \frac{n!}{n^k (n-k)!} \approx \frac{(n/e)^n}{n^k ((n-k)/e)^{n-k}} = \frac{1}{((n-k)/n)^{n-k} e^k} = \frac{1}{(1-k/n)^{n-k} e^k} \approx \frac{1}{(e^{-1})^{(n-k)k/n} e^k}$$

Первое приближение сделано по формуле Стирлинга, второе по замечательному пределу.

• **Поллард**

Пусть у n минимальный делитель p , тогда $p \leq \sqrt{n}$. Сгенерируем $\sqrt[4]{n} \geq \sqrt{p}$ случайных чисел.
 По парадоксу дней рождений какие-то два (a и b) дадут одинаковый остаток по модулю p .
 При этом с большой вероятностью у a и b разный остаток по модулю $n \Rightarrow \gcd(a - b, n)$ – нетривиальный делитель n . Осталось придумать, как найти такую пару a, b .

```

1 x = random [2..n-1]
2 k = pow(n, 1.0 / 4)
3 for i=0..k-1: x = f(x)
    
```

Здесь функция f – псевдорандом на $[0..n)$.

Например (без док-ва) нужными нам свойствами обладает функция $f(x) = (x^2 + 1) \bmod n$.

Так мы получили x , теперь переберём y .

```

1 y = f(x)
2 for i=0..k-1:
3     g = gcd(x - y, n)
4     if (g != 1 && g != n) return g
5     y = f(y)
    
```

Почему так можно? Рассмотрим последовательность $x_i = f^i(x_0)$. Она закичивается по модулю p , что выглядит как буква « p ». Наша цель — найти две точки друг над другом. Сделав достаточно много шагов из стартовой точки окажемся на цикле. После чего остаётся пройти ещё один раз весь цикл.

Полученный нами алгоритм не работает на маленьких n . Все такие n можно проверить за $n^{1/2}$. Также стоит помнить, что вероятность его успеха $\approx \frac{1}{2}$, поэтому для вероятности ≈ 1 всю конструкцию нужно запустить несколько раз.

Сейчас асимптотика — $\mathcal{O}(n^{1/4} \cdot T(\gcd))$, чтобы получить чистое $\mathcal{O}(n^{1/4})$, воспользуемся тем, что $\gcd(a, n) = 1 \wedge \gcd(b, n) = 1 \Rightarrow \gcd(ab \bmod n, n) = 1 \Rightarrow$ можно вместо $\gcd(x-y, n)$ смотреть на группы по $\log n$ разностей: $\gcd(\prod_i (x-y_i), n)$, и только если какой-то $\gcd \neq 1$, проверять каждый y_i отдельно.

5.9. 3-SAT и random walk

• Детерминированное решение за $3^{n/2}$

Пусть решение X^* существует и в нём нулей больше чем единиц. Начнём с $X_0 = \{0, 0, \dots, 0\}$, чтобы из X_0 попасть в X^* нужно сделать не более $\frac{n}{2}$ шагов. Если X_i не решение, то какой-то клоз не выполнен, значит в X^* одна из трёх переменных этого клоза имеет другое значение. Переберём, какая. Получили рекурсивный перебор глубины $\frac{n}{2}$, с ветвлением 3. Если в X^* единиц больше, начинать нужно с $X_0 = \{1, 1, \dots, 1\}$. Нужно перебрать оба варианта.

• Рандомизированное решение за $3^{n/2}$

Упростим предыдущую идею: начнём со случайного X_0 . Матожидание расстояния от ответа равно $E = n/2$, применим неравенство Маркова: $\Pr[d(x_0, x_{ans}) \geq (n+1)/2] \leq E / ((n+1)/2) = 1 - 1/(n+1)$, то есть с вероятностью $\geq \frac{1}{n+1}$ мы на расстоянии $< (n+1)/2$ от ответа, то есть на расстоянии $\leq n/2$. Сделаем $\frac{1}{2}n$ шагов, выбирая каждый раз случайное из трёх направлений. На каждом шаге с вероятностью $\frac{1}{3}$ мы приближаемся к ответу. Итого с вероятностью $\frac{1}{n+1} \cdot \frac{1}{3^{n/2}}$ мы придём в ответ. Повторим $\mathcal{O}(3^{n/2}(n+1)) \approx \mathcal{O}(1.73^n)$ раз.

• Рандомизированное решение за 1.5^n

Почти такой же алгоритм. Сделаем теперь не $\frac{n}{2}$, а n шагов.

В процессе доказательства нам пригодится знание $\forall \alpha \sum_k \binom{n}{k} \alpha^k = (1 + \alpha)^n$.

Доказательство: раскроем скобочки.

Анализ вероятностей. Если перебор за $3^{n/2}$ перебирал все варианты, то мы перебираем 1 вариант и угадываем с вероятностью $\geq p = \sum_k \Pr[k] \frac{1}{3^k}$, где k — расстояние Хэмминга от X_0 до X^* , а

$\Pr[k] = \binom{n}{k} / 2^n$ — вероятность того, что при выборе случайного X_0 расстояние равно k .

$p = \sum_k \frac{1}{2^n} \binom{n}{k} \frac{1}{3^k} = \frac{1}{2^n} (1 + \frac{1}{3})^n = (\frac{2}{3})^n \Rightarrow$ повторим процесс 1.5^n раз.

Кстати, поскольку $\Pr[k \leq \frac{n}{2}] = \frac{1}{2}$, достаточно делать даже не n , а $\frac{n}{2}$ шагов.

• Рандомизированное решение за 1.334^n

Schoning's algorithm (1999). Получается из предыдущего решения заменой $\frac{n}{2}$ шагов на $3n$ шагов.

Вероятность успеха будет не менее $(3/4)^n \Rightarrow$ время работы $\mathcal{O}^*(1.334^n)$. **Доказательство. Статья.**

В той же статье дан более общий результат для k -SAT: $(\frac{2}{1+1/(k-1)})^n = (\frac{2(k-1)}{k})^n$.

Теорема 5.9.1. Вероятность успеха $(3/4)^n$

Доказательство. В прошлой серии мы находились на расстоянии k от ответа и хотели дви-

гаться всё время к ответу. А теперь посчитаем вероятность p того, что мы за первых $3k$ шагов сделали k шагов в неверную сторону, $2k$ шагов в верную сторону (итого пришли в ответ).

$$p = \binom{3k}{k} \left(\frac{1}{3}\right)^{2k} \left(\frac{2}{3}\right)^k$$

Приблизим $n! \approx \left(\frac{n}{e}\right)^n \Rightarrow \binom{3k}{k} \approx \frac{(3k)^{3k}}{(2k)^{2k} k^k} = \left(\frac{27}{4}\right)^k \Rightarrow p \approx \left(\frac{27}{4}\right)^k \left(\frac{1}{9}\right)^k \left(\frac{2}{3}\right)^k = \left(\frac{1}{2}\right)^k$.

Последний шаг такой же, как в 1.5^n : $Pr[\text{success}] = 2^{-n} \sum_k \binom{n}{k} 2^{-k} = 2^{-n} (1 + \frac{1}{2})^n = \left(\frac{3}{4}\right)^n$. ■

Лучшее сегодня решение 3-SAT: **алгоритм PPSZ** за 1.3067^n от Paturi, Pudlak, Saks, Zani (2005).

• Поиск хороших кафешек

Пусть Вы в незнакомом городе, хотите найти хорошее кафе. Рассмотрим следующие стратегии:

1. Осматривать окрестность в порядке удаления от начальной точки
2. Random walk без остановок
3. Random walk на фиксированную глубину с возвратом

Первый, если Вы начали в не очень богатом на кафе районе не приведёт ни к чему хорошему. У второго будут проблемы, если в процессе поиска вы случайно зайдёте в промышленный район. Третий же в среднем не лучше, зато минимизирует риски, избавляет нас от обеих проблем. Для решения 3-SAT мы использовали именно третий вариант.

5.10. Лемма Шварца-Зиппеля

Lm 5.10.1. Пусть дан многочлен P от нескольких переменных над полем \mathbb{F} .

$$(P \neq 0) \Rightarrow \Pr_x [P(x) = 0] \leq \frac{\deg P}{|\mathbb{F}|}$$

Доказательство. Индукция по числу переменных.

База: многочлен от 1 переменной имеет не более $\deg P$ корней. Переход: [wiki](#). ■

Следствие 5.10.2. Задача проверки тождественного равенства многочлена нулю $\in \text{coRP}$.

Доказательство. Подставим случайный x в поле \mathbb{F}_q , где $(q - \text{простое}) \wedge (q > 2 \deg P)$. ■

• (*) Совершенное паросочетание в произвольном графе

Дан неорграф, заданный матрицей смежности c . Нужно проверить, есть ли в нём **совершенное паросочетание**.

Def 5.10.3. Матрица Тамта $T: T_{ij} = -T_{ji}, T_{ij} = \begin{cases} 0 & c_{ij} = 0 \\ x_{ij} & c_{ij} = 1 \end{cases}$

Здесь x_{ij} – различные переменные. Пример: $c = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & x_{12} & x_{13} \\ -x_{12} & 0 & 0 \\ -x_{13} & 0 & 0 \end{bmatrix}$

Теорема 5.10.4. $\det T \neq 0 \Leftrightarrow \exists$ совершенное паросочетание

Теорему вам докажут в рамках курса дискретной математики, нам же сейчас интересно, как проверить тождество. Если бы матрица состояла из чисел, определитель можно было бы посчитать Гауссом за $\mathcal{O}(n^3)$. В нашем случае определитель – многочлен степени n . Подставим во все x_{ij} случайные числа, посчитаем определитель по модулю $p = 10^9 + 7$. Получили вероятностный алгоритм с ошибкой $\frac{n}{p}$ и временем $\mathcal{O}(n^3)$.

Следствие 5.10.5. Мы умеем искать размер максимального паросочетания за $\mathcal{O}(n^3 \log n)$

Доказательство. Сделаем бинпоиск по ответу. Чтобы проверить, есть ли в графе паросочетание размера хотя бы k рёбер ($2k$ вершин), добавим $n - 2k$ фиктивных вершин, которые соединены со всеми и проверим в полученном графе наличие совершенного паросочетания за $\mathcal{O}(n^3)$. ■

- (*) **Гамильтонов путь**

В 2010-м году в неорграфе научились проверять наличие гамильтонова пути за $\mathcal{O}^*(1.657^n)$. Можно почитать в оригинальной [статье](#) Бьёркланда (Björklund).

5.11. Random shuffle

```
1 void Shuffle(int n, int *a)
2   for (int i = 0; i < n; i++)
3     swap(a[i], a[rand() % (i + 1)]);
```

Утверждение 5.11.1. После процедуры `Shuffle` все перестановки a равновероятны.

Доказательство. Индукция. База: после фазы $i = 0$ все перестановки длины 1 равновероятны. Переход: выбрали равновероятно элемент, который стоит на i -м месте, после этого часть массива $[0..i)$ по индукции содержит случайную из $i!$ перестановок. ■

- **Применение**

Если на вход даны некоторые объекты, полезно их первым делом перемешать.

Пример: построить **бинарное дерево поиска**, содержащее данные n различных ключей. Если добавлять их в пустое дерево в данном порядке, итоговая глубина может быть n , а время построения соответственно $\Theta(n^2)$. Если перед добавлением сделать `random shuffle`, то матожидание глубины дерева $\Theta(\log n)$, матожидание времени работы $\Theta(n \log n)$. Оба факта мы докажем при более подробном обсуждении деревьев поиска.

5.12. Дерево игры [stanford.edu]

Пусть есть игра, в которую играют два игрока. Можно построить дерево игры (возможно бесконечное) – дерево всех возможных ходов. Теперь игра – спуск по дереву игры. Для некоторых простых случаев хорошо известны “быстрые” способы сказать, кто выиграет.

- **Игра на 0-1-дереве**

Задача: в листьях записаны числа 0 и 1, первый победил, если игра закончилась в 1, иначе победил второй. Пусть дерево – полное бинарное глубины n .

Решение: простой рандомизированный алгоритм «*сперва сделаем случайный ход, а второй сделаем только если не выиграли первым ходом*». Асимптотика $\mathcal{O}(1.69^n)$. Анализ в практике.

- **Игра на OR-AND-дереве**

Близкая по смыслу, но не игровая, а вычислительная задача. Дано полное бинарное дерево, в листьях 0 и 1. В промежуточных вершинах чередующиеся AND и OR гейты от детей.

Задача равносильна игре на 0-1-дереве: AND означает “чтобы первый выиграл, второй в обеих ветках должен проиграть”, OR означает “чтобы второй проиграл, первый в одной из веток

должен выиграть”. Чередование AND/OR значит “ходят по очереди”.

• Игра на min-max-дереве

Полное бинарное дерево. В листьях записаны \mathbb{Z} числа из $[0, m)$.

Первый игрок максимизирует результат игры, второй минимизирует.

Решение #1: бинарный поиск по ответу и игра на 0-1-дереве

Решение #2: (*) **альфа-бета отсечение**

Решение #3: (*) комбинация этих идей – алгоритм MTD-f ([статья](#), [wiki](#)). MTD-f в своё время стал прорывом в шахматных стратегиях.

5.13. (*) Квадратный корень по модулю

Задача: даны простое p и a , найти $x: x^2 = a \pmod p$.

Хорошо известны два решения: [алгоритм Тоннелли-Шенкса'1973](#) и [алгоритм Циполла'1907](#). Оба алгоритма вероятностны и требуют угадать квадратичный невычет.

В этой главе мы изучим только второй [\[1\]\[2\]](#), как более быстрый и простой в реализации.

Алгоритм: возьмём $P(x) = (x + i)^{(p-1)/2} - 1$ и $A(x) = x^2 - a = (x - x_1)(x - x_2)$.

Посчитаем $\gcd(P(x), A(x))$, если он содержит ровно один из $x - x_1$ и $x - x_2$, мы победили.

Теорема 5.13.1. Вероятность успеха $\geq \frac{1}{2}$. Время работы – $\mathcal{O}(\log n)$ делений чисел порядка p .

Корни существуют \Leftrightarrow символ Лежандра $a^{(p-1)/2} = 1$. В реализации вместо \gcd будем вычислять $R(x) = P(x) \pmod A(x)$ – возведение в степень с умножением похожим на комплексные числа:

$$(p_1x + q_1)(p_2x + q_2) = (p_1q_2 + p_2q_1)x + (p_1p_2a + q_1q_2)$$

Пусть $R(x) = bx + c$, тогда при $b \neq 0$ пробуем корень $-cb^{-1}$.

```

1 def root(a, p):
2     if pow_mod(a, (p-1)/2, p) != 1: return -1 # символ Лежандра
3     while True:
4         i = random [0..p)
5         bx+c = pow_mod(x+i, (p-1)/2, x^2-a) # многочлен в степени по модулю
6         if b != 0:
7             z = -c/b
8             if z^2 = a \bmod p: return z

```

Доказательство Thm 5.13.1. Для начала посмотрим на символ Лежандра и заметим, что $(\alpha/\beta - \text{вычет}) \Leftrightarrow (\text{ровно один из } \{\alpha, \beta\} - \text{вычет})$. Теперь исследуем корни многочлена из решения $P(x) = (x + i)^{(p-1)/2} - 1$. Ими являются такие z , что $(z - i)$ – квадратичный вычет. Мы хотим оценить вероятность того, что «ровно один из $\{x_1, x_2\}$ является корнем $P(x)$ » \Leftrightarrow « $(i - x_1)/(i - x_2)$ – невычет». $\forall x_1 \neq x_2 \ i \xrightarrow{f} \frac{i-x_1}{i-x_2} \Rightarrow f$ – биекция. Невычетов $(p - 1)/2$. ■

Замечание 5.13.2. Если длина p равна n , то $E(\text{времени работы})$ при “умножении по модулю” за $\mathcal{O}(n \log n)$ получается $\mathcal{O}(n^2 \log n)$.

Замечание 5.13.3. Алгоритм можно обобщить на извлечение корня k -й степени.

Лекция #6: Кратчайшие пути

25 февраля 2022

Def 6.0.1. Взвешенный граф – каждому ребру соответствует вещественный вес, обычно обозначают w_e (weight) или c_e (cost). Вес пути – сумма весов рёбер.

Задача SSSP (single-source shortest path problem):

Найти расстояния от выделенной вершины s до всех вершин.

Задача APSP (all pairs shortest path problem):

Найти расстояния между всеми парами вершин, матрицу расстояний.

Обе задачи мы будем решать в ориентированных графах.

Любое решение также будет работать и в неориентированном графе.

6.1. Short description

Приведём результаты, за сколько умеют решать SSSP для разных типов графов:

Задача	Время	Название	Год, Автор	Изучим?
ациклический граф	$V + E$	Динамика	–	+
$w_e = 1$	$V + E$	Поиск в ширину	1950, Moore	+
$w_e \geq 0$	$V^2 + E$	–	1956, Dijkstra	+
$w_e \geq 0$	$V \log V + E$	Dijkstra + fib-heap	1984, Fredman & Tarjan	+
$w_e \in \mathbb{N}$, неорграф	$V + E$	–	2000, Thorup	–
$w_e \geq 0$	$A^* \leq \text{Dijkstra}$	(A^*) A-star	1968, Nilsson & Raphael	+
любые веса	VE	–	1956, Bellman & Ford	+
$w_i \in \mathbb{Z} \cap [-N, \infty)$	$E\sqrt{V} \lceil \log(N+2) \rceil$	–	1994, Goldberg	+

Можно решать APSP запуском SSSP-решения от всех вершин. Время получится в V раз больше.

Кроме того, есть несколько алгоритмов специально для APSP:

Задача	Время	Название	Год, Автор	Изучим?
любые веса	V^3	–	1962, Floyd & Warshall	+
любые веса	$VE + V^2 \log V$	–	1977, Johnson	+
любые веса	$V^3 / 2^{\Omega(\log^{1/2} V)}$	–	2014, Williams	–

К поиску в ширину и алгоритму Флойда применима оптимизация “битовое сжатие”:

$V + E \rightarrow \frac{V^2}{w}$, $V^3 \rightarrow \frac{V^3}{w}$, где w – размер машинного слова.

Беллман-Форд, Флойд – динамическое программирование.

Дейкстра и A^* – жадные алгоритмы. A^* в худшем случае не лучше алгоритма Дейкстры, но на графах типа “дорожная сеть страны” часто работает за $o(\text{размера графа})$.

Алгоритмы Гольдберга и Джонсона основаны на не известной нам пока идее потенциалов.

6.2. bfs

Ищем расстояния от s . Расстояние до вершины v обозначим $dist_v$. $A_d = \{v : dist_v = d\}$.

Алгоритм: индукция по d .

База: $d = 0, A_0 = \{s\}$.

Переход: мы уже знаем A_0, \dots, A_d , чтобы получить A_{d+1} переберём $N(A_d)$ – соседей A_d , возьмём те из них, что не лежат в $A_0 \cup \dots \cup A_d$: $A_{d+1} = N(A_d) \setminus (A_0 \cup \dots \cup A_d)$.

Чтобы за $\mathcal{O}(1)$ проверять, лежит ли вершина v в $A_0 \cup \dots \cup A_d$, используем $mark_v$.

```

1.  A0 = {s}
2.  mark ← 0, marks = 1
3.  for d = 0..|V|
4.      for v ∈ Ad
5.          for x ∈ neighbors(v)
6.              if markx = 0 then
7.                  markx = 1, Ad+1 ← x

```

Время работы $\mathcal{O}(V + E)$, так как в строке 4 каждую v мы переберём ровно один раз для связного графа. Соответственно в строке 5 мы по разу переберём каждое ребро.

• Версия с очередью

Обычно никто отдельно не выделяет множества A_d , так как исходная задача всё-таки в том, чтобы найти $dist_v$. Обозначим $q = A_0 A_1 A_2 \dots$, т.е. выпишем подряд все вершины в том порядке, в котором мы находили до них расстояние. Запишем элементы q с нуля. Заодно заметим, что массив $mark$ не нужен, так как проверку $mark[x] = 0$ можно заменить на $dist[x] = +\infty$.

```

1.  q = {s}
2.  dist ← +∞, ds = 0
3.  for (i = 0; i < |q|; i++)
4.      v = q[i]
5.      for x ∈ neighbors(v)
6.          if distx = +∞ then
7.              distx = distv + 1, q ← x

```

Заметим, что q – очередь.

6.3. Модификации bfs

6.3.1. 1-k-bfs

Задача: веса рёбер целые от 1 до k . Найти от одной вершины до всех кратчайшие пути.

Решение раскатерением рёбер: ребро длины k разделим на k рёбер длины 1. В результате число вершин и рёбер увеличилось не более чем в k раз, время работы $\mathcal{O}(k(V + E))$.

Решение несколькими очередями

```

1 for (int d = 0; d < (V-1)k; d++) // (V-1)k = max расстояние
2     for (int x : A[d])
3         if (dist[x] == d)

```

```

4   for (Edge e : edges[x])
5       if (dist[e.end] > (D = dist[x] + e.weight))
6           A[D].push_back(e.end), dist[e.end] = D;

```

Этот же код верно работает и для весов $0-k$. Правда в таком случае мы можем вставлять новые вершины в список, по которому прямо сейчас итерируемся, поэтому `for (int x : A[d])` некорректно сработает для `vector<int>`. Его следует заменить на `for (size_t i = 0; i < A[d].size(); i++)`.

Lm 6.3.1. Такой bfs работает за $\mathcal{O}(Vk + E)$

Доказательство. Внешний цикл совершает Vk итераций, список смежности каждой вершины смотрим не более одного раза. Суммарное количество вершин во всех списках не более E . ■

Lm 6.3.2. При работе кода выше для $0-k$ -графа каждая вершина добавляется не более $k+1$ раза.

Доказательство. Пусть мы впервые добавили вершину v перейдя по ребру из u веса x . Тогда $d_v = d_u + x$, а значит v лежит в списке d_u .

В какие ещё списки её могут добавить? Только в списки от d_u до $d_u + x$. ■

Замечание 6.3.3. Если убрать из кода выше строчку 3, то он все ещё будет работать верно, разве что список смежности одной вершины может просматриваться $k+1$ раз, а не 1. Получится время работы $\mathcal{O}(k(V + E))$.

6.3.2. 0-1-bfs

Возьмём обычный bfs с одной очередью, заменим очередь на дек. Изменение кода: если перешли по ребру и ответ улучшился, добавим вершину в начало дека.

Алгоритм делает то же самое, что и $0-k$ bfs в случае $k = 1$: мы можем представлять наш дек как две очереди – вершины с расстоянием d , за ними вершины с расстоянием $d+1$. Значит, и время работы такое же: $\mathcal{O}(k(V+E)) = \mathcal{O}(V+E)$.

6.4. Дейкстра

Алгоритм Дейкстры решает SSSP в графе с неотрицательными весами. Будем от стартовой вершины s идти “вперёд”, перебирать вершины в порядке возрастания d_s (кстати, так же делает bfs). На каждом шаге берём v : $d_v = \min$ среди всех ещё не рассмотренных v . Прорелаксируем все исходящие из неё ребра, обновим d для всех соседей. Поскольку веса рёбер неотрицательны на любом пути $s = v_1, v_2, v_3, \dots$ величина $d_{v_i} \nearrow$. Алгоритм сперва знает только d_{v_1} и выберет v_1 , прорелаксирует d_{v_2} , через некоторое число шагов выберет v_2 , прорелаксирует d_{v_3} и т.д.

Формально. Общее состояние алгоритма:

Мы уже нашли расстояния до вершин из множества A , $\forall x \notin A \ d_x = \min_{v \in A} (d_v + w_{vx})$.

Начало алгоритма: $A = \emptyset$, $d_s = 0$, $\forall v \neq s \ d_v = +\infty$.

Шаг алгоритма: возьмём $v = \operatorname{argmin}_{i \notin A} d_i$, прорелаксируем все исходящие из v рёбра.

Утверждение: d_v посчитано верно. Доказательство: рассмотрим кратчайший путь в v , D – длина этого пути, пусть a – последняя вершин из A на пути, пусть b следующая за ней.

Тогда $D \geq d_a + w_{ab} \geq d_v \Rightarrow d_v$ – длина кратчайшего пути до v .

Время работы Дейкстры = $V \cdot \text{ExtractMin} + E \cdot \text{DecreaseKey}$.

- (a) Реализация без куч даст время работы $\Theta(E + V^2)$.
- (b) С бинарной кучей даст время работы $\mathcal{O}(E \log V)$. Обычно пишут именно так.
- (c) С кучей Фибоначчи даст время работы $\mathcal{O}(E + V \log V)$.
- (d) С деревом Ван-Эмбде-Бозса даст время работы $\mathcal{O}(E \log \log C)$ для целых весов из $[0, C)$.
- (e) Существуют более крутые кучи, дающие время $\mathcal{O}(E + V \log \log V)$ на целочисленных весах.

Замечание 6.4.1. Можно запускать Дейкстру и на графах с отрицательными рёбрами. Если разрешить вершину доставать из кучи несколько раз, алгоритм останется корректным, но на некоторых тестах будет работать экспоненциально долго. В среднем, кстати, те же $E \log V$.

6.5. A^* (А-звездочка)

Представьте, что едете из Петербурга в Москву. Представьте себе дорожную сеть страны. Перед поездкой ищем кратчайший маршрут. Что делает Дейкстра? Идёт из Петербурга во все стороны, перебирает города в порядке удаления от Петербурга. То есть, попытается в частности ехать в Москву через Петрозаводск (427 км) и Выборг (136 км). Логичней было бы в процессе поиска пути рассматривать только те вершины (города, населённые пункты, развязки), проезжая через которые, теоретически можно было бы быстрее попасть в Москву...

Почему так? Дейкстра – алгоритм для SSSP. Цель Дейкстры – найти расстояния до всех вершин. При поиске расстояния до конкретной вершины t в Дейкстру можно добавить естественную оптимизацию – **break** после того, как достанем t из кучи.

Алгоритм A^* можно воспринимать, как “модификацию Дейкстру, которая не пытается ехать Петербург \rightsquigarrow Москва через Выборг”. Обозначим начальную вершину s , конечную t , расстояние между вершинами a и b за d_{ab} . $\forall v$ оценим снизу d_{vt} из физического смысла задачи, как f_v .

Алгоритм A^* : в Дейкстре ключ кучи d_v заменим на $d_v + f_v$.

То есть, вершины будем перебирать в порядке возрастания не d_v , а $d_v + f_v$.

Остановим алгоритм, когда вынем из кучи вершину t .

Лм 6.5.1. Если f никогда не переоценивает расстояние ($\forall v: 0 \leq f_v \leq \text{dist}(v, t)$), то алгоритм A^* найдёт корректные расстояния. Правда при этом он может работать экспоненциально долго, так как будет доставать одну и ту же вершину из кучи много раз.

Теорема 6.5.2. Если функция f удовлетворяет неравенству треугольника: $\forall e: a \rightarrow b$ верно $f_a \leq f_b + w_{ab}$, то алгоритм A^* достанет каждую вершину из кучи не более одного раза.

Доказательство. Рассмотрим вершину a : $d_a + f_a = \min$. Возьмём $\forall v$ и путь из неё в a : $v \rightarrow u \rightarrow \dots \rightarrow c \rightarrow b \rightarrow a$. Нужно доказать, что $d_a + f_a \leq (d_v + w_{vu} + \dots + w_{cb} + w_{ba}) + f_a = F$.

Из неравенства Δ имеем $w_{ba} + f_a \geq f_b \wedge w_{cb} + f_b \geq f_a \wedge \dots \Rightarrow F \geq d_v + f_v \geq d_a + f_a$ ■

Следствие 6.5.3. Если для f верно неравенство треугольника, алгоритм можно остановить сразу после вынимания t из кучи.

Замечание 6.5.4. Заметим, что в доказательстве мы нигде не пользовались неотрицательностью весов. Для отрицательных обычно сложно найти функцию f с неравенством треугольника. Скоро мы введём технику потенциалов, узнав её, полезно ещё раз перечитать это место.

Замечание 6.5.5. “Оценка снизу” – лишь физический смысл функции f_v . Запускать алгоритм мы можем взяв любые вещественные числа.

$\forall f$, увеличив все f_v на константу, можно сделать $f_t = 0$. Будем рассматривать только такие f .

Теорема 6.5.6. Алгоритм A^* на функции f , удовлетворяющий трём условиям: (a) неравенству треугольника, (b) $f_t = 0$, (c) $f_v \geq 0$ переберёт подмножество тех вершин, которые перебрала бы Дейкстра с `break`.

Доказательство. Дейкстра переберёт $v: d_v \leq d_t$.

A^* переберёт $v: d_v + f_v \leq d_t + f_t = d_t \Rightarrow d_v \leq d_t - f_v \leq d_t$. ■

6.6. Флойд

Алгоритм Флойда – простое и красивое решение для APSP в графе с отрицательными рёбрами:

```

1 // Изначально d[i,j] = вес ребра между i и j, или ++∞, если ребра нет
2 for (int k = 0; k < n; k++)
3     for (int i = 0; i < n; i++)
4         for (int j = 0; j < n; j++)
5             relax(d[i,j], d[i,k] + d[k,j])

```

На самом деле мы считаем динамику $f[k, i, j]$ – минимальный путь из i в j , при условии, что ходить можно только по вершинами $[0, k)$, тогда

$$f[k+1, i, j] = \min(f[k, i, j], f[k, i, k] + f[k, k, j]).$$

Наша реализация использует лишь двумерный массив и чуть другой инвариант: после k шагов внешнего цикла в $d[i, j]$ учтены все пути, что и а $f[k, i, j]$ и, возможно, какие-то ещё.

Время работы $O(V^3)$. На современных машинах в секунду получается успеть $V \leq 1000$.

6.6.1. Восстановление пути

Также, как в динамике. Если вам понятны эти слова, дальше можно не читать ;-)

В алгоритмах `bfs`, `Dijkstra`, A^* при релаксации расстояния $d[v]$ достаточно сохранить ссылку на вершину, из которой мы пришли, в v . В самом конце, чтобы восстановить путь, нужно от конечной вершины пройти по обратным ссылкам.

Флойд. Способ #1. Можно после релаксации $d[i, j] = d[i, k] + d[k, j]$ сохранить ссылку $p[i, j] = k$ – промежуточную вершину между i и j . Тогда восстановление ответа – рекурсивная функция: `get(i, j) { k = p[i, j]; if (k != -1) get(i, k), get(k, j); }`

Флойд. Способ #2. А можно хранить $q[i, j]$ – первая вершина в пути $i \rightsquigarrow j$. Тогда изначально $q[i, j] = j$. После релаксации $d[i, j] = d[i, k] + d[k, j]$ нужно сохранить $q[i, j] = q[i, k]$.

6.6.2. Поиск отрицательного цикла

Что делать Флойду, если в графе есть отрицательный цикл? Хорошо бы хотя бы вернуть информацию о его наличии. Идеально было бы для каждой пары вершин (i, j) , если между ними есть кратчайший путь найти его длину, иначе вернуть `IND`.

Lm 6.6.1. $(\exists \text{ отрицательный цикл, проходящий через } i) \Leftrightarrow (\text{по окончании Флойда } d[i, i] < 0)$.

Lm 6.6.2. $(\nexists \text{ кратчайшего пути из } a \text{ в } b) \Leftrightarrow (\exists i \text{ и пути "из } a \text{ в } i" \text{ и "из } i \text{ в } b": d[i, i] < 0)$.

Замечание 6.6.3. О переполнениях. Веса всех кратчайших путей по модулю не больше $M = (V - 1)W$, где W – максимальный модуль веса ребра. Если в графе нет отрицательных циклов, Флойд ухватает типа, вмещающего числа из $[-M, M]$. При наличии отрицательных циклов, могут получиться числа меньше $-M$, поэтому будем складывать с корректировкой:

```

1 int sum(int a, int b):
2   if (a < 0 && b < 0 && a < -M - b)
3     return -M;
4   return a + b;

```

Как найти хотя бы один отрицательный цикл? Хочется попробовать восстановить путь от i до i (где $d[i, i] < 0$ также, как любой другой путь. К сожалению, именно так не получится, возможно ссылки заиклятся. Однако если поймать первый момент, когда образуется i , что $d[i, i] < 0$, то обычное восстановление ответа сработает.

6.7. (*) Кратчайшие пути в прикладных задачах

Рассмотрим задачу: у нас есть (a, b) -конь в точке $(0, 0)$ который хочет за минимальное число прыжков попасть в точку (x, y) . $a, b \leq 100$, $|x|, |y| \leq 10^9$.

Медленное решение: bfs от $(0, 0)$, пока не придём в (x, y) .

Быстрое решение.

1. Жадность. Пока $|\Delta x| + |\Delta y| > 1000$ делаем ход, который минимизирует $|\Delta x| + |\Delta y|$. 2. bfs, пока не придём в (x, y) .

Мораль. Интересности происходят только в окрестности концов пути.

• Узловые точки

Пусть мы путешествуем по большому графу (страна, система дорог). Мы уже решали задачу «как проехать из Петербурга в Москву?» Дейкстрой и A^* . Но любой автомобилист вам и так скажет «выехать на трассу М11, далее по трассе, причём тут Дейкстра?».

Мораль. Между точками в разных городах поиск кратчайшего пути может работать так: доехать до некой узловой точки (например, въезд на трассу) применить заранее сделанный предподсчёт «кратчайшие пути между всеми парами узловых точек», попасть в последнюю узловую точку (съезд с трассы), доехать до пункта назначения. Части алгоритма «доехать до некой узловой точки» и «доехать до пункта назначения» мы можем решить той же A^* .

Систему можно делать многоуровневой – узловые точки федерального масштаба, регионального масштаба, городского масштаба и т.д.

Самое сложное – выбирать узловые точки, многоуровневую схему, делать быстрый предподсчёт.

6.8. (*) Перебор и кратчайшие пути

Рассмотрим задачу: дана перестановка p длины $n \leq 20$, нужно отсортировать её за минимальное число ходов, ход – взять p_i и поставить на позицию p_i (на своё место).

Медленное решение. bfs за $n! \cdot n$ ($n!$ вершин, из каждой n рёбер).

Мы всегда можем отсортировать за n ходов – ставим последнего на своё место и т.д. Если мы не тронем какие-то элементы (оставим их на своих местах), они должны образовывать

возрастающую подпоследовательность. Итого: $n - LIS(p) \leq ans \leq n$.

• **Решение через A^***

$f[v] = n - LIS(p)$, куча по $d[v] + f[v]$. Этот подход точно не хуже чем bfs:

bfs переберёт $v: d[v] \leq d[end]$, а A^* переберёт $v: d[v] + f[v] \leq d[end] + f[end] = d[end]$.

• Решение перебором

Переборы часто пишут рекурсивно.

```

1 void go(state p, int depth) {
2     if (is_sorted(p))
3         ans = min(ans, depth);
4         return;
5     if (depth + f(v) >= ans) return; // отсечение по ответу
6     if (last_depth[v] <= depth) return; // запоминание
7     last_depth[v] = depth; // запоминание требует много памяти
8     for (move)
9         go(move(p), depth + 1)
10 }
```

В переборе самое главное – отсечение по ответу.

В принципе, можно ещё добавить запоминание, но не обязательно.

• Iterative Deepening

Чтобы сделать отсечение по ответу более мощным нужно угадать **ans**. Его угадывают линейным поиском, который называют «iterative deepening».

```

1 for (ans = 1;; ans++)
2     go(start, 0)
```

Предполагается, что `go` работает за $exp(ans) \Rightarrow$ этот `for` не ухудшил асимптотику.

• Сравнение решений

Решение #1: A^* .

Решение #2: Перебор с Iterative Deepening, но без запоминания.

Даже с идеальным отсечением по ответу (2) переберёт как минимум всё то, что и (1).

(2) точно не быстрее (1).

(2) использует $\mathcal{O}(depth)$ памяти, в (1) $time \approx space$.

Замечание 6.8.1. Функцию $f(v)$, мы, конечно, не храним (экономия памяти), и не пытаемся считать заранее (мы же не знаем, от каких вершин считать). И в A^* , и в переборе $f(v)$ считается при входе в вершину.

6.9. (*) Beam search

[\[wiki\]](#)

Решим предыдущую задачу новым способом. Возьмём самое простое решение – bfs.

Напишем его слоистым образом A_0, A_1, A_2, \dots

Идея. В каждом слое будем хранить *BEAM* лучших вершин v . Лучших, конечно, в смысле $f(v)$. Можно отсекасть чуть иначе: берём лучшую вершину в слое и храним лишь те, что не более чем на D хуже.

```

1 nth_element(A[d].begin(), A[d].begin() + BEAM, A[d].end())
2 A[d].resize(BEAM);
```

Чем *BEAM* больше, тем дольше работает, но больше вероятность найти ответ. В большинстве практических задач, нужен не оптимальный ответ, а близкий к оптимальному \Rightarrow чем больше

BEAM, тем ответ точнее.

- **Применение beam-search к calculator**

Задача из прошлого семестра: $x \rightarrow x+a, x+b, x \cdot c$, получить из 1 число $n \leq 10^{18}$ за минимальное число шагов. Тогда предлагался bfs в обратную сторону (из n в 1). Теперь мы можем ускорить этот bfs через beam-search: в слое будем хранить $BEAM \approx 60$ минимальных чисел.

- **Применение beam-search к редакционному расстоянию**

Рассмотрим задачу «найти редакционное расстояние между s и t ».

Решается слоистой динамикой $dp[i, j]$. Один переход $dp[i] \rightarrow dp[i+1]$.

Применяем beam-search: давайте в слое $dp[i]$ хранить только $BEAM$ лучших. То же можно проделать с НОП и другими слоистыми динамиками.

- **Распознавание речи**

Задача. Дан фрагмент речи a , даны образцы слов b_1, b_2, b_3, \dots

Определить, какое одно слово было произнесено?

Простейшее решение – свести задачу к редакционному расстоянию с небольшим уточнением. Каждый ход мы делаем одно из смещений $(i, j) \rightarrow (i+1, j), (i+1, j+1), (i, j+1)$ и в любом случае прибавляем к штрафу $dist(a_i, b_j)$. Можно минимизировать не сумму штрафов, а среднее арифметическое (нормировка по длине).

Заметьте, что внутри beam-search мы можем параллельно идти по всем образцам слов b_1, b_2, \dots и хранить $BEAM$ меньших по штрафу состояний. Те слова, что останутся в последнем слое динамики – и есть лучшие кандидаты.

Лекция #7: Кратчайшие пути

4–18 марта 2022

Мы уже знаем поиск ширину и алгоритм Дейкстры. Для графов с отрицательными рёбрами у нас пока есть только Флойд, который решает задачу APSP, для SSSP ничего лучше нам не известно. Цель сегодняшней лекции – научиться работать с графами с отрицательными рёбрами. Перед изучением нового попробуем модифицировать старое.

Берём Дейкстру и запускаем её на графах с отрицательными рёбрами. Когда до вершины улучшается расстояние, кладём вершину в кучу. Теперь одна вершина может попасть в кучу несколько раз, но на графах без отрицательных циклов алгоритм всё ещё корректен. В среднем по тестам полученный алгоритм работает $\mathcal{O}(VE)$ и даже быстрее, но \exists тест, на котором время работы экспоненциально. Возможность такой тест придумать будет у вас в дз.

Теперь берём bfs на взвешенном графе с произвольными весами. Вершину кладём в очередь, если до неё улучшилось расстояние. Опять же одна вершина может попасть в очередь несколько раз. На графах без отрицательных циклов алгоритм всё ещё корректен. Оказывается, что мы методом «а попробуем» получили так называем «алгоритм Форд-Беллмана с очередью», который работает за $\mathcal{O}(VE)$, а в среднем по тестам даже линейное время.

Теперь всё по порядку.

7.1. Алгоритм Форд-Беллмана

Решаем SSSP из вершины s .

Насчитаем за $\mathcal{O}(VE)$ динамику $d[k, v]$ – минимальный путь из s в v из не более чем k рёбер.

База: $d[0, v] = (v == s ? 0 : +\infty)$.

Переход: $d[k+1, v] = \min(d[k, v], \min_x d[k, x] + w[x, v])$, где внутренний минимум перебирает входящие в v рёбра, вес ребра $w[x, v]$. Получили “динамику назад”.

Ответ содержится в $d[n-1, v]$, так как кратчайший путь содержит не более $n-1$ ребра.

Запишем псевдокод версии “динамика вперёд”.

```
1 vector<vector<int>> d(n, vector<int>(n, INFINITY));
2 d[0][s] = 0;
3 for (int k = 0; k < n - 1; k++)
4     d[k+1] = d[k];
5     for (Edge e : all_edges_in_graph)
6         relax(d[k+1][e.end], d[k][e.start] + e.weight);
```

Соптимизируем память до линейной:

```
1 vector<int> d(n, INFINITY);
2 d[s] = 0;
3 for (int k = 0; k < n - 1; k++) // n-1 итерация
4     for (Edge e : all_edges_in_graph)
5         relax(d[e.end], d[e.start] + e.weight);
```

После k первых *итераций* внешнего цикла в $d[v]$ содержится минимум из некоторого множества путей, в которое входят все пути из не более чем k рёбер \Rightarrow алгоритм всё ещё корректен. Полученный псевдокод в дальнейшем мы и будем называть алгоритмом Форда-Беллмана. Его можно воспринимать так “взять массив расстояний d и улучшать, пока улучшается”.

7.2. Выделение отрицательного цикла

Изменим алгоритм Форд-Беллмана: сделаем +1 итерацию внешнего цикла.

∃ отрицательный цикл \Leftrightarrow на последней n -й итерации произошла хотя бы одна релаксация.

Действительно, если нет отрицательного цикла, релаксаций не будет. С другой стороны:

Lm 7.2.1. \forall итерации \forall отрицательного цикла произойдёт релаксация хотя бы по одному ребру.

Доказательство. Обозначим номера вершин v_1, v_2, \dots, v_k и веса рёбер w_1, w_2, \dots, w_k . Релаксаций не произошло $\Leftrightarrow (d[v_1] + w_1 \geq d[v_2]) \wedge (d[v_2] + w_2 \geq d[v_3]) \wedge \dots$. Сложим все неравенства, получим $\sum_i d[v_i] + \sum_i w_i \geq \sum_i d[v_i] \Leftrightarrow \sum_i w_i \geq 0$. Противоречие с отрицательностью цикла. ■

Осталось этот цикл восстановить. Пусть на n -й итерации произошла релаксация $a \rightarrow b$. Для восстановления путей для каждой вершины v мы поддерживаем предка $p[v]$.

• **Алгоритм восстановления:** откатываемся из вершины b по ссылкам p , пока не зациклимся. Обязательно зациклимся. Полученный цикл обязательно отрицательный.

Lm 7.2.2. \forall момент времени \forall вершины v верно $d[p_v] + w[p_v, v] \leq d[v]$.

Доказательство. В момент сразу после релаксации верно $d[p_v] + w[p_v, v] = d[v]$. До следующей релаксации $d[v]$ и p_v не меняются, а $d[p_v]$ может только уменьшаться. ■

Следствие 7.2.3. После релаксации v произошла релаксация $p_v \Rightarrow d[p_v] + w[p_v, v] < d[v]$.

Lm 7.2.4. Откат по ссылкам p из вершины b зациклится.

Доказательство. Пусть не зациклился \Rightarrow остановился в вершине s . При этом $p_s = -1 \Rightarrow d[s]$ не менялось $\Rightarrow d[s] = 0$. Обозначим вершины пути $s = v_1, v_2, \dots, v_k = b$. Последовательно для всех рёбер пути используем неравенство из леммы **Lm 7.2.2**, получаем $d[s] + (\text{вес пути}) \leq d[b] \Leftrightarrow (\text{вес пути}) \leq d[b]$. Но $d[b]$ уже обновился на n -й итерации, значит строго меньше веса любого пути из не более чем $(n-1)$ ребра. Противоречие. ■

Lm 7.2.5. Вес полученного цикла отрицательный.

Доказательство. Сложим по всем рёбрам цикла неравенство из леммы **Lm 7.2.2**, получим $\sum d_i + \sum w_i \leq \sum d_i \Leftrightarrow \sum w_i \leq 0$. Чтобы получить строгую отрицательность рассмотрим y – последнюю вершину цикла, для которой менялось расстояние. Тогда для вершины цикла z : $p_z = y$ неравенство из леммы **Lm 7.2.2** строгое. ■

Из доказанных лемм следует:

Теорема 7.2.6. Алгоритм восстановления отрицательного цикла корректен.

7.3. Модификации Форд-Беллмана

Теперь наш алгоритм умеет всё, что должен.

Осталось сделать так, чтобы он работал максимально быстро.

7.3.1. Форд-Беллман с break

Мы уверены, что после $n-1$ итерации массив d меняться перестанет. На случайных тестах это произойдёт гораздо раньше. Оптимизация: делать итерации, пока массив d меняется. Факт: на случайных графах в среднем $\mathcal{O}(\log V)$ итераций \Rightarrow время работы $\mathcal{O}(E \log V)$.

7.3.2. Форд-Беллман с очередью

Сперва заметим, что внутри итерации бесполезно просматривать некоторые рёбра. Ребро $a \rightarrow b$ может дать успешную релаксацию, только если на предыдущей итерации поменялось $d[a]$.

Пусть B_k – вершины, расстояние до которых улучшилось на k -й итерации. Псевдокод:

```

1.   $d = \{+\infty, \dots, +\infty\}$ ,  $d[s] = 0$ 
2.   $B_0 = \{s\}$ 
3.  for ( $k = 0$ ;  $B_k \neq \emptyset$ ;  $k++$ )
4.      for  $v \in B_k$ 
5.          for  $x \in \text{neighbors}(v)$ 
6.              if  $d[x] + w[v,x] < d[v]$  then
7.                   $d[v] = d[x] + w[v,x]$ ,  $B_{k+1} \cup = \{x\}$ 

```

Последние две оптимизации могли только уменьшить число операций в Форд-Беллмане.

• Добавляем очередь

Сделаем примерно то же самое, что с поиском в ширину.

Напомним, bfs мы сперва писали через множества A_d , а затем ввели очередь $q = \{A_0, A_1, \dots\}$.

```

1.   $d = \{+\infty, \dots, +\infty\}$ ,  $d[s] = 0$ 
2.   $q = \{s\}$ ,  $\text{inQueue}[s] = 1$ 
3.  while (! $q.empty()$ )
4.       $v = q.pop()$ ,  $\text{inQueue}[v] = 0$ 
5.      for  $x \in \text{neighbors}(v)$ 
6.          if  $d[x] + w[v,x] < d[v]$  then
7.               $d[v] = d[x] + w[v,x]$ 
8.              if (! $\text{inQueue}[v]$ )  $\text{inQueue}[v] = 1$ ,  $q.push(v)$ 

```

Алгоритм остался корректным. Докажем, что время работы $\mathcal{O}(VE)$. Пусть в некоторый момент t_k для всех вершин из A_k расстояние посчитано верно. В очереди каждая вершина встречается не более 1 раза \Rightarrow все вершины из очереди мы обработаем за $\mathcal{O}(E)$. После такой обработки корректно посчитаны расстояния до всех вершин из A_{k+1} .

7.3.3. (*) Форд-Беллман с random shuffle

Сейчас мы оценим, сколько в худшем и лучшем случае работает Форд-Беллман с break.

Здесь и далее одна итерация – один раз перебрать за $\Theta(E)$ все рёбра графа.

Lm 7.3.1. На любом тесте в лучшем случае сделает 1 итерацию.

Доказательство. Рассмотрим дерево кратчайших путей. Пусть в массиве рёбер, который просматривает Форд-Беллман, все рёбра упорядочены от корня к листьям. Тогда при просмотре ребра $a \rightarrow b$ расстояние до вершины a по индукции уже посчитано верно. ■

Lm 7.3.2. На любом тесте Форд-Беллман сделает итераций не больше, чем глубина дерева кратчайших путей.

Lm 7.3.3. \exists тест: Форд-Беллман в худшем случае сделает $V-1$ итерацию.

Доказательство. Возьмём граф, в котором дерево кратчайших путей – один путь длины $V-1$. Пусть в массиве рёбер, который просматривает Форд-Беллман, рёбра упорядочены от листьев к корню. Тогда после i итераций верно посчитано расстояние для ровно $i+1$ вершины. ■

• Форд-Беллманом с random_shuffle #1

```

1 vector<int> d(n, INFINITY);
2 d[s] = 0;
3 random_shuffle(all_edges_in_graph); // перемешали рёбра только 1 раз в начале
4 while (расстояния улучшаются)
5     for (Edge e : all_edges_in_graph)
6         relax(d[e.end], d[e.start] + e.weight);

```

Lm 7.3.4. На любом тесте матожидание числа итераций не больше $\frac{V}{2}$.

Доказательство. Худший тест: дерево кратчайших путей – путь e_1, e_2, \dots, e_{V-1} ; до каждой вершины $\exists!$ кратчайший путь. Пусть $e_i: v_i \rightarrow v_{i+1}$ и расстояние до v_i найдено на k_i -й фазе. $Pr[k_i \neq k_{i+1}] = \frac{1}{2}$ (зависит только от порядка рёбер e_i и e_{i-1} в перестановке). Далее пользуемся линейностью матожидания. $\mathbb{E}[\#\{i: k_i \neq k_{i+1}\}] = \sum_i Pr[k_i \neq k_{i+1}] = \frac{V-1}{2}$. ■

• Форд-Беллманом с random_shuffle #2

Забавно, но весьма похожий код имеет уже другую оценку числа фаз.

```

1 vector<int> d(n, INFINITY);
2 d[s] = 0;
3 while (расстояния улучшаются)
4     random_shuffle(all_edges_in_graph); // переупорядочиваем рёбра каждый раз
5     for (Edge e : all_edges_in_graph)
6         relax(d[e.end], d[e.start] + e.weight);

```

Lm 7.3.5. На любом тесте матожидание числа итераций не больше $\frac{V}{e-1} \approx \frac{1}{1.72}$.

Доказательство. Пусть перед началом очередной итерации мы верно знаем расстояния до вершин из множества A . Рассмотрим $v \notin A$ и путь p_v из A до v в дереве кратчайших путей. Чтобы на текущей итерации найти расстояние до v , все рёбра p_v должны быть в перестановке в том же порядке, что и в p_v . Вероятность по всем $n!$ перестановкам этого события равна $\frac{1}{|p_v|!}$. Обозначим за k число вершин, до которых мы найдём расстояние.

Пользуемся линейностью матожидания $\mathbb{E}[k] = \sum_{v \notin A} \frac{1}{|p_v|}$. Худший случай: «дерево кратчайших путей – бамбук», $\{|p_v|\} = \{1, 2, \dots\} \Rightarrow \mathbb{E}[k] = \sum_{i \geq 1} \frac{1}{i!} = e - 1$. ■

7.4. Потенциалы Джонсона

Чтобы воспользоваться алгоритмом Дейкстры на графах с отрицательными рёбрами, просто сделаем веса неотрицательными...

Def 7.4.1. Любой массив вещественных чисел p_v можно назвать потенциалами вершин. При этом потенциалы задают новые веса рёбер: $e: a \rightarrow b, w'_e = w_e + p_a - p_b$.

Самое важное: любые потенциалы сохраняют кратчайшие пути.

Lm 7.4.2. $\forall s, t$ кратчайший путь на весах w_e перейдёт в кратчайший на весах w'_e

Доказательство. Рассмотрим любой путь $s \rightsquigarrow t: s = v_1 \rightarrow v_2 \rightarrow \dots \rightarrow v_k = t$.

Его новый вес равен $w'_{v_1 v_2} + w'_{v_2 v_3} + \dots = (w_{v_1 v_2} + p_{v_1} - p_{v_2}) + (w_{v_2 v_3} + p_{v_2} - p_{v_3}) + \dots$

Заметим, что почти все p_v сокращаются, останется $W + p_s - p_{p_t}$, где W – старый вес пути.

То есть, веса всех путей изменились на константу \Rightarrow минимум перешёл в минимум. ■

Осталось дело за малым – подобрать такие p_v , что $\forall e w'_e \geq 0$.

Для этого внимательно посмотрим на неравенство $w'_e = w_e + p_a - p_b \geq 0 \Leftrightarrow p_b \leq p_a + w_e$ и поймём, что расстояния d_v , посчитанные от любой вершины s отлично подходят на роль p_v .

Чтобы все v были достижимы из s , введём фиктивную s и из неё во все вершины нулевые рёбра.

Если в исходном графе $\forall e w_e \geq 0$, получим $\forall v d_v = 0$, в любом случае $\forall v d_v \leq 0$.

Как найти расстояния? Форд-Беллманом.

Можно не добавлять s , а просто начать работу Форд-Беллмана с массива $\mathbf{d} = \{0, 0, \dots, 0\}$.

Получается мы свели задачу “поиска расстояний” к “задаче поиска потенциалов”, а её обратно к “задаче поиска расстояний”. Зачем?

Алгоритм Джонсона решает задачу APSP следующим образом: один раз запускает Форда-Беллмана для поиска потенциалов, затем V раз запускает Дейкстру от каждой вершины, получает матрицу расстояний в графе без отрицательных циклов за время $VE + V(E + V \log V) = VE + V^2 \log V$. Заметим, что Форд-Беллман не является узким местом этого алгоритма.

Лекция #8: DSU, MST и Ёен

11–18 марта 2022

8.1. DSU: Система Непересекающихся Множеств

Цель: придумать структуру данных, поддерживающую операции:

- `init(n)` – всего n элементов, каждый в своём множестве.
- `get(a)` – по элементу узнать уникальный идентификатор множества, в котором лежит a .
- `join(a, b)` – объединить множества элемента a и элемента b .

Для удобства считаем, что элементы занумерованы числами $0 \dots n-1$.

8.1.1. Решения списками

Напишем максимально простую реализацию, получится примерно следующее:

```

1 list<int> sets[n]; // храним множества в явном виде
2 int id[n]; // для каждого элемента храним номер множества
3 void init(int n) {
4     for (int i = 0; i < n; i++)
5         id[i] = i, sets[i] = {i};
6 }
7 int get(int i) {
8     return id[i]; // O(1)
9 }
10 void join(int a, int b) {
11     for (int x : sets[id[b]]) // долгая часть
12         id[x] = id[a];
13     sets[id[a]] += sets[id[b]]; // O(1), += не определён... но было бы удобно =>
14 }

```

Чтобы долгая часть работала быстрее будем «перекрашивать» меньшее множество (у нас же есть выбор!). Для этого перед `for` добавим `if + swap(a, b)`.

Теорема 8.1.1. Тогда суммарное число операций всех `for` во всех `join` не более $n \log n$.

Доказательство. Пусть x сейчас перекрашивается \Rightarrow размер множества, в котором живёт x , увеличился как минимум вдвое. Для каждого x произойдёт не более $\log_2 n$ таких событий. ■

Следствие 8.1.2. Суммарное время работы m `join`-ов $O(m + n \log n)$.

8.1.2. Решения деревьями

Каждое множество – дерево. Отец корня дерева – он сам. Функция `get` вернёт корень дерева. Функция `join` соединит корни деревьев. Для ускорения используют две эвристики:

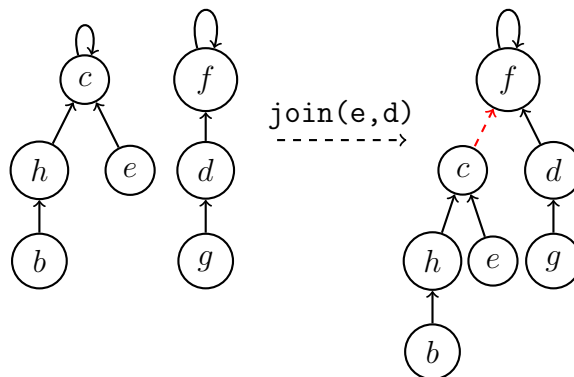
- **Эвристика сжатия путей:** все рёбра, по которым пройдёт `get` перенаправить в корень.
- **Ранговая эвристика:** подвешивать меньшее к большему. Можно выбирать меньшее по глубине, рангу, размеру. Эти идеи дают идентичный результат.

Def 8.1.3. Ранг вершины – глубина её поддерева, если бы не было сжатия путей. Сжатие путей не меняет ранги. Ранг листа равен нулю.

Вот базовая реализация без ранговой эвристики и без сжатия путей:

```

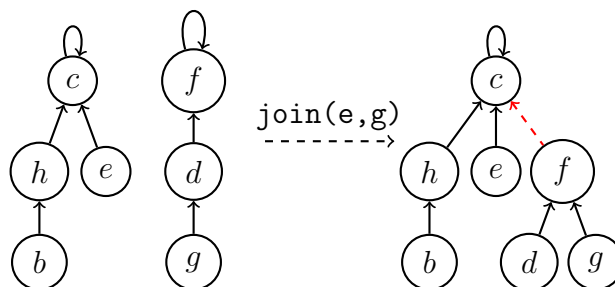
1 int p[n]; // для каждого элемента храним отца
2 void init(int n) {
3     for (int i = 0; i < n; i++)
4         p[i] = i; // каждый элемент - сам себе корень
5 }
6 int get(int i) {
7     // подняться до корня, вернуть корень
8     return p[i] == i ? i : get(p[i]);
9 }
10 void join(int a, int b) {
11     a = get(a), b = get(b); // перешли к корням
12     p[a] = b; // подвесили один корень за другой
13 }
    
```



А вот версия и с ранговой эвристикой и со сжатием путей:

```

1 int p[n], rank[n];
2 void init(int n) {
3     for (int i = 0; i < n; i++)
4         p[i] = i, rank[i] = 0;
5 }
6 int get(int i) {
7     return p[i] == i ? i : (p[i] = get(p[i]))
8     ;
9 }
10 void join(int a, int b) {
11     a = get(a), b = get(b);
12     if (not (rank[a] <= rank[b])) swap(a, b);
13     if (rank[a] == rank[b]) rank[b]++;
14     p[a] = b; // a - вершина с меньшим рангом
    
```



Время работы join во всех версиях равно Time(get) + 1. Теперь оценим время работы get.

Теорема 8.1.4. Ранговая эвристика без сжатия путей даст $\mathcal{O}(\log n)$ на запрос.

Для доказательства заметим, что глубина не больше ранга, а ранг не больше $\log_2 n$, так как:

Lm 8.1.5. Размер дерева ранга k хотя бы 2^k .

Доказательство. Индукция. База: для нулевого ранга имеем ровно одну вершину.

Переход: чтобы получить дерево ранга $k + 1$ нужно два дерева ранга k . ■

Напомним пару определений:

Def 8.1.6. Пусть v – вершина, p – её отец, $size$ – размер поддерева.

Ребро $p \rightarrow v$ называется тяжёлым если $size(v) > \frac{1}{2}size(p)$
лёгким если $size(v) \leq \frac{1}{2}size(p)$

Теорема 8.1.7. Сжатие путей без ранговой эвристики даст следующую оценку: m запросов get в сумме пройдут не более $m + (m + n) \log_2 n$ рёбер.

Доказательство. $\forall v$, поднимаясь вверх от v , мы пройдём не более $\log_2 n$ лёгких рёбер, так как при подъёме по лёгкому ребру размер поддерева хотя бы удваивается. Рассмотрим любое тяжёлое ребро ($v \rightarrow p$) кроме самого верхнего. Сжатие путей отрезает v у p , это уменьшает размер поддерева p хотя бы в два раза, что для каждого p случится не более $\log_2 n$ раз. ■

8.1.3. Оценка $\mathcal{O}(\log^* n)$

Теорема 8.1.8. *Сжатие путей вместе с ранговой эвристикой дадут следующую оценку: m запросов `get` в сумме пройдут не более $m(1 + 2 \log^* n) + \Theta(n)$ рёбер.*

Доказательство. Обозначим $x = 1.7$, заметим $x^x \geq 2$. Назовём ребро $v \rightarrow p$ крутым, если $\text{rank}[p] \geq x^{\text{rank}[v]}$ (при подъёме по этому ребру ранг сильно возрастает). Обозначим время работы i -го `get`, как $t_i = 1 + a_i + b_i =$ одно верхнее ребро $+ a_i$ крутых $+ b_i$ не крутых.

$\forall v$ имеем $\text{rank}[p_v] > \text{rank}[v] \Rightarrow a_i \leq 2 \log^* n$. Теперь исследуем жизненный цикл не крутых рёбер. $\forall v$, если v не корень, $\text{rank}[v]$ уже не будет меняться. При сжатии путей у всех рёбер кроме самого верхнего возрастёт разность $(\text{rank}[p_v] - \text{rank}[v]) \Rightarrow$ каждое некрутое ребро уже через $x^{\text{rank}[v]}$ проходов по нему вверх навсегда станет крутым $\Rightarrow \sum_i b_i \leq \sum_v x^{\text{rank}[v]} = \sum_r \text{count}(r)x^r$, где $\text{count}(r)$ – число вершин с рангом r . Вершины с рангом r имеют поддеревья размера $\geq 2^r$, эти поддеревья не пересекаются $\Rightarrow \text{count}(r) \leq \frac{n}{2^r} \Rightarrow \sum_i b_i \leq \sum_r \frac{n}{2^r} x^r = n \sum_r \left(\frac{x}{2}\right)^r = \Theta(n)$. ■

Теорема 8.1.9. *Сжатие путей вместе с ранговой эвристики дадут следующую оценку: m запросов `get` в сумме пройдут не более $\Theta(m + n \log^* n)$ рёбер.*

Доказательство. Крутое ребро, которое k раз поучаствовало в сжатии путей в качестве не самого верхнего крутого, назовём *ребром крутизны k* . Крутизна любого ребра не более $\log^* n$. При сжатии пути, крутизны всех крутых кроме самого верхнего растут $\Rightarrow \sum a_i \leq n \log^* n$. ■

8.2. (*) Оценка $\mathcal{O}(\alpha^{-1}(n))$

8.2.1. (*) Интуиция и $\log^{**} n$

Теорема 8.2.1. *Сжатие путей вместе с ранговой эвристикой дадут следующую оценку: m запросов `get` в сумме пройдут не более $\Theta(m(1 + \log^{**} n) + n)$ рёбер.*

Доказательство. Ребро ($v \rightarrow p$) крутизны хотя бы $\text{rank}[v]$ назовём *дважды крутым*.

Если обычное ребро $x^{\text{rank}[v]}$ раз поучаствует в сжатии, оно станет крутым \Rightarrow проходов по не крутым рёбрам $\sum_v x^{\text{rank}[v]} = \Theta(n)$.

Если крутое ребро $\text{rank}[v]$ раз поучаствует в сжатии, оно станет дважды крутым \Rightarrow проходов по не дважды крутым рёбрам $\sum_v \text{rank}[v] \leq 2n + \sum_v x^{\text{rank}[v]} = \Theta(n)$.

Осталось оценить число дважды крутых рёбер на пути. Для дважды крутого ребра имеем:

$$\text{rank}[p_v] \geq x^{x^{x^{\dots^{\text{rank}[v]}}}}$$

Высота степенной башни – $\text{rank}[v]$, поэтому $(\text{rank}[v] > 2 \log^* n \Rightarrow \text{rank}[p_v] > n) \Rightarrow$ на любом пути вниз $\mathcal{O}(\log^{**} n)$ дважды крутых рёбер (проход по такому ребру меняет ранг r на $2 \log^* r$). ■

Замечание 8.2.2. Аналогично можно рассмотреть *трижды крутые* и даже k -раз крутые рёбра. Тогда суммарное время всех `get`-ов будет $\Theta((m + n)k + m \log^{** \dots} n)$.

8.2.2. (*) Введение обратной функции Аккермана

Def 8.2.3. *Функция Аккермана*

$$\begin{cases} A_0(n) = n + 1 \\ A_k(n) = A_{k-1}^{n+1}(n) = A_{k-1}(A_{k-1}(\dots(n))) \text{ (взять функцию } n + 1 \text{ раз)} \end{cases}$$

Lm 8.2.4. $A_k(n) > n$

Lm 8.2.5. $f(t) = A_t(1)$ монотонно возрастает

Доказательство. $A_{t+1}(1) = A_t(A_t(1)) = A_t(x) > A_t(1)$, так как $x > 1$. ■

Выпишем несколько первых функций явно:

$$A_0(t) = t + 1$$

$$A_1(t) = A_0^{(t+1)}(t) = 2t + 1$$

$$A_2(t) = A_1^{(t+1)}(t) = 2^{t+1}(t + 1) - 1 \geq 2^t \text{ (последнее равенство доказывает по индукции)}$$

$$A_3(t) = A_2^{(t+1)}(t) \geq A_2^t(2^t) \geq \dots \geq 2^{2^{\dots^t}} \text{ (высота башни } t + 1).$$

Def 8.2.6. *Обратная функция Аккермана* $\alpha(t) = \min\{k \mid A_k(1) \geq t\}$

Посчитаем несколько первых значений функции α

$$A_0(1) = 1 + 1 = 2$$

$$A_1(1) = 2 \cdot 1 + 1 = 3$$

$$A_2(1) = 2^2 * 2 - 1 = 7$$

$$A_3(1) = A_2(A_2(1)) = A_2(7) = 2^8 \cdot 8 - 1 = 2047$$

$$A_4(1) = A_3(A_3(1)) = A_3(2047) > 2^{2^{\dots^{2047}}} \text{ (башня высоты 2048).}$$

Получившаяся оценка снизу на $A_4(1)$ – невероятно большое число, превышающее число атомов в наблюдаемой части Вселенной. Поэтому обычно предполагают, что $\alpha(t) \leq 4$.

8.2.3. (*) Доказательство

Для краткости записей ранг вершины e СНМ будем обозначать r_e , а родителя e в СНМ p_e .

У каждого множества в СНМ есть ровно один корень (представитель множества).

Элемент, который не является корнем, будем называть обычным.

Мы уже знаем, что $\forall e: e \neq p_e \ r_{p_e} > r_e$. Интересно, на сколько именно больше:

Def 8.2.7. *Порядок элемента.* \forall обычного элемента e $level(e) = \max\{k \mid r_{p_e} \geq A_k(r_e)\}$.

Def 8.2.8. *Итерация порядка.* \forall обычного элемента e $iter(e) = \max\{i \mid r_{p_e} \geq A_{level(e)}^{(i)}(r_e)\}$.

Получили пару $\langle level(e), iter(e) \rangle$. Чем больше пара, тем у e больше разница с рангом отца.

Lm 8.2.9. \forall обычного элемента e верно, что $0 \leq level(e) < \alpha(n)$.

Доказательство. $level(e) \geq 0$, так как $A_0(r_e) = r_e + 1 \leq r_{p_e}$.

$n > r_{p_e} \geq A_{level(e)}(r_e) \geq A_{level(e)}(1) \Rightarrow n > A_{level(e)}(1) \Rightarrow level(e) < \alpha(n)$. ■

Lm 8.2.10. \forall обычного элемента e верно, что $1 \leq \text{iter}(e) \leq r_e$.

Доказательство. $r_{p_e} \geq A_{\text{level}(e)}(r_e) \Rightarrow \text{iter}(e) \geq 1$.

$r_{p_e} < A_{\text{level}(e)+1}(r_e) = A_{\text{level}(e)}^{(r_e+1)}(r_e) \Rightarrow \text{iter}(e) < r_e + 1 \Leftrightarrow \text{iter}(e) \leq r_e$. ■

Будем доказывать время работы СНМ методом амортизационного анализа.

Def 8.2.11. *Потенциал элемента.* Определим для любого элемента e величину $\varphi(e)$.

$$\varphi(e) = \begin{cases} \alpha(n) \cdot r_e & \text{если } e \text{ – представитель своего множества} \\ (\alpha(n) - \text{level}(e)) \cdot r_e - \text{iter}(e) & \text{если } e \text{ – обычный элемент} \end{cases}$$

Def 8.2.12. *Потенциал.* Возьмём $\varphi = \sum_e \varphi(e)$.

Обозначим через φ_t потенциал после первых t операций.

Lm 8.2.13. $\varphi_0 = 0, \forall i \varphi_i \geq 0$.

Доказательство. Изначально все элементы – представители, и все $r_e = 0$. Далее все $r_e \geq 0$, появляются обычные элементы. Используем [Lm 8.2.9](#) и [Lm 8.2.10](#), получаем неотрицательность потенциалов обычных элементов. ■

Lm 8.2.14. Если у обычной вершины v увеличить ранг отца, то $\varphi(v) \searrow$.

Теорема 8.2.15. Для операции типа *join* амортизированное время $a_i = t_i + \Delta\varphi = \mathcal{O}(\alpha(n))$.

Доказательство. Пусть мы присвоили $p_b = a$. Тогда потенциал мог измениться только у b , a и детей a . Итого $\Delta\varphi = (\sum_{c \in \text{children}(a)} \Delta\varphi(c)) + \Delta\varphi(b) + \Delta\varphi(a)$. Первые два слагаемых неположительны, последнее равно $\alpha(n)$. ■

Теорема 8.2.16. Для операции типа *get* амортизированное время $a_i = t_i + \Delta\varphi = \mathcal{O}(\alpha(n))$.

Доказательство. После сжатия путей по [Lm 8.2.14](#) потенциалы вершин не могли увеличиться. Пусть i -ая операция *get* проходит $w_i + x_i + 2$ вершин. Здесь w_i – число вершин $v: \Delta\varphi(v) < 0$, x_i – число, вершин $v: \Delta\varphi(v) = 0$, но $r_{p_v} \uparrow$, а 2 – корень и его сын. Если у вершины v не меняется потенциал, не меняются и $\text{level}(v), \text{iter}(v)$.

Осталось показать, что $x_i \leq \alpha(n)$. От противного: $(x_i > \alpha(n)) \wedge$ (по [Lm 8.2.9](#) \exists не более $\alpha(n)$ различных значений level) $\Rightarrow \exists$ две вершины a и b на пути: $\Delta\varphi(a) = \Delta\varphi(b) = 0, \text{level}(a) = \text{level}(b)$. После сжатия путей у нижней из $\{a, b\}$ как минимум изменится iter . Противоречие. ■

Теорема 8.2.17. Суммарное время работы m произвольных операций с СНМ – $\mathcal{O}(m \cdot \alpha(n))$.

Доказательство. $\sum_i t_i = \sum_i a_i - \varphi_m + \varphi_0$ ($\varphi_0 = 0, \varphi_m \geq 0$) $\Rightarrow \sum_i t_i \leq \sum_i a_i = m \cdot \alpha(n)$. ■

8.3. MST: Минимальное Остовное Дерево

Def 8.3.1. *Остовное дерево связного графа – подмножество его рёбер, являющееся деревом.*

Def 8.3.2. *Вес остова – сумма весов рёбер.*

Задача: дан граф, найти остов минимальной стоимости.

Задача поиска максимального остова равносильна данной (домножение весов на -1).

8.3.1. Алгоритм Краскала

Начнём с пустого остова. Будем перебирать рёбра в порядке возрастания веса.

Если текущее ребро при добавлении в остов не образует циклов, добавим ребро в остов.

Проверку на циклы будем делать СНМ-ом. Время работы алгоритма: $\mathcal{O}(\text{sort}(E) + (V+E)\alpha)$.

8.3.2. Алгоритм Прима

Поддерживаем остовное дерево множества вершин A . База: $A = \{0\}$, рёбер в остове нет. Для перехода $\forall v$ поддерживаем $d_v = \min_{A \rightarrow v} \langle w_e, e \rangle$. Переход: выбрать $v \notin A$ $d_v = \min$, подсоединить v к остову за ребро d_v . **second**. Чтобы быстро выбирать вершину v , поддерживаем кучу всех d_v .

Время работы = $E \cdot \text{DecreaseKey} + V \cdot \text{ExtractMin}$, то есть, $\mathcal{O}(E + V \log V)$ и $\mathcal{O}(V^2)$.

Замечание 8.3.3. Алгоритм Прима – Дейкстра с операцией «max» вместо «+».

8.3.3. Алгоритм Борувки

Вес ребра w_e поменяем на пару $\langle w_e, e \rangle$. Теперь все веса различны.

Фаза алгоритма: для каждой вершины выберем минимальное по весу ребро.

Выбранные рёбра не образуют циклов (от противного: пользуемся тем, что веса различны).

Добавим все выбранные рёбра в остов. Сожмём компоненты связности. Удалим кратные рёбра.

Алгоритм: пока граф не сожмётся в одну вершину выполняем очередную фазу алгоритма.

Удалять кратные рёбра будем сортировкой рёбер.

Рёбра – пары чисел от 1 до $V \Rightarrow$ цифровая сортировка справится за $\mathcal{O}(V + E)$.

Lm 8.3.4. *Время работы алгоритма Борувки $\mathcal{O}((V + E) \log V)$.*

Доказательство. Фаза работает за $V + E$. Число вершин уменьшится в два раза. ■

Lm 8.3.5. *Время работы алгоритма Борувки $\mathcal{O}(E + V^2)$.*

Доказательство. $\mathcal{O}(V + E)$ нужно на первое удаление кратных рёбер.

Теперь всегда $E \leq V^2 \Rightarrow$ алгоритм работает за $\mathcal{O}(V^2 + (\frac{V}{2})^2 + (\frac{V}{4})^2 + \dots) = \mathcal{O}(V^2)$. ■

Lm 8.3.6. *Время работы алгоритма Борувки $\mathcal{O}(E \cdot \lceil \log V^2/E \rceil)$.*

Доказательство. V^2 за фазу уменьшается в 4 раза \Rightarrow за $\lceil \log_4 V^2/E \rceil$ фаз V^2 станет не больше исходного E . По предыдущей лемме оставшаяся часть алгоритма отработает за $\mathcal{O}(E)$. ■

8.3.4. Сравнение алгоритмов

Асимптотика: из изученных нами куч и сортировок получаем, что лучше всего

Прим за $\mathcal{O}(E + V \log V)$, Прим за $\mathcal{O}(E \log_{E/V} V)$ и Краскал за $\mathcal{O}(E \log \log C)$.

- Краскал просто пишется, быстро работает.
- На очень плотных графах быстрее V^2 через Прима.
- Алгоритм Борувки мы оценили лишь сверху. В среднем по тестам он даёт линию.

Замечание 8.3.7. На всех планарных графах Борувка работает за $\mathcal{O}(V + E)$.

8.3.5. Лемма о разрезе и доказательства

Доказательства трёх алгоритмов очень похожи, все они опираются на *лемму о разрезе*:

Lm 8.3.8. Для любого разбиения множества вершин $V = A \sqcup B$, существует минимальный остов, содержащий e – минимальное по весу ребро, проходящее через разрез $\langle A, B \rangle$.

Доказательство. Возьмём минимальный остов без e . Добавим e , получится цикл.

Два ребра этого цикла проходят через разрез $\langle A, B \rangle$. Старое не меньше e , удалим его. ■

Если же в процессе построения \min остова уже известно подмножество рёбер будущего \min остова, эти рёбра задают компоненты связности. Мы можем сжать компоненты в вершины и для конденсации применить лемму о разрезе. Итого:

Следствие 8.3.9. X – подмножество рёбер некоего минимального остова. Зафиксируем разрез $V = A \sqcup B$ такой, что все рёбра X не пересекают разрез. Тогда \exists минимальный остов, содержащий $X \cup \{e\}$, где e – минимальное по весу ребро, проходящее через разрез $\langle A, B \rangle$.

Для доказательства описанных выше алгоритмов нужно лишь применить правильный разрез.

• Доказательство алгоритма Прима.

Разрез – текущее множество A и дополнение $V \setminus A$.

• Доказательство алгоритма Краскала.

Добавляем ребро (a, b) . Разрез – любой такой, что (a, b) через него проходит.

• Доказательство алгоритма Борувки.

$\forall v$ ребро минимальное для вершины v можно добавить. Разрез $\langle \{v\}, V \setminus \{v\} \rangle$.

Почему все эти рёбра можно добавить одновременно? Верно обобщение леммы о разрезе.

Lm 8.3.10. Пусть $\forall e \in D$ ребро e можно добавить в остов X по лемме о разрезе, и множество $X \cup D$ не содержит циклов, тогда \exists минимальный остов, содержащий $X \cup D$.

Для нашего конкретного случая подойдут следующие разрезы.

Рёбра, которые хочет добавить Борувка, разобьём на компоненты связности. Каждая компонента – дерево, рёбра компоненты будем добавлять по одному от корня к листьям. Добавляя очередное ребро ведущее в лист v , используем разрез $\langle \{v\}, V \setminus \{v\} \rangle$ и лемму о разрезе. ■

8.4. (*) Алгоритм Йена

Мы научились искать кратчайший простой путь, MST.

Идея Йена помогает искать ещё и k -й минимальный (простой путь, остовное дерево, поток...)

Оригинальный алгоритм ищет именно k -й путь.

Алгоритм. Запустим Дейкстру, получим кратчайший путь p_1 .

Второй простой путь p_2 отличается от p_1 , переберём, совпадающий префикс путей, запретим ходить по следующему ребру, запустим Дейкстру. Итого p_2 мы нашли за $|p_1|$ запусков Дейкстры.

Чтобы найти p_3 , заметим, что при поиске p_2 у нас было несколько кандидатов вида $\langle P, E, W \rangle$ – множество путей, у которых начало P , после P запрещено ходить по рёбрам из E (пока $|E| = 1$, в дальнейшем будет больше), Дейкстра из множества $\langle P, E \rangle$ нашла кратчайший, его вес – W .

p_2 – минимум по W_i , а также минимум в своём множестве путей $\langle P_i, E_i \rangle$. Оствшиеся в $\langle P_i, E_i \rangle$ пути разделятся на множества такого же вида – пройдем P_i , зафиксируем ещё несколько (возможно, 0) вершин по p_2 и свернём (запретим следующее ребро из p_2).

Нужно хранить кучу k минимальных по W_i множеств $\langle P_i, E_i \rangle$ и k раз “выбирать следующий минимальный путь”: вытаскивать из кучи минимум, затем $\leq V$ раз запускать Дейкстру.

Итого: 2-й путь нашли за $\mathcal{O}(V \cdot \text{Dijkstra})$, а k -й за $\mathcal{O}(kV \cdot \text{Dijkstra})$.

8.5. (*) Алгоритм Эпштейна для k -го пути

[Eppstein'98]

Задача: дан орграф, на котором мы умеем быстро строить дерево кратчайших путей (например, $w_e \geq 0$ или ациклический), найти среди всех (необязательно простых) путей из s в t (концы фиксированы) k -ую статистику по суммарному весу рёбер.

Возможное применение: возьмём стандартную линейную динамику «дойти из 1 в n , переходя каждый раз вперёд или на 2, или на 3, собрать максимальную сумму». Попросим найти из всего множества способов дойти из 1 в n сразу k максимумов. Решение: строим ациклический граф динамики, на нём Эпштейна.

Решение за $\mathcal{O}((m+n+k) \log(mk))$.

1. Построим дерево кратчайших путей из t по обратным рёбрам. $d[v]$ – расстояние до t , p_v – отец в дереве кратчайших путей.

2. $\forall v$ будем поддерживать кучу $h[v]$ – возможные ответвления от кратчайшего пути $v \rightsquigarrow t$.

$h[v] = h[p_v] \cup \{(d[x] + w_e) - d[v] \mid e: v \rightarrow x\}$, в $h[p_v]$ уже лежат ответвления сделанные дальше по пути, $(d[x] + w_e) - d[v]$ это то, на сколько вес ответвления больше веса кратчайшего пути.

3. Изначально множество кандидатов $H = h[s]$, далее k раз достаём очередной минимум и добавляем новых кандидатов.

```

1 d[], p[] = Dijkstra(t)
2
3 for v: // динамика, перебираем вершины от корня дерева
4     h[v] = h[p[v]].copy() // используем персистентность
5     for e in g[v]:
6         h[v].add({d[b[e]]+w[e] - d[v], e}) // запомним Δ веса и номер ребра
7 H = h[s]
8
9 for k:
10     <Delta,e> = H.extractMin() // e : a[e] ->b[e]
11     H = H.merge(h[b[e]].c(Delta)) // увеличить все элементы на Delta

```

В качестве $h[v]$ нам нужна структура данных, которая быстро умеет

`add, extractMin, merge, copy, incAll`

Подойдёт leftist heap (левацкая куча) и хранение числа-увеличителя.

Лекция #9: Жадность и приближённые алгоритмы

6 апреля 2022

The point is, ladies and gentleman, greed is good. Greed works, greed is right. Greed clarifies, cuts through, and captures the essence of the evolutionary spirit. Greed in all its forms, greed for life, money, love, knowledge has marked the upward surge in mankind. And greed-mark my words-will save not only Teldar Paper but the other malfunctioning corporation called the USA.

Gordon Gekko [Michael Douglas], *Wall Street (1987)*

There is always an easy solution to every human problem-neat, plausible, and wrong.

H. L. Mencken, *“The Divine Afflatus”, New York Evening Mail (November 16, 1917)*

9.1. Хаффман

Для начала заметим печальный факт: идеального архиватора не существует.

Lm 9.1.1. \nexists архиватора f и целого $n: \forall x (|x| = n \Rightarrow f(x) < n)$

Доказательство. Пусть \exists . Входов 2^n , выходов $1 + 2 + \dots + 2^{n-1} < 2^n \Rightarrow f$ – не инъекция. ■

Принцип сжатия текста: сопоставить каждому символу битовый код.

Кодирование: записать сами коды и выписать общую последовательность бит.

Раскодирование: жадно откусить от кода 1 символ.

Когда раскодирование однозначно? iff коды *безпрефиксные*.

Алгоритм Хаффмана – наиболее известный алгоритм сжатия текста.

Хотим придумать префиксные коды, минимизирующие величину

$$F = \sum_i (len_i \cdot cnt_i)$$

где len_i – длина кода, cnt_i – частота i -го символа, F – количество бит в закодированном тексте.

На практике нужно ещё сохранить сами коды, и длина закодированного текста будет

$F + |store_codes|$. Префиксные коды удобно представлять в виде дерева. Чтобы раскодировать очередной символ закодированного текста, спустимся по дереву кодов до листа.

• Алгоритм построения префиксных кодов

Будем строить дерево снизу. Изначально каждому символу с ненулевой частотой сопоставлен лист дерева. Пока в алфавите больше одного символа, выберем символы x и y с минимальными cnt , создадим новый символ xy и вершину дерева для него, из которой ведут рёбра в x и y . Сделаем $cnt_{xy} = cnt_x + cnt_y$, удалим из алфавита x , y , добавим xy .

• Реализация

Если для извлечения минимума использовать кучу, мы получим время $O(\Sigma \log \Sigma)$, где Σ – размер алфавита.

Можно изначально отсортировать символы по частоте и заметить, что новые символы по частоте только возрастают, поэтому для извлечения минимума достаточно поддерживать две очереди – “исходные символы” и “новые символы”. Минимум всегда содержится в начале одной из двух очередей $\Rightarrow O(sort(\Sigma) + \Sigma)$.

Теорема 9.1.2. Алгоритм корректен.

Доказательство. Если упорядочить символы $len_i \searrow$, то $cnt_i \nearrow$ (иначе можно поменять местами два кода и уменьшить F). У каждой вершины дерева ровно два ребёнка (если ребёнок только один, код можно было бы укоротить). Значит, существует два брата, две самых глубоких вершины, и это ровно те символы, у которых cnt минимален. Осталось сделать переход по индукции к задаче того же вида: заменим два самых глубоких символа-брата на их отца, и будем искать код отца. Длина кода отца домножается как раз на сумму частот братьев. ■

9.1.1. Хранение кодов

Есть несколько способов сохранить коды.

1. Сохранить не коды, а частоты. Декодер строит по ним коды так же, как кодер.
2. Рекурсивный код дерева: если лист, пишем бит 0 и символ, иначе пишем бит 1.
Длина = $2 + \lceil \log \Sigma \rceil$ бит на каждый ненулевой символ.
3. Сделать dump памяти = $\text{sizeof(Node)} * (2\Sigma - 1)$ байт.

Какой бы способ мы не выбрали, можно попробовать рекурсивно сжать коды Хаффманом. Чтобы рекурсия не была бесконечной добавим в код один бит – правда ли, что сжатый текст меньше. Если этот бит ноль, вместо кода запишем исходный текст.

9.2. Компаратор и сортировки

Решим несколько простых задач.

1. Сортировка по времени

Даны n заданий, у каждого есть время выполнения t_i . К моменту T выполнить как можно больше заданий. Решение: выполняем в порядке возрастания t_i . Доказательство: не важен порядок \Rightarrow выполняем в порядке возрастания t_i . Пусть на первом месте не t_{min} , поставим $t_{min} \Rightarrow$ сумма уменьшилась \Rightarrow сумма всё ещё $\leq T \Rightarrow$ перейдём по индукции к аналогичной задаче, но без задания t_{min} и временем $T - t_{min}$.

2. Сортировка по дедлайну

Даны n заданий, у каждого есть время выполнения t_i и дедлайн момента сдачи d_i . Нужно успеть выполнить все задания. Решение: выполнять в порядке возрастания дедлайна. Способ придумать решение: посмотрим, кого мы можем выполнить последним? Любого, у кого $d_i \geq \sum_j t_j \Rightarrow$ выполним в конце d_{max} . Доказательство: можем выполнить множество заданий I , выполнив последним $d_{last} \Rightarrow$ (можем выполнить $I_1 = I \setminus \{d_{max}\} \wedge (d_{max} \geq d_{last} \geq \sum_j t_j \Rightarrow$ можем выполнить последним $d_{max}) \Rightarrow$ возьмём по индукции решение для I_1 , допишем к нему d_{max} .

На практике разобраны похожие задачи – с решением “сортировка по сумме”.

Ход мыслей в (1) и (2): как только мы показали, кого ставить на первое/последнее место, отсортируем в таком порядке, получим решение за $\mathcal{O}(sort)$.

3. Сортировка по частному

Даны файлы размеров s_i и частоты обращения к ним cnt_i . Лента – хранилище данных, позволяющее к файлу, начинающемуся на позиции pos обратиться за время pos .

Расположить файлы на ленте, минимизировать суммарное время обращения $\sum_i pos(i)cnt_i$.

Решение для случая $s_i = 1$: сортируем по убыванию cnt_i .

Решение для случая $cnt_i = 1$: сортируем по возрастанию s_i .

Общий случай: сортируем по убыванию $\frac{cnt_i}{s_i}$. Чтобы придумать такое решение предположим, что задача решается сортировкой \Rightarrow у сортировки есть компаратор и компаратор решает задачу для $n = 2 \Rightarrow$ чтобы придумать компаратор полезно исследовать решение при $n = 2$. В порядке $\{1, 2\}$ получаем $cnt_2 s_1$, в порядке $\{2, 1\}$ получаем $cnt_1 s_2 \Rightarrow \text{less}(1, 2): cnt_2 s_1 < cnt_1 s_2 \Leftrightarrow \frac{cnt_1}{s_1} > \frac{cnt_2}{s_2}$. Решение придумано, доказываем: для оптимального ответа верно, что отличимые компаратором less элементы идут в том же порядке, что задаёт less ; осталось понять, что неотличимые ($\frac{cnt_1}{s_1} = \frac{cnt_2}{s_2}$) можно расставить в любом порядке, ответ $\sum_i pos(i)cnt_i$ не изменится.

Алгоритм 9.2.1. *Перестановочный метод:* π – искомая перестановка, $F(\pi) \rightarrow \min$.

В поисках оптимальной π пробуем поменять соседние элемента: $F(\pi) \leq F(\pi[\text{swap}(i, i+1)])$.

Полученное условие используем, как компаратор.

Часто это условие совпадает с более простым: $\text{less}(i, j) = (F([i, j]) < F([j, i]))$

Необходимые условия к компаратору

Когда можно пользоваться перестановочным методом [Algo 9.2.1](#) из (3)?

Проще сформулировать это в терминах компаратора `lessOrEquals`: он должен задавать **линейный предпорядок** (**total preorder**), то есть быть транзитивным и любые два элемента должны быть сравнимы хотя бы в одну сторону.

Если говорить в терминах отношения компаратора `less`, то он должен задавать **strict weak ordering**, условия на который немного сложнее: он должен быть антирефлексивным, транзитивным, и несравнимость должна быть транзитивной. Поясним последнее: если $\neg \text{less}(a, b)$ и $\neg \text{less}(b, a)$, то будем говорить, что a и b несравнимы и писать $a \sim b$. Такое отношение \sim должно быть транзитивным.

Проблемы обычно только с транзитивностью и транзитивностью несравнимости. Чтобы показать транзитивность в (3), мы переписывали компаратор: $(cnt_2 s_1 < cnt_1 s_2) \rightarrow (\frac{cnt_1}{s_1} > \frac{cnt_2}{s_2})$.

Чтобы сортировка по такому компаратору минимизировала $F(\pi)$, достаточно выполнения следующего условия: для всех перестановок π и для каждого i , если $\text{lessOrEquals}(\pi_i, \pi_{i+1})$, то $F(\pi) \leq F(\pi')$, где π' получена из π операцией $\pi_i \leftrightarrow \pi_{i+1}$. (Упражнение: проверьте, что этого действительно достаточно.)

9.2.1. Задача про 2 станка

Задача: даны 2 станка, n деталей, время обработки деталей на первом станке a_i и время обработки деталей на втором станке b_i . Выбрать порядок обработки деталей π , обработать каждую сперва на первом, затем на втором станке. Минимизировать время окончания работ.

Решение: разбить на $P_1 = \{i: a_i \leq b_i\}$ и $P_2 = \{i: a_i > b_i\}$, отсортировать P_1 по возрастанию a , P_2 по убыванию b . $\pi = P_1 P_2$. Оригинальная **статья** (S.M.Johnson'1953).

• (*) Вспомогательные рассуждения

Время окончания выполнения на первом станке $A_i = A_{i-1} + a_{\pi_i}$

Время окончания выполнения на втором станке $B_i = \max(A_i, B_{i-1}) + b_{\pi_i}$

Пусть x_i – ожидание второго станка перед обработкой i -й детали $\Rightarrow B_k = \sum_{i \leq k} (x_i + b_i)$.

Обозначим $F(k) = \sum_{i \leq k} x_i$. $F(n)$ – суммарный простой второго станка. Задача: $F(n) \rightarrow \min$.

При этом $x_k = \max(0, A_k - B_{k-1}) = \left(\sum_{i=1}^k a_i - \sum_{i=1}^{k-1} b_i \right) - \sum_{i=1}^{k-1} x_i = D_k - F_{k-1}$, где $D_i = \sum_{j \leq i} a_j - \sum_{j < i} b_j$.

Докажем по индукции, что $F(k) = \max(0, \max_{i=1..k} D_i)$.

База: $k = 0$. Переход: $x_k > 0 \Rightarrow x_k = D_k - F_{k-1} \wedge F_k = D_k$; $x_k = 0 \Rightarrow F(k) = F(k-1) \wedge D_k \leq F_{k-1}$.

• (*) Корректность алгоритма

Осталось показать, что для минимизации $F(n) = \max_{i=1..n} D_i$, нужно 3 вещи:

1. Если рядом стоят $a_i > b_i$ и $a_{i+1} \leq b_{i+1}$, можно поменять местами, $F(n) \searrow$
2. Если рядом стоят $a_i \leq b_i$ и $a_{i+1} \leq b_{i+1}$: $a_i > a_{i+1}$, можно поменять местами, $F(n) \searrow$
3. Если рядом стоят $a_i > b_i$ и $a_{i+1} > b_{i+1}$: $b_i < b_{i+1}$, можно поменять местами, $F(n) \searrow$

• Неверное решение

Выведем компаратор, как в [Algo 9.2.1](#). $F(1, 2) = a_1 + \max(a_2, b_1) + b_2 = (a_1 + a_2 + b_1 + b_2) - \min(a_1, b_2)$.

Получили “less(i, j): return min(a[i], b[j]) > min(a[j], b[i])”.

Получили транзитивное отношение меньше, которое не является строгим линейным порядком.

Если его передать функции `std::sort`, получится почти всегда рабочий алгоритм.

Стресс-тестом можно найти контрпримеры. Подробнее на [codeforces](#) и на [github](#).

• Внешние ссылки

Можно ознакомиться с [e-maxx](#) и [ИТМО-конспектами](#). По ссылкам рассуждения (доказательства) не полны. При этом по обоим ссылкам рабочая реализация (хоть и переусложнённая).

9.2.2. Выбор максимума

Пример задачи («сортировка по дедлайну»): есть n задач, у каждой есть время, требуемое для выполнения t_i и дедлайн d_i , к которому нужно завершить выполнение задачи.

Мы уже умеем проверить, можно ли выполнить все задачи: отсортировать по d_i и выполнять в таком порядке. Теперь мы хотим успеть выполнить максимальное количество задач.

Замечание 9.2.2. Полученный в этом разделе метод часто позволяет перейти от «выполнить все» сортировкой за $\mathcal{O}(n \log n)$ к «выполнить как можно больше» жадностью за $\mathcal{O}(n \log n)$.

Решение за $\mathcal{O}(n^2)$: если мы зафиксируем множество заданий, которое хотим выполнить, выполнять их нужно в порядке $d_i \uparrow \Rightarrow$ отсортируем по d_i и напишем динамику $T[i, k]$ – минимальное суммарное время, нужное, чтобы выполнить какие-то k заданий из i первых. Переходы:

1. `relax(T[i+1, k], T[i, k]);`
2. `if d[i] >= T[i, k] + t[i] then relax(T[i+1, k+1], T[i, k] + t[i]);`

Решение за $\mathcal{O}(n \log n)$: будем жадно брать в порядке $d_i \uparrow$, если не получается взять, попытаемся заменить на работу с максимальным t_j .

1. `sumTime = 0`
2. `for пары $\langle d_i, t_i \rangle$ в порядке $d_i \uparrow$:`
3. `if sumTime + t_i <= d_i then берём: heap.add(t_i), sumTime += t_i`
4. `else if t_i < heap.max then делаем замену:`
5. `sumTime += t_i - heap.max, heap.delMax, heap.add(t_i)`

В строке 4, когда условие выполнено, верно и $d_i \geq d_j \Rightarrow$ мы лучше и по t , и по $d \Rightarrow$ после

замены порядок выполнения оставим прежним, а себя вставим ровно на место удалённого.

• **Корректность решения за $\mathcal{O}(n \log n)$**

Пусть на момент рассмотрения пары $\langle d_i, t_i \rangle$ жадность считает, что нужно выполнять задания $t_{j_1} \leq t_{j_2} \leq \dots \leq t_{j_m}$. Мы специально записали эти задания в порядке $t_j \uparrow$. Докажем по индукции, что, $\forall k \ T[i, k] = t_{j_1} + \dots + t_{j_k}$, более того множество этих k предметов есть $\{j_1, \dots, j_k\}$.

База: $i = 0 \Rightarrow m = 0$.

Переход: пользуемся корректностью динамики $T[i+1, k] = T[i, k] \vee T[i, k-1] + t_i$.

Заметим, что для всех таких k , что $t_{j_k} \leq t_i$ оптимально $T[i, k]$ и наше решение как раз его выберет. А теперь заметим **TODO**, что для всех таких k , что $t_{j_k} > t_i$ второе возможно и оптимальнее первого.

Наше решение его и выберет. Для удобства считаем $t_{j_{m+1}} = +\infty$.

В сумме получаем, что $\forall k \ T[i, k+1]$ – префикс $t_{j_1} \leq t_{j_2} \leq \dots \leq t_i \leq \dots \leq t_{j_m}$.

Замечание 9.2.3. То же применимо к задаче про спортсменов:

$s_i + m_i \geq s_j + m_j$, $m_i < m_j \Rightarrow s_i > s_j \Rightarrow$ можно вместо j вставить i на то же место \Rightarrow наше рассуждение (доказательство) выше работает и для этой задачи.

9.3. Жадность для гамильтонова пути. Варнсдорф

Варнсдорф решил задачу гамильтонова цикла для шахматного коня. Правило Варнсдорфа говорит: нужно идти в вершину минимальной остаточной степени, а при равных прижиматься к краю (то ли по $\min(dx, dy) \rightarrow \min$, то ли $|dx| + |dy| \rightarrow \min$). В какой вершине начать: в случайной точно подойдёт.

Варнсдорф, конечно, решал для 8×8 , но его решение подходит и для $n \times n$.

Как это счастье применить к общему случаю?

1. Жадно идти в вершину минимальной остаточной степени!
2. Можно хитрее: random walk (тоже жадность), который пытается жадно идти в вершину минимальной остаточной степени, а из равных выбирает случайную.
3. Можно ещё хитрее: random walk (всё ещё жадность), который выбирает случайную вершину куда идти с вероятностью обратно пропорциональной остаточной степени. $p_v = \frac{1}{deg_v}$.

Лекция #10: Приближённые алгоритмы

15–20 апреля 2022

Как мы знаем, некоторые задачи NP-трудны, вряд ли их возможно решить за полином. Но на практике решать их всё равно нужно. В таких случаях часто ищут приближённое решение, причём ищут за полиномиальное или даже линейное время.

Def 10.0.1. Алгоритм M называется α -оптимальным (α -OPT) для задачи P , если на любом входе для P ответ, выдаваемый M , не более, чем в α раз хуже оптимального.

10.1. Коммивояжёр

Задача коммивояжёра заключается в поиске гамильтонова цикла минимального веса. Решим задачу в случае полного графа с неравенством Δ .

Lm 10.1.1. $\forall \alpha > 1 \nexists$ полиномиального α -OPT алгоритма для задачи коммивояжёра.

Доказательство. Если такой существует, то он в частности ищет гамильтонов путь. ■

Замечание 10.1.2. Любой граф можно сделать полным, добавив вместо отсутствия ребра ребро веса $+\infty$. А вот неравенство Δ – важное свойство.

10.1.1. 2-OPT через MST

Оптимальный маршрут коммивояжёра = цикл = путь + ребро = остовное дерево + ребро \Rightarrow **MST \leq OPT**. Возьмём MST графа. Каждое ребро заменим на два ориентированных (туда и обратно), получили эйлеров орграф (число входящих равно числу исходящих).

Возьмём эйлеров обход, его вес в два раза больше MST. Если в полученном пути некоторая вершина встречается второй раз, удалим её второе вхождение.

По нер-ву Δ вес пути только уменьшится. Получили маршрут коммивояжёра веса $\leq 2 \cdot$ MST.

10.1.2. 1.5-OPT через MST (Кристофидес)

Чтобы превратить MST в эйлеров граф, добавим к нему M – совершенное паросочетание минимального веса на вершинах нечётной степени. $w(MST) \leq$ OPT, докажем что $w(M) \leq \frac{1}{2}$ OPT.

$\upharpoonright X$ – маршрут коммивояжёра только на нечётных вершинах. По нер-ву Δ $w(X) \leq$ OPT.

Цикл X можно разбить на два паросочетания – A (чётные рёбра) и B (нечётные рёбра).

$$w(A) + w(B) = w(X) \Rightarrow \min(w(A), w(B)) \leq \frac{1}{2}w(X) \leq \frac{1}{2}OPT$$

Lm 10.1.3. \exists полиномиальный алгоритм для поиска паросочетания минимального веса в произвольном графе. Кстати, умеют это делать весьма быстро, за $\mathcal{O}(E\sqrt{V} \log(VN))$. *PDF 2017.*

10.2. Set cover

Задача: даны множества A_1, A_2, \dots, A_m , хотим покрыть U : $|U| = n$ минимальным числом A_i .

Шаг жадности: добавить в ответ множество A_i , покрывающее максимальное число ещё не покрытых элементов U .

Упражнение 10.2.1. Придумать реализацию за $\mathcal{O}(n + \sum |A_i|)$.

Теорема 10.2.2. Получили $(\ln n)$ -OPT алгоритм.

Доказательство. Пусть оптимальный ответ равен k .

За один шаг жадность покрывает $\geq \frac{1}{k}n$ элементов $U \Rightarrow$

размер U уменьшится в $\frac{k-1}{k}$ раз \Rightarrow за t шагов размер U уменьшится в $(\frac{k-1}{k})^t$ раз.

Осталось взять t : $(\frac{k-1}{k})^t < \frac{1}{n} \Leftrightarrow (\frac{k}{k-1})^t > n \Leftrightarrow t > \frac{1}{\ln k - \ln(k-1)} \ln n > k \ln n$. ■

Упражнение 10.2.3. Найдите $(\ln n)$ -ОПТ приближение взвешенной задачи – вес A_i равен w_i .

10.3. Рюкзаки

Будем пользоваться англоязычными обозначениями задач:

1. **Partition** – разбиение множества предметов на два равных по весу.
Даны a_i , выбрать I : $\sum_{i \in I} a_i = \frac{1}{2} \sum a_i$.
2. **Knapsack** – поместить в рюкзак размера S предметы максимальной суммарной стоимости.
Даны $a_i > 0$, $p_i > 0$ и S , выбрать $0 \leq x_i \leq 1$: $\sum a_i x_i \leq S$, при этом $\sum p_i x_i \rightarrow \max$.
 $x_i \in \mathbb{R} \Rightarrow$ задача называется непрерывным рюкзаком, решается сортировкой по $\frac{p_i}{a_i}$.
 $x_i \in \mathbb{Z} \Rightarrow$ получили классический NP-трудный дискретный рюкзак (knapsack).
3. **Bin packing** – уложить предметы в минимальное число рюкзаков размера 1.
Даны $0 < a_i \leq 1$, выбрать I_1, I_2, \dots, I_k : $\cup I_j = \{1..n\} \wedge \forall j \sum_{i \in I_j} a_i \leq 1$. При этом $k \rightarrow \min$.

• NP-трудность

Базовой NP-трудной у нас будет задача partition.

Knapsack труден, так как можно взять все $p_i = a_i$, $S = \frac{1}{2} \sum a_i$ и решить partition.

Bin packing труден, так как если научимся отличать ответы 2 и 3, решим partition.

Следствие 10.3.1. $\forall \varepsilon > 0$ для bin packing $\nexists (1.5 - \varepsilon)$ -ОПТ алгоритма.

10.4. Схемы приближений

Для некоторых задач доступны сколь угодно точные приближения за полиноми. **PTAS/FPTAS**.

Def 10.4.1. Алгоритм $A(\varepsilon, x)$ – PTAS (polynomial-time approximation scheme) для задачи M , если $\forall \varepsilon > 0$ он отработает за полином от $|x|$ и выдаст решение F такое, что:

$$\begin{cases} F \geq (1-\varepsilon)OPT & \text{если } M \text{ – задача максимизации} \\ F \leq (1+\varepsilon)OPT & \text{если } M \text{ – задача минимизации} \end{cases}$$

Def 10.4.2. FPTAS (fully PTAS) обязывает A работать за полином и от $|x|$, и от ε^{-1} .

10.5. Knapsack

Здесь можно почитать про PTAS и FPTAS для **knapsack**.

Задача: даны $a_i > 0$, $p_i > 0$, выбрать $0 \leq x_i \leq 1$: $\sum a_i x_i \leq S \wedge \sum p_i x_i \rightarrow \max$.

Если $x_i \in \mathbb{R}$ задача называется непрерывной и решается жадно сортировкой по $\frac{p_i}{a_i}$.

Если $x_i \in \mathbb{Z}$ задача называется дискретной или 0-1 и жадное решение ничего не приближает.

Пример: $a_1 = 1$, $p_1 = 2$, $a_2 = S$, $p_2 = S \Rightarrow$ жадность возьмёт первый предмет, оптимально брать второй. Ошибка в $S/2$ раз \Rightarrow не ограничена.

Замечание 10.5.1. Во всех решениях дальше будем считать, что мы сразу выкинули предметы, которые даже по отдельности не помещаются в рюкзак, то есть у которых $x_i > S$.

На практике разобрали 2-приближение: выбираем лучшее из жадности и первого предмета, который уже не влез.

Но можно приблизить и PTAS, и FPTAS.

- **PTAS:** переберём, какие k предметов возьмём за $\binom{n}{k}$, оставшиеся наберём жадностью.

Теорема 10.5.2. Мы получили $(1 - \frac{1}{k})$ -OPT решение, работающее за $\mathcal{O}(n^{k+1})$

Доказательство. Обозначим оптимальное решение I . Рассмотрим момент, когда мы угадали k max по p_i предметов, присутствующих в I : $i_1 \dots i_k$. Посмотрим теперь, как жадность лихо добавляет лучшие по $\frac{p_i}{a_i}$ предметы и остановимся на первом событии «жадность не взяла предмет $j \in I$ ». В этот момент остановим жадность, полученное решение обозначим I' . $I' \cup \{j\}$ не хуже OPT: i_1, \dots, i_k, j – общая часть, остальные предметы I' мажорируют I по $\frac{p_i}{a_i}$. Осталось заметить, что $p_j \leq \frac{1}{k}(p_{i_1} + \dots + p_{i_k}) \leq \frac{1}{k}\text{OPT} \Rightarrow P(I') \geq (1 - \frac{1}{k})P(I)$, где $P(I) = \sum_{i \in I} p_i$. ■

Замечание 10.5.3. На самом деле, перебирать надо и меньшие множества тоже, поскольку в ответе может быть меньше, чем k предметов.

- **FPTAS:** возьмём псевдополиномиальную динамику `minWeight[i, sumP]`, и чтобы она работала за строгий полином. сделаем все $p_i \in \{0, 1, \dots, k\}$: $p'_i = \lfloor \frac{p_i}{\max_j p_j} k \rfloor = \lfloor \frac{1}{x} p_i \rfloor$, где $x = \frac{\max_j p_j}{k}$. Решение работает за $\mathcal{O}(n^2 k)$, оценим ошибку.

$$\text{Пусть } P(\text{set}) = \sum_{i \in \text{set}} p_i, \text{ а } P'(\text{set}) = \sum_{i \in \text{set}} p'_i = \sum_{i \in \text{set}} \lfloor \frac{p_i}{x} \rfloor.$$

Сразу заметим, что $P'(\text{set}) \leq \frac{P(\text{set})}{x} \leq P'(\text{set}) + n$, так как мы теряем на округлении, но не более, чем единицу на каждый предмет.

Пусть opt это оптимальное решение, а мы нашли sol . Тогда $P'(sol) \geq P'(opt)$, так как мы нашли самое лучшее решение в профитах p' .

$$\begin{aligned} P(sol) &\geq xP'(sol), P'(sol) \geq P'(opt), P'(opt) \geq \frac{P(opt) - nx}{x} \\ \implies P(sol) &\geq xP'(sol) \geq xP'(opt) \geq P(opt) - nx \end{aligned}$$

Получили решение, которое хуже не более чем на nx . Хотим: $nx \leq \varepsilon P(opt)$.

Знаем, что $P(opt) \geq \max_j p_j$. Выберем x , так что $\varepsilon(\max_j p_j) \geq nx$, тогда получаем требуемое.

$$x \leq \frac{\varepsilon(\max_j p_j)}{n} \implies k \geq \frac{n}{\varepsilon}$$

Замечание 10.5.4. Аналогичная динамика `maxProfit[i, sumW]` не дала бы нам хорошее приближение: чтобы решение оставалось подходящим веса нужно округлять вверх, после минимального округления вверх предметы могут перестать влезать \implies ответ ухудшился в 2 раза.

10.6. Partition

Задача NP-трудна. Эквивалентная ей – **subset sum**.

Оптимизационная версия partition – *balanced partition*: разбить на два максимально близких по сумме множества.

Обе задачи мы умеем решать за $\mathcal{O}^*(2^{n/2})$ методом meet-in-the-middle.

⚡ приближённых по величине $D = |\sum a_i - \sum b_i|$ решений (NP-трудно отличить $D = 0$ от $D > 0$). Но делать что-то надо, поэтому строят приближённые решения по величине $\max(A, B)$, $A = \sum a_i$, $B = \sum b_i$.

Тривиальное жадное решение за $\mathcal{O}(n \log n)$: перебирать предметы в порядке $a_i \downarrow$, класть очередной предмет в множество с меньшей суммой.

Lm 10.6.1. При $n \leq 4$ решение даст оптимальный ответ

Доказательство. Пусть $a_1 > a_2 > a_3 > a_4$, тогда варианты $a_1 + a_2$ и $a_1 + a_3$ точно хуже $\max(a_1 + a_4, a_2 + a_3)$, а наша жадность за первые три хода как раз получит $\{a_1\}, \{a_2, a_3\}$, четвертое положит оптимально. ■

Пусть $n \geq 5$. Будем следить за величиной $D = |A - B|$. Заметим, что если после обработки $k \geq 4$ предметов D стало не больше a_5 , то и дальше оно будет всегда не больше a_5 . Тогда уменьшить величину $\max(A, B)$ можно не больше чем на $d = \frac{1}{2}a_5$. Обозначим $X = \max(A, B)$, $X \geq 3a_5 \Rightarrow d \leq \frac{1}{6}X \Rightarrow X \leq \text{OPT} + \frac{1}{6}X \Rightarrow X \leq \frac{6}{5}\text{OPT}$.

Однако, могло случиться так, что D так и не стало достаточно маленьким (после обработки четвертого предмета и дальше). В этом случае все предметы, начиная с пятого, мы клали в одну и ту же коробку, пусть это B , и всё равно $B_n < A_n$. Тогда $\max(A_n, B_n) = A_n = A_4 = \max(A_4, B_4)$, и при этом для любого размещения предметов $A_n \geq A_4$. Мы знаем, что A_4 было минимально, поэтому A_n тоже минимально. То есть в этом случае жадность находит оптимальный ответ.

Таким образом, наша жадность $\frac{6}{5}$ -оптимальна.

Замечание 10.6.2. Более аккуратным анализом можно показать, что это $\frac{7}{6}$ -OPT решение.

PTAS схема: переберём, куда кладём $2k$ максимальных по весу предметов, остаток разложим, следуя предыдущей жадности. Получили $(1 + \frac{1}{k})$ -OPT решение. Время $\mathcal{O}(2^k n) \Rightarrow$ не FPTAS.

Упражнение 10.6.3. Доказательство осталось на практику

10.6.1. Алгоритм Кармаркара-Карпа (LDM)

[\[wiki\]](#)

Задача partition: разбить предметы на два рюкзака, минимизировать $D = |\sum a_i - \sum b_i|$. Мы уже умеем решать её жадностью [Lm 10.6.1](#).

Алгоритм Кармаркара-Карпа (1982)

```

1 def solve(a):
2     while size(a) > 1:
3         x = a.extractMax()
4         y = a.extractMax()
5         a.push(x - y)
6     return a.extractMax()

```

Описанный алгоритм ещё называют LDM (largest difference method). Он даёт $\frac{7}{6}$ -приближение (как и жадность).

Теперь посмотрим на величину $D = |A - B|$. На n случайных равномерно распределённых величинах в $\mathbb{R} \cap [0, 1]$ жадность даст $E(D) = \Theta(\frac{1}{n})$, Кармаркар-Карп $n^{-\Theta(\log n)}$. Ту же асимптотику

матожидания разности Кармаркар-Карп даст на любых независимых одинаково распределённых случайных величинах.

[Yakir'1996] – доказательство.

[Mertens'1999], [Boettcher, Mertens'2008] – более глубокий анализ.

Интуиция, почему Кармаркар-Карп даёт $n^{-\Theta(\log n)}$.

Представим a в отсортированном порядке. Сейчас там n чисел из $[0, 1]$, за первые $\frac{n}{2}$ шагов мы заменим их на разности соседних, получим $\frac{n}{2}$ чисел из $[0, \frac{1}{n} + o(\frac{1}{n})]$.

Повторим процесс $\log n$ раз, последнее число будет $\approx \frac{1}{n} \cdot \frac{1}{n/2} \cdot \frac{1}{n/4} \dots = n^{-\Theta(\log n)} = 2^{-\Theta(\log^2 n)}$.

10.6.2. (*) Применение для $\langle P, M \rangle$ -антихеш-теста

Найдём 2 строки длины n над алфавитом $\{a, b\}$ с равными полиномиальными $\langle P, M \rangle$ -хешами.

$$\text{hash}(s) = (s_0 + s_1 P^1 + s_2 P^2 + \dots) \bmod M$$

$$\text{hash}(s) = \text{hash}(t) \Leftrightarrow \text{hash}(s) - \text{hash}(t) = \sum x_i P^i \bmod M = 0, \text{ где } x_i \in \{0, \pm 1\}$$

Итого: нам даны числа $P^i \bmod M$, хотим некоторые из них разбить на два равных рюкзака.

Применяем Кармаркар-Карпа, пока в множестве a не встретится 0.

Оценка чисел в a в конце Кармаркар-Карпа:

$M \cdot 2^{-\Theta(\log^2 n)} \Rightarrow \log \log n = \Theta(\sqrt{\log M})$ возьмём $n = 2^{\Theta(\sqrt{\log M})}$, на практике константа будет 1.5. Для $M \approx 10^{18}$ имеем $\log M = 64$, $\sqrt{\log M} = 8$, $n = 2^{1.5 \cdot 8} = 4096$.

10.7. Bin packing

[Garey, Johnson'1997] Оценки на first fit (FFD), best fit (BFD, BBF).

[PTAS] Более-менее случайная презентация.

[Hoberg'2015] $OPT + \mathcal{O}(\log OPT \log \log OPT)$ приближение.

[Korf'1999] Решение задачи на практике.

[Jansen et al.'2010] FPT-решение за $\mathcal{O}(2^{k \log^2 k} + n)$, ошибающееся \leq чем на 1 корзину.

Напомним, формулировку задачи Section 10.3 и, что отличить ответы 2 и 3 уже трудно Cons 10.3.1.

Распределять по k корзинам сложнее, чем по двум (partition).

Динамикой мы умеем решать задачу лишь для \mathbb{Z} весов, и то за $\mathcal{O}(n \cdot W^{k-1})$.

• First fit

Перебираем предметы в произвольном порядке.

Для каждого предмета перебираем корзины, кладём предмет в первую подходящую.

Теорема 10.7.1. *First fit – 2-OPT алгоритм для bin packing*

Доказательство. Если наш ответ m , то $m - 1$ корзина заполнена больше чем на половину. Итого $\sum a_i > \frac{1}{2}(m - 1)$. С другой стороны $OPT \geq \sum a_i \Rightarrow 2OPT > m - 1 \Rightarrow 2OPT \geq m$. ■

First fit ещё можно оценить как $1.7OPT + 2$.

Конечно, есть более удачные порядки перебора предметов чем “произвольный”.

И в теории, и на практике хорошо работает порядок “по убыванию”.

Теорема 10.7.2. *First fit decreasing (FFD) для задачи bin packing даёт ответ $\leq \frac{11}{9}OPT + 1$*

Доказательство. Можно прочесть по ссылкам выше. Кстати, это была PhD Джонсона [Johnson'73]. Чуть более сильный ($\frac{11}{9}OPT + \frac{6}{9}$) результат [Dosa'2007].

При этом в Cons 10.3.1 мы уже говорили, что нет $(1.5-\varepsilon)$ -OPT алгоритма.

• Best fit

Эвристика: класть в корзину, в которой после добавления предмета останется поменьше места.

BBF – перебирать при этом предметы в произвольном порядке даёт $1.7OPT + 1$.

BFD – перебирать при этом предметы в убывающем порядке даёт $\frac{11}{9}OPT + 1$.

• Практически эффективное решение

Обе жадности уже хороши тем, что работают за $\mathcal{O}(n \log n)$.

Есть чуть менее быстрое, но более эффективное решение:

- зафиксируем число корзин k (если задача минимизации k , сделаем бинпоиск по ответу);
- начнём со случайного распределения предметов по k корзинам
- минимизируем максимальный размер корзины локальными оптимизациями двух типов – пытаемся перекинуть предмет из одной корзины в другую; пытаемся поменять два предмета в двух корзинах.
- для эффективной реализации второй оптимизации перебираем предметы в порядке $a_i \uparrow$ и пытаемся менять местами только соседние – a_i и a_{i-1} .

```

1 // p[i] - номер корзины по предмету, sumA[j] - суммарный вес в корзине
2 while optimizing // одна фаза оптимизации работает за O(n)
3   min_index <-- номер корзины с минимальным весом
4   for i = 1..n
5     if sumA[p[i]] - a[i] > sumA[min_index]
6       p[i] = min_index, recalc sumA
7   for i = 2..n // a[i] возрастают
8     if sumA[p[i]] - a[i] > sum[p[i-1]] - a[i-1]
9       swap(p[i], p[i-1]), recalc sumA

```

• (*) PTAS для bin packing

Напомним задачу: даны $0 < a_i \leq 1$, разбить их на min число корзин размера 1.

• План решения

1. Сперва рассмотрим ситуацию, когда все $a_i \geq \varepsilon$ и различных a_i не более K
2. Отбросим ограничение “не более K ” (округлим все a_i к одному из K значений).
3. Отбросим ограничение $a_i \geq \varepsilon$ (в начале выкинем $a_i < \varepsilon$, в конце жадно добавим).

1 Итак, все $a_i \geq \varepsilon$ и различных a_i не более $K \Rightarrow$ в каждую корзину влезает не более $M = \frac{1}{\varepsilon}$ предметов \Rightarrow различных типов корзин не более $R = K^M$. Нам не нужно больше n корзин, есть не более $P = \binom{n+R-1}{R}$ вариантов выбрать n корзин R типов.

Заметим, что $1/\varepsilon, K, M, R$ – константы. Огромные, но константы.

$P \leq \frac{1}{R!}(n+R)^R \Rightarrow P$ – полином от n , переберём все P вариантов, выберем оптимальный.

2 Выберем $K = \lceil \frac{1}{\varepsilon^2} \rceil$. Отсортируем a_i , разобьём их на K групп, в каждой не более $\lfloor n\varepsilon^2 \rfloor$ элементов. Изменим в каждой группе все числа на максимальное, теперь различных $\leq K$.

Исходная задача – A , новая – A' . A' поэлементно больше $A \Rightarrow \forall$ распределение для A' является распределением для A . Рассмотрим A'' , полученную из A' удалением $\lfloor n\varepsilon^2 \rfloor$ максимальных.

A'' поэлементно меньше $A \Rightarrow k(A'') \leq k(A) \Rightarrow k(A') \leq k(A'') + \lfloor n\varepsilon^2 \rfloor \leq k(A) + \lfloor n\varepsilon^2 \rfloor$.

$\forall i a_i \geq \varepsilon \Rightarrow OPT \geq n\varepsilon \Rightarrow k(A') \leq k(A) + \varepsilon \cdot OPT = (1 + \varepsilon)OPT$.

3 I – исходная задача, I' – задача без $a_i < \varepsilon$. Мы уже получили решение для I' не хуже $(1 + \varepsilon)k(I')$. Добавим, используя любую жадность, например FF, все $a_i < \varepsilon$.

Если ответ не изменился, отлично $k(I') \leq k(I) \Rightarrow (1 + \varepsilon)k(I') \leq (1 + \varepsilon)k(I)$.

Если ответ изменился, то все корзины кроме последней заполнены больше чем на $1 - \varepsilon \Rightarrow (1 - \varepsilon)(k(I') - 1) + 1 \leq \sum_{\text{все}} a_i \leq k(I) \Rightarrow k(I') \leq \frac{1}{1 - \varepsilon}(k(I) - 1) + 1 \leq (1 + 2\varepsilon)k(I) + 1$.

Замечание 10.7.3.

В части (1) мы получили точное решение.

В части (2) мы получили $(1 + \varepsilon)$ приближении, и лишь

В части (3) приближение стало $(1 + 2\varepsilon)$ -OPT + 1.

10.8. Надстрока

TODO: гипотеза о надстроке

10.9. Литература

[Книга по приближённым алгоритмам]. Approximation algorithms, Vazirani '2001.

[Knapsack | PTAS FPTAS]. Лекция от Goeman-а.

[Euclidian TSP | PTAS]. Sanjeev Arora '1998.

[Permanent FPRAS]

Лекция #11: Центроидная декомпозиция

20 апреля 2022

11.1. Построение и минимум на пути дерева

Пусть дано дерево с весами на рёбрах или в вершинах. Рассмотрим три задачи вида «построить структуру данных, которая умеет за $\mathcal{O}(1)$ вычислять минимум...»

1. на пути от любой вершины до корня;
2. на любом пути, проходящем через корень;
3. на любом пути

Для решения первой задачи поиском в глубину по дереву предподсчитаем все минимумы. Вторую задачу сведём к первой, разделив корнем путь на две половины.

Def 11.1.1. *Центроид* – вершина, при удалении которой, размеры всех компонент $\leq \frac{n}{2}$.

Чтобы решить третью задачу найдём за $\mathcal{O}(n)$ центроид v (задача с практики); все пути, проходящие через v обработаем, используя (2); удалим v из дерева и рекурсивно построим структуру от оставшихся компонент.

В итоге каждая вершина ровно один раз будет найдена, как центроид. У каждого центроида v есть центроид-предок p_v в дереве рекурсии, глубина в дереве рекурсии d_v и компонента связности $C(v)$, в которой он был найден, как центроид.

Def 11.1.2. *Центроидной декомпозицией* будем называть два массива – p_v и d_v . Первый из массивов задаёт так называемое “дерево центроидной декомпозиции”.

Замечание 11.1.3. Зная только глубины d_v , можно восстановить компоненты $C(v)$. Для этого нужно от центра v запустить dfs, который проходит только по вершинам большей глубины. Поэтому $C(v)$ хранить **не нужно**.

Lm 11.1.4. Для центроидной декомпозиции нужно $\Theta(n)$ памяти.

Lm 11.1.5. Глубина дерева центроидной декомпозиции не более $\log_2 n$.

Lm 11.1.6. Суммарный размер компонент $\mathcal{O}(n \log n)$.

Следствие 11.1.7. В задаче поиска min на пути суммарный размер предподсчёта $\mathcal{O}(n \log n)$.

Если декомпозиция уже построена, можем посчитать все минимумы от u до вершин $C(u)$ с помощью dfs:

```

1 // c - центроид
2 void calc_min(int c, int v, int parent = -1, int curMin = INT_MAX) {
3     curMin = min(curMin, w[v]); // пусть веса в вершинах
4     f[d[c]][v] = curMin; // на уровне d[c] каждая вершина v встретится ≤ 1 раза
5     for (int x : graph[v])
6         if (x != parent && d[x] >= d[c])
7             calc_min(c, x, v, curMin);
8 }
9 calc_min(u, u);

```


Лемма 11.1.8. Для любого пути $[a..b]$ на дереве, есть такой центр v : $v \in \text{path}[a..b]$ и $a, b \in C(v)$.

Чтобы найти такой центр v рассмотрим пути от a и b вверх в дереве центроидной декомпозиции и возьмём наименьшего общего предка (LCA) a и b – самый нижний такой центр, что $a, b \in C_v$.

$$\text{минимум на пути}[a..b] = \min(f[d_v, a], f[d_v, b])$$

Сейчас мы умеем LCA искать только за \mathcal{O} (глубины дерева). В нашем случае это $\mathcal{O}(\log n)$.

На практике научимся за $\mathcal{O}(\log \log n)$, а в скором будущем вообще за $\mathcal{O}(1)$.

11.2. Реализация

```

1 // Эта функция нужна только, чтобы посчитать размер компоненты связности
2 int calc_size(int v, int parent = -1) {
3     int size = 1;
4     for (int x : graph[v])
5         if (x != parent && d[x] == -1)
6             size += calc_size(x, v);
7     return size;
8 }
9 // Зная размер дерева n, считаем размеры поддеревьев, параллельно ищем центроид
10 int get_centroid(int v, int parent, int n, int &centroid) {
11     int size = 1;
12     for (int x : graph[v])
13         if (x != parent && d[x] == -1)
14             size += get_centroid(x, v, n, centroid);
15     if (size * 2 >= n && centroid == -1)
16         centroid = v;
17     return size;
18 }
19 // Главная функция построения, v - любая вершина компоненты
20 void build(int v, int parent, int depth) {
21     int centroid = -1;
22     get_centroid(v, -1, calc_size(v), centroid);
23     d[centroid] = depth, p[centroid] = parent;
24     precalc_min(depth, centroid); // собственно насчитали минимумы, см. ниже
25     for (int x : graph[centroid])
26         if (d[x] == -1)
27             build(x, centroid, depth + 1);
28 }
29 d = vector<int>(n, -1);
30 p = vector<int>(n, -1);
31 build(0, -1, 0);

```

На самом деле, чтобы решать задачи, нам не нужна прям центроидная декомпозиция, нам нужно выбирать вершину `centroid` так, чтобы глубина дерева декомпозиции была $\leq \log_2 n$.

Благодаря этому, можно обойтись без `calc_size`:

```

1 void build(int v, int parent, int depth, int n) {
2     int centroid = -1;
3     get_centroid(v, -1, n, centroid); // n - оценка сверху на размер компоненты
4     // если оценка наглая, if мог не сработать, но тогда v подходит
5     if (centroid == -1) centroid = v;
6     ...
7     build(x, centroid, depth + 1, (n + 1) / 2);
8 }

```

И, наконец, часть, которая ищет минимум (немного иначе написанная, чем было выше). Нужно для каждого центроида насчитать dfs-ом минимумы на путях до корня.

```
1 f = vector<vector<int>>(LOG, vector<int>(n));
2 // depth - глубина центроида
3 void precalc_min(int depth, int v, int parent = -1, int curMin = INT_MAX) {
4     curMin = min(curMin, weight[v]); // пусть веса будут в вершинах
5     f[depth][v] = curMin; // на каждой глубине v встретится не более раза
6     for (int x : graph[v])
7         if (x != parent && d[x] == -1) // не ходим в уже помеченные вершины
8             calc_min(depth, x, v, curMin);
9 }
10 int get_min(int a, int b) {
11     int lca = get_lca(a, b); // в дереве центроидной декомпозиции, заданной p[]
12     return min(f[d[lca]][a], f[d[lca]][b]);
13 }
```

Лекция #12: Модели вычислений

19 марта

12.1. RAM

TODO

12.2. RAM-w

TODO

12.3. Машина Тьюринга

TODO

Лекция #13: BST и AVL

27 апреля 2022

13.1. BST, базовые операции

Хеш-таблицы идеально умеют хранить неупорядоченные множества: `find`, `add`, `del` за $O(1)$. Для хранения упорядоченных по ключу x_i пар $\langle x_i, data_i \rangle$ будем использовать BST:

Def 13.1.1. *BST – binary search tree (бинарное дерево поиска). В каждой вершине пара $\langle x, data \rangle$. Левое поддерево содержит пары $\langle x, data \rangle$ со строго меньшими x . Правое поддерево содержит пары $\langle x, data \rangle$ со строго большими x .*

Как видно из определения, мы пока предполагаем, что все x_i различны.

$data_i$ – просто дополнительные данные. Например $\langle x_i, data_i \rangle$ – студент, x_i – имя студента.

Хранить вершины дерева будем, как `struct Node { int x; Data data; Node *p, *l, *r; };`

- `v->l (left)` – левый сын
- `v->r (right)` – правый сын
- `v->p (parent)` – отец

Отсутствие сына/отца будем обозначать нулевым указателем. Если мы не пользуемся отцом, то для экономии памяти и времени, хранить и поддерживать его не будем.

Find(x). Основная операция в дереве поиска – поиск. Чтобы проверить, присутствует ли ключ в дереве, спускаемся от корня вниз: если искомый меньше, идём налево, иначе направо.

Add(x). Операция добавления – делаем то же, что и `find`, в итоге или найдём x , или выйдем за пределы дерева. В то место, где мы вышли за пределы дерева, и вставим новое значение.

По BST можно за линию построить сортированный массив значений. Для этого нужно сделать так называемый «симметричный обход дерева» – рекурсивно обойти левое поддерево, выписать x , рекурсивно обойти правое поддерево. Через полученный массив можно определить:

- `next/prev(v)` – следующая/предыдущая в отсортированном порядке вершина дерева

Next(v). Если есть правый сын, спуститься в него и до упора влево, иначе идти вверх, пока не найдём больший ключ.

Del(x). Сперва сделаем `find`. Если у вершины не более одного ребёнка, её очень просто удалить – переподвесим своего ребёнка отцу. Если у вершины v два ребёнка, то перейдём в `next(v)` (идём вправо и до упора влево), сделаем `swap(v->x, next(v)->x)` и удалим `next(v)`. Заметим, что у `next(v)` точно нет левого ребёнка, так как, чтобы её найти, мы спустились до упора влево.

Пока преимуществом над хеш-таблицей является только `next`.

Вскоре появятся и другие операции, например, на практике сделаем `Node* lower_bound(x)`.

Все описанные операции: `add`, `del`, `find`, `next`, `lower_bound` работают за глубину дерева.

Чтобы в этом был толк, скоро изучим способы поддерживать глубину $O(\log n)$.

• Ускорение некоторых операций

По сути BST можно спарить с хеш-таблицей и двусвязным списком.

- `find` за $O(1)$: хеш-таблица, для каждого x хранящая `Node* v`, содержащую x .
- `next/prev` за $O(1)$: добавляем на вершинах двусвязный список, это называют *прошивкой*.

- `del` за $\mathcal{O}(1)$: он у нас выражался, как `find + next + $\mathcal{O}(1)$` .
- А ещё удалять можно лениво: `find + $\mathcal{O}(1)$` .

Итого все операции кроме `add` и `lower_bound` можно сделать за $\mathcal{O}(1)$.

• `std::set<int>`

`std::set<int>` – какое-то BST (на самом деле RB-tree).

`std::set<int>::iterator` – указатель на `Node` этого BST.

Заметим, что при модификации BST (`add, del`) меняются только указатели между вершинами \Rightarrow если у нас был итератор (`Node* it`), после модификаций он останется корректным итератором. С помощью прошивки мы теперь умеем делать быстрые `++` и `-` для рукописного итератора.

• Равные ключи

Как поступать при желании хранить равные ключи? Есть несколько способов.

1. Самый общий – сказать, что равных не бывает: задем x_i на пару $\langle x_i, i \rangle$.
2. Можно в вершине хранить пару $\langle x, count \rangle$ – сколько раз встречается ключ x . Если рядом с ключом есть дополнительная информация (например, мы храним в дереве студентов, а ключ – имя студента), то нужно хранить не число `count`, а список равных по ключу объектов (равных по имени студентов).
3. Можно ослабить условие из определения, хранить в правом поддереве «больше либо равные ключи». К сожалению, это работает не для всех описанных ниже реализаций.
4. Можно ещё сильнее ослабить определение – равные разрешать хранить и слева, и справа.

13.2. Немного кода

```

1 Node* next(Node* v) {
2     if (v->r != 0) { // от правого сына спускаемся до упора влево
3         v = v->r;
4         while (v->l)
5             v = v->l;
6         return v;
7     } else { // поднимаемся вверх, пока не окажемся левым сыном
8         while (v->p && v->p->r == v)
9             v = v->p;
10        return v->p; // возвращает null, если next не существует
11    }
12 }
```

• Добавление

```

1 Node* add(Node* v, int x) { // v -- корень поддерева
2     if (!v)
3         return new Node(x); // единственная новая вершина
4     if (x < v->x)
5         v->l = add(v->l, x);
6     else
7         v->r = add(v->r, x);
8     return v;
9 }
```

Более короткая версия, использующая ссылку на указатель:

```

1 void add(Node* &v, int x) {
2     if (!v)
3         v = new Node(x);
4     else
5         add(x < v->x ? v->l : v->r, x);
6 }

```

Алгоритм 13.2.1. *Персистентная версия. Версия, которая не портит старые вершины.*

```

1 Node* add(Node* v, int x) {
2     if (!v)
3         return new Node(x);
4     else if (x < v->x)
5         return new Node(v->x, add(v->l, x), v->r);
6     else
7         return new Node(v->x, v->l, add(v->r, x));
8 }

```

Эта версия прекрасна тем, что уже созданные Node-ы никогда не изменяются (**immutable**). Поэтому можно писать так: `Node *root2 = add(root1, x)`, после чего у вас есть два корректных дерева – старое с корнем в `root1` и новое с корнем в `root2`. Да, эти деревья пересекаются по вершинам. Но любую операцию `add/find/lower_bound` мы можем проделать с любым из них.

• Для любителей экзотики

Реклама Node**. Полюбуйтесь на нерекурсивные `find`, `del`, `add`:

```

1 Node** find(Node** root, int x) {
2     while (*root && (*root)->x != x)
3         root = (x < (*root)->x ? &(*root)->l : &(*root)->r);
4     return root;
5 }
6 bool isIn(Node *root, int x) {
7     return *find(&root, x) != 0;
8 }
9 void add(Node** root, int x) { // предполагает, что x точно нет в дереве
10    *find(root, x) = new Node(x);
11 }
12 void del(Node **root, int x) { // предполагает, что x точно есть в дереве
13    Node **v = find(root, x);
14    if (!(*v)->r) {
15        *v = (*v)->l; // подцепили левого сына v к её отцу
16        return;
17    }
18    Node **p = &(*v)->r;
19    while ((*p)->l)
20        p = &(*p)->l;
21    swap((*v)->x, (*p)->x);
22    *p = (*p)->r; // подцепили правого сына p к её отцу
23 }

```

13.3. AVL (Адельсон-Вельский, Ландис'1962)

Def 13.3.1. *Дерево будем называть **сбалансированным** iff глубина дерева $\mathcal{O}(\log n)$.*

Замечание 13.3.2. Тут нужно быть аккуратными. $\mathcal{O}(\log n)$ выше может быть рандомизированным или амортизированным. В таком случае мы всё равно будем называть дерево сбалансированным. В слове BST буква B = binary, а не balanced.

Def 13.3.3. *Высота вершины v – максимальное из расстояний от v до листьев в поддереве v .*

Def 13.3.4. *BST называют AVL-деревом, если оно удовлетворяет AVL-инварианту: в каждой вершине разность высот детей не более одного: $\forall v |h(v.l) - h(v.r)| \leq 1$.*

Lm 13.3.5. *Глубина AVL-дерева $\mathcal{O}(\log(n))$*

Доказательство. Обозначим S_h – минимальный размер дерева высоты h , тогда $n = S_h \geq S_{h-1} + S_{h-2} \Rightarrow S_h \geq h$ -е число Фибоначчи $\geq 1.618^h \Rightarrow h \leq \log_{1.618} n$. ■

13.3.1. Добавление в AVL-дерево

• **Краткое изложение**

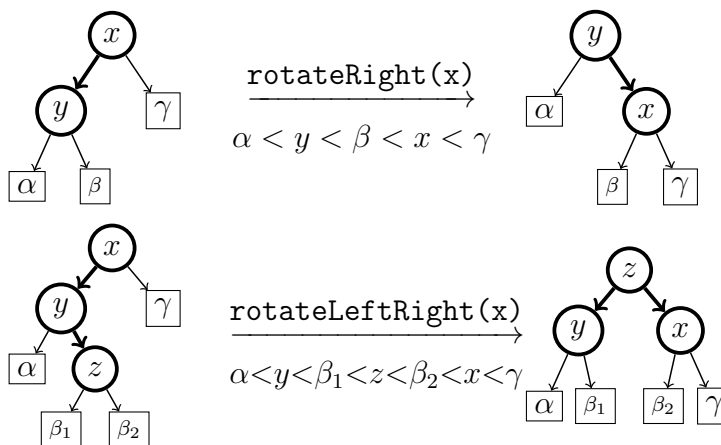
Добавление в AVL-дерево начинается также, как в обычное BST – спускаемся до упора, вставляем. Что могло пойти не так? Высоты некоторых вершин увеличились на 1, если после этого $|v.l.h - v.r.h| = 2$, то AVL-свойство нарушено. Что делать? На обратном ходу рекурсии, поднимаясь снизу вверх, если видим некошерную вершину, «повернём» дерево (см. картинки).

(a) $h(v.l.l) \geq h(v.l.r)$.

В этом случае достаточно сделать малое вращение (single rotation) по ребру $(v \rightarrow v.l)$.

(b) $h(v.l.l) = h(v.l.r) - 1$.

В этом случае нужно делать большое вращение (double rotation) по рёбрам $(v \rightarrow v.l \rightarrow v.l.r)$.



Греческими буквами обозначены поддеревья, мы ничего не предполагаем про их непустоту. Расставляя высоты поддеревьев, убеждаемся, что AVL-свойство теперь выполнено.

• Немного кода

```

1 Node* rebalance(Node* v) {
2     if (!v)
3         return v;
4
5     int hl = height(v->l), hr = height(v->r), d = hl - hr;
6     assert(abs(d) <= 2);
7     if (d == 2)
8         if (height(v->l->l) >= height(v->l->r)) // v->l существует, так как hl ≥ 2
9             v = rotateRight(v); // Малое правое вращение
10        else
11            v = rotateLeftRight(v); // Большое правое вращение
12    else if (d == -2)
13        ...
14
15    return v;
16 }
17
18 Node* add(Node* v, int x) {
19     if (!v)
20         return new Node(x);
21     if (x < v->x)
22         v->l = add(v->l, x);
23     else
24         v->r = add(v->r, x);
25     return rebalance(v);
26 }

```

Остаётся вопрос поддержки актуальных высот поддеревьев. Его мы затронем в [Section 13.6](#).

• Более подробное изложение

Описание перебалансировки в AVL-дереве можно прочитать по [ссылке](#). Там же есть большой кусок реализации AVL-деревя. Доказательство корректности происходящего там не очень аккуратное, ниже доказано более аккуратно.

• (*) Более аккуратное доказательство

(Возможно, тут написано что-то переусложнённое; если где-то найдётся что-то более простое, можно будет заменить.)

TODO рисунок (пока рисунки можно смотреть всё по той же [ссылке](#))

Rebalance.

Обозначим за h_1 высоты до вызова `rebalance`, h_2 — после.

Гарантии на вход: $v.l$ и $v.r$ являются корнями корректных AVL-деревьев, и $|h_1(v.l) - h_1(v.r)| \leq 2$.

Гарантии на выход. Пусть $u = \text{rebalance}(v)$. Тогда u является корнем корректного AVL-деревя (с теми же ключами/данными), и $h_1(v) - 1 \leq h_2(u) \leq h_1(v)$.

Доказательство гарантий на выход: не умаляя общности, рассматриваем только случаи, когда левое поддерево глубже (то есть $d = 2$) и мы делаем правое вращение — малое или большое.

Пусть $h := h_1(E)$. Тогда $h_1(b) = h + 2$ и $h_1(d) = h + 3$.

Нужно проверить корректность AVL-деревя. Кроме того, мы хотим доказать, что $h_1(v) - 1 \leq h_2(u) \leq h_1(v)$, то есть что $h + 2 \leq h_2(u) \leq h + 3$ (*).

○ Первый случай. Пусть $h_1(A) \geq h_1(C)$. Тогда $h_1(A) = h_1(b) - 1 = h + 1$, и $h \leq h_1(C) \leq h + 1$. Мы выполняем правое малое вращение.

1. Проверяем AVL-условие в вершине d : $|h_2(C) - h_2(E)| = |(h \dots h + 1) - h| \leq 1$.
2. Высота d : $h_2(d) = 1 + \max(h_2(C), h_2(E)) = 1 + \max((h \dots h + 1), h)$, то есть $h + 1 \leq h_2(d) \leq h + 2$.
3. Проверяем AVL-условие в вершине b : $|h_2(A) - h_2(d)| = |(h + 1) - (h + 1 \dots h + 2)| \leq 1$.
4. Высота b : $h_2(b) = 1 + \max(h_2(A), h_2(d)) = 1 + \max(h + 1, (h + 1 \dots h + 2))$, то есть $h + 2 \leq h_2(b) \leq h + 3$, что и требовалось в (*).

о Второй случай. Пусть $h_1(A) = h_1(c) - 1$. Тогда $h_1(c) = h_1(b) - 1 = h + 1$, и $h_1(A) = h$. Кроме того, $h - 1 \leq h_1(C') \leq h$ и $h - 1 \leq h_1(C'') \leq h$.

Мы выполняем правое большое вращение.

1. Проверяем AVL-условие в вершине b : $|h_2(A) - h_2(C')| = |h - (h - 1 \dots h)| \leq 1$.
2. Высота b : $h_2(b) = 1 + \max(h_2(A), h_2(C')) = 1 + \max(h, (h - 1 \dots h))$, то есть $h_2(b) = h + 1$.
3. Аналогично проверяем AVL-условие в вершине d .
4. Аналогично вычисляем $h_2(d) = h + 1$.
5. Проверяем AVL-условие в вершине c : $|h_2(b) - h_2(d)| = |(h + 1) - (h + 1)| = 0$.
6. Высота c : $h_2(c) = 1 + \max(h_2(b), h_2(d)) = 1 + \max(h + 1, h + 1) = h + 2$, то есть (*) выполнено.

Add.

Обозначим за h_0 высоты до вызова `add`, h_1 — до вызова `rebalance`, h_2 — после.

Гарантии на вход: v является корнем корректного AVL-дерева, и $|h_1(v.l) - h_1(v.r)| \leq 2$.

Гарантии на выход. Пусть $u = \text{add}(v, x)$. Тогда u является корнем корректного AVL-дерева (с теми же ключами/данными плюс новый ключ), и $h_0(v) \leq h_2(u) \leq h_0(v) + 1$.

Доказательство гарантий на выход. Не умаляя общности, пусть рекурсивный запуск был вызван от левого поддерева. После рекурсивного запуска левое и правое поддерево v — корректные AVL-деревья: правое не менялось, левое — по предположению индукции. Высота правого не менялась, а левого увеличилась не больше, чем на единицу, поэтому $h_0(v) \leq h_1(v) \leq h_0(v) + 1$. Теперь хотим запустить `rebalance(v)`. Нужно проверить гарантию на вход. Было: $|h_0(v.l) - h_0(v.r)| \leq 1$. Рекурсивный запуск увеличил одну из глубин не больше, чем на 1 (по предположению индукции), поэтому теперь разница высот не больше двух и мы имеем право запустить $u = \text{rebalance}(v)$.

Если в вершине v уже выполнено AVL-условие, то `rebalance(v)` ничего не сделает, и всё доказано. Если же оно перестало выполняться, это значит, что высота $v.l$ стала на два больше, чем высота $v.r$. Отметим, что в этом случае $h_1(v) = h_0(v) + 1$.

По гарантии `rebalance` на выход, u — корень корректного AVL-дерева, и, кроме того, $h_1(v) - 1 \leq h_2(u) \leq h_1(v)$. Поэтому $(h_0(v) + 1) - 1 \leq h_2(u) \leq h_0(v) + 1$, что и требовалось доказать.

• При добавлении вращений мало

Можно убедиться, что если в вершине v происходит перебалансировка, то высота поддерева вершины v и до добавления, и после добавления и вращения равна $h(\gamma) + 2 \Rightarrow$ сразу же после первого вращения операцию перебалансировки можно прервать.

Однако это верно только для операции `add` (например, после операции `del` может понадобиться много перебалансировок на пути).

Lm 13.3.6. При добавлении в AVL-дереве происходит $\mathcal{O}(1)$ присваиваний указателей.

К сожалению, высоты могут меняться у $\mathcal{O}(\log n)$ вершин.

На практике мы покажем, что тем не менее амортизированное число изменений высот $\mathcal{O}(1)$.

• Удаление

Удаление из AVL-деревя происходит также, как удаление из обычного BST.

На обратном ходу рекурсии от удалённой вершины происходит перебалансировка.

Подробнее про удаление из AVL на практике.

• Компактный код (персистентных) поворотов

Замечание 13.3.7. И в добавлении, и в удалении при подъёме вверх и перебалансировке можно пользоваться ссылками на родителя. Но удобнее всю перебалансировку делать именно на обратном ходу рекурсии.

Замечание 13.3.8. Раз отцы не нужны, дерево можно сделать персистентным (см. [Section 13.5](#)).

Получается очень простой и короткий код вращения.

Большое вращение выражается через два малых.

Недостаток: тратим $\mathcal{O}(\log n)$ памяти на каждую операцию изменения дерева.

```

1 // Node = {x, l, r}
2 Node* rotateRight(Node* v) { // персистентная версия
3     return new Node {v->l->x, v->l->l, new Node {v->x, v->l->r, v->r}};
4 }
5
6 Node* rotateLeftRight(Node* v) { // не персистентная версия
7     v->l = rotateLeft(v->l); // для персистентности здесь нужно создать новую вершину
8     return rotateRight(v);
9 }
10
11 Node* rotateLeftRight(Node* v) { // персистентная версия
12     return rotateRight(new Node {v->x, rotateLeft(v->l), v->r});
13 }

```

13.4. Split/Merge

Def 13.4.1. $split(t, x)$ делит t на $l = \{a \mid a < x\}$ и $r = \{a \mid a \geq x\}$.

Def 13.4.2. Пусть нам даны деревья l, r : все ключи в l меньше всех ключей в r , $merge(l, r)$ – BST, полученное объединением ключей l и r .

Замечание 13.4.3. $merge$ – операция, обратная $split$.

В разборе практики показано, как можно в AVL дереве сделать $split$ и $merge$ за $\mathcal{O}(\log n)$.

Замечание 13.4.4. По умолчанию $split$ и $merge$ ломают деревья, получаемые на вход.

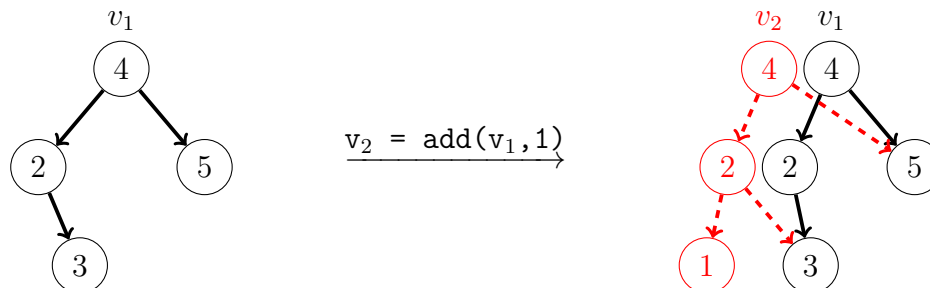
Но их, как и все другие процедуры работы с BST, можно сделать персистентными.

13.5. Персистентность

Def 13.5.1. *Персистентной* называется структура данных, которая при модификации сохраняет доступ ко всем предыдущим версиям.

У нас уже есть персистентный `add` в обычное BST [Algo 13.2.1](#). Отличие от не персистентной версии в том, что на обратном ходу рекурсии вместо того, чтобы менять старые вершины, мы создаём новые. По аналогии можно реализовать `del`.

Замечание 13.5.2. Каждая отдельная версия персистентного дерева – дерево. Все версии в совокупности образуют ациклический граф.



Замечание 13.5.3. У вершины теперь несколько “отцов”, поддерживать их не получается.

Замечание 13.5.4. Есть несколько способов сделать структура персистентной. Рассмотренный нами называется `path copying`. Также полезна `fat nodes`.

13.6. Дополнительные операции, групповые операции

Можно поддерживать в вершине v размер поддерева `v->size`.

Всегда, когда поддерево вершины v меняется, считаем `v->size = v->l->size+v->r->size+1`. Чтобы не обрабатывать отдельно случай “`v->l = ∅`”, будем как пустое поддерево использовать специальную фиктивную вершину `Node *nullNode = {size:0, l:nullNode, r:nullNode}`.

BST используется для хранения упорядоченного множества. Структура дерева задаёт на элементах порядок, поэтому у каждого элемента дерева есть позиция (номер).

Теперь, когда у каждой вершины хранится размер поддерева, мы умеем делать 4 операции:

`getValue(i)` – получать значение ключа по номеру
`getPosition(x)` – по ключу находить его номер в дереве
`add(i, data)` – вставить вершину с записью `data` на i -ю позицию в дереве
`del(i)` – удалить вершину, находящуюся на i -й позиции.

В процедурах `getValue(i)`, `add(i,data)`, `del(i)` используется спуск по дереву, в котором сравнение идёт не по ключу, а по размеру поддерева:

```

1 Node* getValue( Node* v, int i ) { // 'i - нумерация с нуля'
2     if (v->l->size == i)
3         return v;
4     if (v->l->size > i)
5         return getValue(v->l, i);
6     return getValue(v->r, i - v->l->size - 1);
7 }

```

• Операции на отрезке и групповые операции

Пусть в каждой вершине дерева кроме ключа x_v хранится некое число y_v .

Научимся для примера обрабатывать запрос “`getMin(l,r) = min{y_v | l ≤ x_v ≤ r}`”

Также, как мы поддерживали размер поддерева, поддерживаем минимальный y_v в поддереве.

Каждой вершине BST соответствует поддерево, которому соответствует отрезок значений.

Сделаем `getMin(l,r)` за $\mathcal{O}(\log n)$ спуском по дереву.

По ходу спуска, находясь в вершине v , будем поддерживать отрезок значений $[vl, vr]$.

```

1 int getMin( Node* v, int vl, int vr, int l, int r ) {
2     if (v == nullNode || r < vl || vr < l) // '[vl,vr] ∩ [l,r] = ∅'
3         return INT_MAX;
4     if (l <= vl && vr <= r) // '[vl,vr] ⊂ [l,r]'
5         return v->min_y;
6     return min(
7         getMin(v->l, vl, v->x - 1, l, r),
8         getMin(v->r, v->x + 1, vr, l, r),
9         (l <= v->x && v->x <= r) ? v->y : INT_MAX
10    );
11 }
12 int result = getMin(root, INT_MIN, INT_MAX, l, r);

```

Таким образом можно считать любую ассоциативную функцию от y_v в $\{y_v | l \leq x_v \leq r\}$.

• Модификация на отрезке

Хотим со всеми $\{y_v | l \leq x_v \leq r\}$ сделать $y_v += \Delta$ (групповая операция).

Если мы захотим честно обновить все нужные y_v , это займёт линейное время.

Чтобы получить время $\mathcal{O}(\log n)$ на запрос, воспользуемся отложенными операциями:

у каждой вершины v будем хранить `v->add` – число, которое нужно прибавить ко всем вершинам в поддереве v . Запрос “прибавления на отрезке” (`rangeInc`) теперь обрабатывается по аналогии с `getMin`, есть три случая: (1) $[vl, vr] \cap [l, r] = \emptyset$, (2) $vl, vr \subset l, r$ и (3) “отрезки пересекаются”.

Операция называется отложенной, потому что в любом запросе (`add`, `del`, `getMin`, `rangeInc`), проходя через вершину v , у которой `v->add ≠ 0`, мы проталкиваем эту операцию вниз детям:

```

1 void pushTo( Node* v, int x ) {
2     if (v != nullNode) // 'фиктивная вершина всегда должна оставаться в исходном состоянии'
3         v->add += x, v->min += x;
4 }
5 void push( Node* v ) { // 'push = проталкивание вниз'
6     pushTo(v->l, v->add);
7     pushTo(v->r, v->add);
8     v->add = 0; // 'при этом минимум не изменился'
9 }

```

13.7. Неявный ключ

Теперь наша цель – на основе BST сделать структуру данных для хранения массива, которая умеет делать вставку в середину `insert(i,x)`, удаление из середины `del(i)`, считать функцию на отрезке массива и, конечно, как и положено массиву, обращаться к ячейке `a[i]=z, z=a[i]`.

Возьмём дерево по ключу “индекс в массиве”: $x_v = i$, $y_v = a_i$.

Тогда, если в каждой вершине хранить размер поддерева и минимум y_v в поддереве, у нас уже

есть операции `getValue(i)`, `add(i, data)`, `del(i)`, `getMin(l,r)`.

Нужно только после `add(i,data)` и `del(i)` пересчитывать ключи (индексы в массиве)...

Но зачем их пересчитывать и вообще хранить, если мы ими *нигде* не пользуемся?

Идея дерева по неявному ключу: можно просто не хранить ключ.

По неявному ключу также можно делать операции `split(v,i)` и `merge(l,r)`, из них легко выразить циклический сдвиг массива = один `split` + один `merge`.

13.8. Reverse на отрезке

Если у нашего дерева есть `split` и `merge`, можно реализовать функции `getMin(l,r)`, `rangeInc(l,r)` и другие чуть проще: высплитим отрезок $[l, r)$; в его корне будет содержаться минимум на $[l, r)$, к этому корню можно за $\mathcal{O}(1)$ применить отложенную операцию; сделав с корнем $[l, r)$ всё, что хотели, смержим деревья обратно. Пример:

```

1 Node* rangeInc( Node *v, int l, int r, int value ) {
2     Node *a, *b, *c;
3     split(v, b, c, r);
4     split(b, a, b, l);
5     b->add += value; // a = [0, l), b = [l, r), c = [r, end)
6     return merge(merge(a, b), c);
7 }
```

Техника “высплитим отрезок $[l, r)$ ” позволяет простой отложенной операцией `v->isReversed` реализовать `reverse(l, r)` (перевернуть отрезок массива).

Замечание 13.8.1. `getMin`, `rangeInc` можно было делать спуском по дереву, сделать `reverse` без выспливания отрезка не получится.

13.9. Итоги

Идея отложенных операций и идея неявного ключа применимы к любым BST.

Операции `getValue(i)`, `getPosition(x)`, `add(i, data)`, `del(i)`, `getMin(l,r)`, `rangeInc(l,r)`, применимы к любому BST по явному ключу и неявному ключу.

Операция `reverse(l,r)` осмысленна только в дереве по неявному ключу, так как принципиально меняется порядок вершин дерева. В любом BST, где уже есть операции `split` и `merge`, можно получить `reverse(l,r)`.

Не в стандарте есть `tree` и `rope`. **TODO** ссылки

13.10. Персистентность: итоги

Любое дерево поиска, где не обязательно нужны отцы, можно сделать персистентным. В частности, AVL. Время работы останется пропорциональным глубине дерева, для AVL это $\mathcal{O}(\log n)$.

Следствие 13.10.1. \exists персистентное дерево по неявному ключу с операциями за $\mathcal{O}(\log n)$.

Итого у нас есть персистентный массив с обращением к i -й ячейке и кучей ништяков (вставка в середину, `getMin`, `rangeInc`, `reverse` на отрезке и т.д.), который всё это умеет за $\mathcal{O}(\log n)$.

Следствие 13.10.2. Любую структуру данных можно сделать персистентной, с замедлением в $\mathcal{O}(\log n)$ раз. Просто поменяем все массивы на персистентные массивы.

Основной минус персистентных массивов – \forall операция создаёт $\mathcal{O}(\log n)$ новых вершин дерева.

13.11. В-дерево (Bayer, McCreight'1972)

Зафиксируем число k . Вершиной дерева будем называть множество от $k - 1$ до $2k - 2$ ключей. Если в вершине хранится i ключей x_1, x_2, \dots, x_i , то детей у неё $i + 1$: T_1, T_2, \dots, T_{i+1} . При этом верно $T_1 < x_1 < T_2 < \dots < x_i < T_{i+1}$. Вспомним бинарное дерево поиска: $T_1 < x_1 < T_2$, один ключ, два ребёнка.

Def 13.11.1. В-дерево: для всех вершин кроме, возможно, корня количество ключей $[k-1, 2k-1)$, соответственно количество детей $[k, 2k)$, все листья имеют одинаковую глубину.

Лм 13.11.2. При $k \geq 2$ глубина В-дерева, хранящего n ключей, – $\Theta(\log_k n)$

При больших k встаёт вопрос: как хранить множество ключей и детей? В основном используют два варианта: в виде отсортированного массива или в виде дерева поиска.

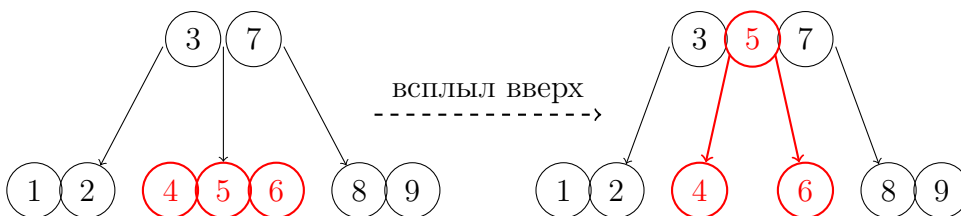
13.11.1. Поиск по В-дереву

Спуск вниз по дереву. Выбор направления, куда спускаться, происходит за $\mathcal{O}(\log k)$, – бинарный поиск или поиск в BST. Итого $\log_k n \log k = \log n$ процессорных операций.

Преимущество В-дерева над другими BST – количество операций чтения с диска. Пусть мы храним в дереве большую базу данных на $10^9 \dots 10^{10}$ записей. Тогда в оперативную память такая база не помещается, и при чтении из базы происходит чтение с жесткого диска. При чтении с жесткого диска за одну операцию чтения мы читаем не 1 байт, а сразу блок, например из 4 килобайт. В-дерево, не ухудшая процессорное время, минимизирует число обращений к диску.

13.11.2. Добавление в В-дерево

Спустимся до листьев, добавим в листе v в пачку из $[k-1..2k-2]$ ключей ещё один ключ. Если листьев в пачке стало $2k-1$, скажем “вершина переполнилась” и разделим её на две вершины по $k-1$ ключу, соединённые общим ключом-отцом (итого $(k-1) + (k-1) + 1 = 2k-1$ ключ). Этот ключ-отец вставим в соответствующее место отца v . Можно представлять эту операцию, как будто средний из ключей всплыл вверх на один уровень.



На уровень выше опять могла возникнуть проблема “слишком много ключей в вершине”, тогда рекурсивно продолжим “толкать ключи вверх”. Пример такой ситуации для $k = 2$ смотрите на картинке. При $k = 2$ в каждой вершине должно быть от 1 до 2 ключей.

При проталкивании ключа вверх важно, как мы храним $\Theta(k)$ ключей. Если в отсортированном массиве, вставка происходит за $\mathcal{O}(k \log_k n)$. Если в дереве поиска, то за $\mathcal{O}(\log k \log_k n) = \mathcal{O}(\log n)$.

13.11.3. Удаление из В-дерева

Как обычно, сперва найдём вершину, затем, если она не лист, swap-нем её с соседним по значению ключом справа. Теперь задача свелась к удалению ключа из листа. Удалим. Сломаться

могло то, что в вершине теперь меньше $k-1$ ключей. Тогда возьмём вершину-брата, и общего ключа-отца для этих двух вершин и “спустим ключ-отец вниз”. Это операция обратная изображённой на картинке выше. В результате операции мы могли получить слишком толстую вершину, тогда обратно разобьём её на две (получится, что мы просто сколько-то ключей забрали от более толстого брата себе). У вершины-отца могла уменьшиться число ключей, тогда рекурсивно пойдём вверх от отца.

Итого. Мы сперва переходим к листу. Затем удаляем ключ из листа (из вершины по середине дерева нельзя просто взять и удалить ключ). И дисбаланс ключей мы исправляем подъёмом от листа к корню дерева.

Замечание 13.11.3. В удалении и добавлении за $\mathcal{O}(\log n)$ мы пользуемся тем, что дерево поиска умеет делать split и merge за логарифмическое время.

13.11.4. Модификации

B^* -tree предлагает уплотнение до $[k, \frac{3}{2}k]$ ключей в вершине. Разобрано на [практике](#).

B^+ -tree предлагает хранить ключи только в листьях. Можно почитать в [wiki](#).

13.11.5. Split/Merge

Делаются за высоту дерева. Будут в практике, как упражнения. У деревьев должно быть одинаковое k , но могут быть произвольные высоты.

13.12. Производные B-деревя

13.12.1. 2-3-дерево (Норcroft’1970)

При $k = 2$ у каждой вершины от 2 до 3 детей. Такая конструкция называется 2-3-деревом.

13.12.2. 2-3-4-дерево и RB-дерево (Bayer’1972)

Можно ещё разрешить вершины, у которых ровно 4 сына. Такая конструкция будет называться 2-3-4-деревом. Поддерживать 2-3-4 свойство при добавлении и удалении также, как и у 2-3 дерева: слишком толстые вершины делим на две, тонкие соединяем с братьями.

2-3-4-дерево ничем не лучше 2-3-дерева. Зато оно эквивалентно красно чёрному дереву (RB-дерево). А RB-дерево это быстрая и популярная реализация сбалансированного дерева поиска. Например, используется в C++.STL внутри `set<>`.

Def 13.12.1. Рассмотрим BST, в котором каждая вершина покрашена в красный или чёрный. Также в каждое место “отсутствия сына” мысленно добавим фиктивного сына. Дерево называется красно чёрным, если:

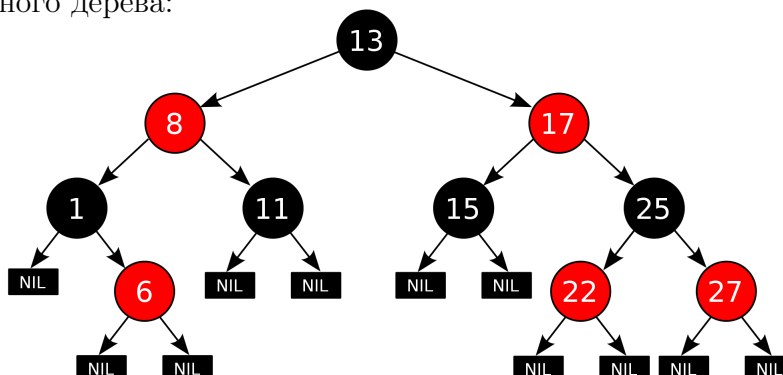
- (a) На пути от корня до любой фиктивной вершины одинаковое число чёрных вершин.
- (b) Сыном красной вершины не может быть красная вершина.
- (c) Корень – чёрная вершина.

Lm 13.12.2. Есть преобразование между красночёрными и 2-3-4-деревьями.

Доказательство. Соединим красные вершины с их чёрными отцами, получим толстую вершину, в которой от 1 до 3 ключей и от 2 до 4 детей. В обратную сторону: у вершины 2-3-4-дерева с двумя ключами создадим правого красного сына, у вершины 2-3-4-дерева с тремя ключами создадим два красных сына. ■

Замечание 13.12.3. Это почти биекция. В случае “один красный сын” при сливании его с корнем теряется информация правый он или левый.

Пример красно-чёрного дерева:



13.12.3. AA-дерево (Arne Anderson'1993)

Из 2-3-4-дерева разделением толстой вершины на “чёрную с красными детьми” можно получить RB-дерево. Но мы знаем, что можно жить без 4-вершин: если 2-3-дерево по тому же принципу преобразовать в RB-дерево, получится так называемое AA-дерево.

Def 13.12.4. *RB-дерево называется AA-деревом, если любая красная вершина является правым сыном своего отца.*

RB-дерево и AA-дерево при аккуратной реализации имеют отличную скорость работы. Также заметим, что оба дерева кроме обычных любому BST ссылок на детей хранят лишь один бит – цвет вершины. Для сравнения их между собой приведём цитату из работы Андерсона:

The performance of an AA tree is equivalent to the performance of a red-black tree. While an AA tree makes more rotations than a red-black tree, the simpler algorithms tend to be faster, and all of this balances out to result in similar performance. A red-black tree is more consistent in its performance than an AA tree, but an AA tree tends to be flatter, which results in slightly faster search times.

AA-дерево известно тем, что у него есть простая/красивая/быстрая **реализация**.

Лекция #14: Декартово дерево

11 мая 2022

14.1. Treap (Seidel, Aragon'1989)

Def 14.1.1. Декартово дерево (*cartesian tree*) от множества пар $\langle x_i, y_i \rangle$ – структура, являющаяся BST по x_i и кучей с минимумом в корне по y_i .

Lm 14.1.2. Если все y_i различны, декартово дерево единственно.

Доказательство. Корень выбирается однозначно – минимальный y_i . Левое поддерево – декартово дерево от всех $x_j < x_i$, оно единственно по индукции. Аналогично правое поддерево. ■

Def 14.1.3. Случайное дерево поиска (RBST) от множества $\{x_i\}$.

Каждый элемент с равной вероятностью $\frac{1}{n}$ может стать корнем.

Аналогичные условия выполнены рекурсивно сверху вниз во всех поддеревьях.

Def 14.1.4. Случайное дерево поиска (RBST) от множества $\{x_i\}$ – результат добавления случайной перестановки x_i в пустое BST (спустились от корня вниз до упора, вставили лист).

Lm 14.1.5. Если x_i различны, определения эквивалентны. (Напомним, что ключи можно всегда сделать различными, например, записав пары $\langle x_i, i \rangle$).

Доказательство. Корень – первый элемент перестановки, выбирается равновероятно.

Смотрим на левое поддерево. Элементы, попавшие в левое поддерево, также образуют случайную перестановку \Rightarrow далее по индукции. ■

Мы хотим понять, что RBST является в каком-то смысле сбалансированным. В идеале мы бы хотели доказать следующее утверждение:

Lm 14.1.6. Матожидание высоты RBST (то есть максимальной глубины вершин) – $\mathcal{O}(\log n)$.

Это утверждение верно, но доказывать мы его не будем (есть в допах в теордз). Вместо этого докажем более простое утверждение, которого нам хватит для оценки времени работы:

Lm 14.1.7. Пусть RBST построено по набору ключей $\{x_1, x_2, \dots, x_n\}$. Для каждого i обозначим за v_i узел, содержащий ключ x_i .

Тогда для каждого i матожидание глубины узла v_i – $\mathcal{O}(\log n)$.

Доказательство. Посмотрим на путь от корня до v_i . При каждом спуске с вероятностью хотя бы $1/3$ размер поддерева уменьшался хотя бы в полтора раза: рассмотрим отсортированное множество ключей, с вероятностью $1/3$ корень будет выбран из второй трети, тогда в поддерево, содержащее v_i , попадёт не больше $2/3$ от всех ключей.

Тогда матожидание количества спусков, при которых размер поддерева уменьшится хотя бы в полтора раза, не больше трёх. Тогда матожидание количества спусков, при которых размер поддерева станет равным единице, не больше $3 \log_{1.5} n = \mathcal{O}(\log n)$. (Возможно, размер поддерева с корнем в v_i даже больше единицы – нам только лучше.) ■

Замечание 14.1.8. Из этого, в частности, следует, что матожидание средней глубины вершин в RBST – $\mathcal{O}(\log n)$.

Def 14.1.9. Декартово дерево (*treap*) от множества ключей $\{x_i\}$ – декартово дерево (*cartesian tree*) нар $\langle x_i, \text{random} \rangle$.

Lm 14.1.10. *Treap*, как *BST*, является *RBST*

Следствие 14.1.11. Матожидание высоты *Treap* – $\mathcal{O}(\log n)$

Следствие 14.1.12. Матожидание глубины любой вершины в *Treap* – $\mathcal{O}(\log n)$

Следствие 14.1.13. Матожидание средней глубины вершины в *Treap* – $\mathcal{O}(\log n)$

14.2. Операции над *Treap*

В основе декартова дерева лежат операции *Split* и *Merge*, *Add* и *Del* выражается через них. И *Split*, и *Merge* – спуск вниз по дереву. Время обеих операций – глубина вершины, до которой мы спускаемся, т.е. $\mathcal{O}(\log n)$ (подробнее на практике).

Напомним, что *Split*(x) разделяет структуру данных на две того же типа – содержащую элементы “ $< x$ ” и содержащую элементы “ $\geq x$ ”. *Merge*(L, R) – обратная к *Split* операция. В частности *Merge* требует, чтобы все ключи в L были меньше всех ключей в R .

Lm 14.2.1. Пусть над некоторым *BST* определены *Split* и *Merge* за $\mathcal{O}(\log n)$. Тогда можно доопределить: $\text{Add}(x) = \text{Split} + \text{Merge} + \text{Merge}$, $\text{Del}(x) = \text{Split} + \text{Split} + \text{Merge}$.

Доказательство. *Add*: разделим старое дерево на меньшие и большие x элементы. Итого, включая сам x , у нас три дерева. Склеим их. *Del*: разделим старое дерево на три – меньше x , сам x , больше x . Склеим крайние. ■

Split в декартовом дереве состоит из двух случаев – в какую из половин попал корень:

```

1 // чтобы вернуть пару, удобно получить в параметры две ссылки
2 void Split(Node* t, Node* &l, Node* &r, int x) {
3     if (!t)
4         l = r = 0; // база индукции
5     else if (t->x < x)
6         Split(t->r, t->r, r, x), l = t;
7     else
8         Split(t->l, l, t->l, x), r = t;
9 }
```

Подробно объясним строку (5). Если корень $t \rightarrow x$ попал в левую половину, то элементы меньше x – это корень, всё его левое поддерево и какая-то часть правого. Поэтому вызовем рекурсивно от $t \rightarrow r$, левую из образовавшихся частей подвесить к корню, правую запишем в r .

Merge(L, R) по определению должен сделать корнем *Node*-у с минимальным y . Это или корень L , или корень R . Пусть это L . у, тогда рекурсивно дёргаем *Merge*($L.R, R$).

14.3. Сборка мусора (garbage collection)

Если наше *BST* персистентно, а нам нужны не все версии, появляются бесполезные вершины, память от которых неплохо бы очистить. Простейший способ: *shared_ptr*, или рукописная «ссылочная сборка мусора» (работает быстрее стандартного *shared_ptr*):

```

1 struct Node {
2     Node *l, *r;
```

```

3   int cnt;
4   void dec() { if (!--cnt) l->dec(), r->dec(), delete this; }
5 }

```

Более эффективный способ — не париться, пока память не закончится, как только закончилась, очистить разом всё лишнее. Тут можно поступить двумя способами:

• Списковый аллокатор

Поддерживаем список свободных Node-ов. В Node добавим бит `used`. В момент «список свободных меньше 100 (больше глубины, запас, которого хватит на 1 операция с деревом)»

1. Пометим сперва вообще все Node как `used=0`.
2. Запустим `dfs` от нужных корней.
3. Переберём все Node-ы и `used=0` добавим в список свободных.

Для организации списка удобно поле `Node* l` использовать как `next`.

• Стековый аллокатор

Пусть нам нужно оставить только последнюю версию дерева. Выделим заранее

TODO

14.4. Дополнение о персистентности

Мы умеем представлять любую структуру данных в виде множества массивов и деревьев поиска. Дерево поиска само по себе может быть персистентным. Сделать массив персистентным можно двумя способами:

1. Уже умеем. Дерево поиска по неявному ключу. Оно кроме всего прочего будет `Copy`.
2. Научимся позже. Дерево отрезков с запросами сверху.

Когда нам от массива не нужно ничего кроме присваивания “`a[i] := x`” и чтения “`x = a[i]`”, второй способ предпочтительней – его реализация проще.

14.4.1. Offline

Если все запросы к персистентной структуре известны заранее, мы можем построить дерево версий и обойти его поиском в глубину. При выходе из рекурсии нам нужно откатывать операции \Rightarrow «амортизация» убивается (пример: `persistent offline СНМ` будет работать за $\mathcal{O}(\log n)$).

14.4.2. Персистентная очередь за $\mathcal{O}(1)$

В прошлом семестре мы прошли “очередь с минимумом без амортизации”. Там использовались две идеи: (а) очередь = два стека, (б) переворачивать стек можно лениво. Из последнего домашнего задания мы знаем, что персистентный стек с операциями `push/pop/size/copy` за $\mathcal{O}(1)$ существует и представляет собой дерево. Осталось собрать все знания в алгоритм.

Пусть наша очередь = `L|R` – стек `L` для `pop` и стек `R` для `push`. В момент `L.size == R.size` начнём ленивое копирование: `L|R \rightarrow LR|`. Для этого создадим ко `L1 = L.copy()`, `R1 = R.copy()`, сделаем `R = empty`, и будем продолжать использовать `L/R` для `pop/push`, а новые нам нужно лишь для копирования. Вся процедура копирования выглядит так:

```

1 while (R1.size()) result.push(R1.pop()) // R1 попал в result в обратном порядке
2 while (L1.size()) tmp.push(L1.pop()) // переворачиваем
3 for (k = 0; k < L.size(); k++) // мы хотим не весь L1, а L, часть, не скушанную pop-ами

```

```
4 result.push(tmp.pop()) // нужная часть L1 попал в result в таком же порядке
```

Если изначально $n == L.size == R.size$, нам нужно $3n$ шагов для копирования \Rightarrow за 1 pop/push достаточно делать 3 шага, чтобы точно успеть, пока L не опустел и пока R не вырос в 2 раза.

Далее можно прочесть полную версию кода на 5 стеках.

```
1 struct Queue {
2     Stack L, R; // L для pop, R для push
3     Stack L1, R1, tmp; // вспомогательные стеки, итого 5 стеков
4     int state, copied;
5
6     Queue Copy() const {
7         return Queue(L.copy(), R.copy(), L1.copy(), R1.copy(), tmp.copy(), state, copied);
8     }
9     int Front() const {
10        return L.front(); // очередь не пуста => L не пуст!
11    }
12    pair<Queue, int> Pop() const {
13        Queue res = Copy();
14        int data = res.L.pop();
15        forn(i, 3) res.Step();
16        return make_pair(res, data);
17    }
18    Queue Push(int data) const {
19        Queue res = Copy();
20        res.R.push(data);
21        forn(i, 3) res.Step();
22        return res;
23    }
24    void Step() { // основной шаг переворачивания; этот метод не const!
25        if (state == DO_NOTHING) {
26            // Если у нас достаточно большой pop-стек, рано волноваться
27            if (L.size > R.size) return;
28            // Иначе L.size == R.size, и мы начинаем копирование
29            R1 = R, R = new Stack(), tmp = new Stack();
30            state = REVERSE_R;
31        }
32        if (state == REVERSE_R) {
33            tmp.push(R1.pop())
34            if (R1.size == 0) L1 = L.copy(), state = REVERSE_L;
35        } else if (state == REVERSE_L) {
36            R1.push(L1.pop());
37            if (L1.size == 0) copied = 0, state = REVERSE_L_AGAIN;
38        } else { // state == REVERSE_L_AGAIN
39            if (L.size > copied) copied++, tmp.push(R1.pop());
40            if (L.size == copied) L = tmp, state = DO_NOTHING;
41        }
42    }
43 };
```

У структуры Stack операции pop и push меняют стек.

Персистентность стека используется только в момент вызова метода Copy.

14.4.3. Простой персистентный дек (pairing)

PairingDeque — рекурсивная структура данных, которая шепчет нам «давайте хранить не более одного элемента слева, не более одного справа, а остальные разобьём на пары и рекурсивно сделаем с ними то же самое».

Время работы операций `push_back`, `pop_back`, `push_front`, `pop_front` — $\mathcal{O}(\log n)$.

Мы уже умели получать такое время (BST, дерево отрезков), но у данной структура проще и имеет меньшую константу. Из крутого: если мы делаем много `push_back` подряд, амортизированное время — $\mathcal{O}(1)$ на операцию. Ниже представлен код на псевдоплюсах.

```
1 struct Deque<T> {
2     T l; // элемент или null
3     Deque<pair<T,T>> m; // все остальные разбиты на парах
4     T r; // элемент или null
5
6     Deque<T> push_back(T x) { // возвращаем новую версию
7         if (r == null) return {l, m, x};
8         return Deque(l, m.push_back(pair(r, x)), null);
9     }
10    pair<T, Deque<T>> pop_back() { // возвращаем то, что достали, и новую версию
11        if (r != null) return pair(r, Deque(l, m, null));
12        if (m == null) return pair(l, Deque(null, null, null));
13        <x, m1> = m.pop_back();
14        return pair(x.second, Deque(l, m1, x.first));
15    }
16 }
```

14.4.4. Треар и избавление от Y

В структуре Треар мы пользовались игреками для того, чтобы оценить полученную структуру как RBST. Попробуем избавиться от хранения y , но всё ещё сохранить случайность дерева.

Обсудим только операции Split и Merge, раз остальные выражаются через них. Функция Split не использовала y , поэтому не поменяется. Как поменять функцию Merge?

Если мы хотим сmerzжить дерево A и дерево B , то кинем монетку: с вероятностью $\frac{|A|}{|A|+|B|}$ выберем в качестве корня корень A , с вероятностью $\frac{|B|}{|A|+|B|}$ выберем в качестве корня корень B .

Почему именно такие вероятности? Каждый из ключей в A является корнем A с вероятностью $1/|A|$, каждый из ключей B является корнем B с вероятностью $1/|B|$.

Получается, что каждый ключ объединения является корнем с вероятностью $\frac{1}{|A|+|B|}$.

По индукции, корни всех оставшихся поддеревьев тоже будут выбраны случайно. Конец.

Плюсы: сэкономили память, теперь всегда можно делать Треар персистентным.

Минусы: работает медленней, random – нетривиальное $\mathcal{O}(1)$.

Замечание 14.4.1. Какая вообще была проблема с персистентностью? Если мы делаем `root = new Node; for 20 раз root = Merge(root, root)`, мы раньше получали дерево из 2^{20} одинаковых y -ов, которые благополучно выстраивались в бамбук.

Если мы не планируем никогда Merge-ить дерево с собой или пересекающиеся части одного дерева, у нас такой проблемы не будет и можно спокойно использовать y -и.

14.5. (*) Частичная персистентность: fat nodes

Def 14.5.1. Структура данных частично персистентна, если допускает модификацию последней версии и *get*-запрос к любой.

Замечание 14.5.2. Частично персистентный массив тривиально сделать за $\mathcal{O}(1)$ памяти/времени на модификацию и $\mathcal{O}(\log n)$ времени на обращение: храним независимо для каждой ячейки вектор пар \langle время модификации, значения \rangle ; обращение = бинпоиск.

Теорема 14.5.3. Если у нас есть ссылочная структура (BST, список, двусвязный список) с ограниченной входящей степенью (во всех перечисленных структурах ≤ 2), можно получить амортизированную оценку $\mathcal{O}(1)$ времени и на модификацию, и на обращение.

Доказательство. Техника fat nodes.

Можно было бы воспользоваться тривиальным решением и в каждой вершине хранить вектор $\langle time_i, version_i \rangle$, при каждом обращении к каждой вершине пришлось бы делать бинпоиск.

Оптимизация: будем для каждой вершины хранить не более k версий. Когда у вершины приходится создавать $(k+1)$ -ую версию, создадим копию вершины из одной новой версии, всем предкам вершины рекурсивно создадим новую версию, направив исходящее ребро в копию.

```

1 # Vertex : parents, versions
2 def newVersion(t, v, newVersion):
3     if len(v.versions) < k:
4         v.versions.append((t, newVersion))
5         return
6     u = Vertex(v.parents, [(t, newVersion)])
7     for p in v.parents:
8         newVersion(t, p, redirectLink(p, v, u))

```

Осталось выбрать число k . Возьмём любое, чтобы $\forall v |v.parents| < k$. Время работы и память оцениваются одинаково. Потенциал $\varphi = \sum |v.versions|$, где сумма берётся только по вершинам, содержащим последнюю версию. Когда мы создаём новую вершину, мы уменьшаем потенциал на k и увеличиваем на $|v.parents| \Rightarrow$ уменьшаем. ■

• **Примеры:**

- Треар. Обычная операция split. Создаст $\mathcal{O}(\log n)$ новых вершин, отработает за $\mathcal{O}(\log n)$. Её fat-node-персистентный аналог создаст асимптотически столько же вершин за то же время.
- AVL по явному ключу. Операция add модифицирует ссылки у $\mathcal{O}(1)$ вершин. Высоты – промежуточная информация, её можно смело портить у старых версий. Значит n персистентных операций add в AVL используют $\mathcal{O}(n)$ памяти и $\mathcal{O}(n \log n)$ времени.
- AVL по неявному ключу. Нужно хранить размеры \Rightarrow поменять $\Theta(\log n)$ вершин.
- Дерево отрезков. Одна модификация в точке изменяет $\log n$ вершин $\Rightarrow n$ модификаций изменят $n \log n$ вершин \Rightarrow создадут $\mathcal{O}(n \log n)$ новых.

Следствие 14.5.4. Если взять scanline с деревом отрезков или, мы не получили преимущества перед обычной персистентностью. Если взять scanline с BST, то важно, какое именно BST использовать, и какие операции с ним. Пусть используем AVL с явным ключом, операции только add \Rightarrow profit.

Пример задачи: online проверка, лежит ли точка внутри невыпуклого многоугольника (обобщение: в какой грани планарного графа лежит точка). Научились её решать за $\langle \mathcal{O}(n), \mathcal{O}(\log n) \rangle$.

Лекция #15: Дерево отрезков

18 мая 2022

Дерево отрезков (range tree) – это и структура данных, и мощная идея. Как структура данных, она не даёт почти ничего асимптотически нового по сравнению с BST. Основные плюсы дерева отрезков – малые константы времени и памяти, а также простота реализации.

Поэтому наш рассказ начнётся с самой эффективной реализации дерева отрезков – “снизу”.

15.1. Общие слова

Дерево отрезков строится на массиве. Каждой вершине соответствует некоторый отрезок массива. Обычно дерево отрезков хранят, как массив вершин, устроенный, как бинарная куча:

$$\text{root} = 1, \text{parent}[v] = v / 2, \text{leftSon}[v] = 2 * v, \text{rightSon}[v] = 2 * v + 1.$$

Если вершине v соответствует отрезок $[vl, vr]$ ¹, её детям будут соответствовать отрезки $[vl, vm]$ и $(vm, vr]$, где $vm = (vl+vr)/2$ ². Листья дерева – вершины с $vl = vr$, в них хранятся элементы исходного массива. Сам массив нигде кроме листьев обычно не хранится.

В вершинах дерева отрезков хранится некая функция от отрезка массива. Простейшие функции – `min`, `sum` чисел на отрезке, но хранить можно совершенно произвольные вещи. Например, `set` различных чисел на отрезке. Единственное ограничение на функцию: зная только значения функции в двух детях, мы должны иметь возможность посчитать функцию в вершине.

Дерево отрезков используют, чтобы вычислять значение функции на отрезке.

Обычно (но не всегда!) дерево отрезков допускает модификацию в точке, на отрезке.

15.2. Дерево отрезков с операциями снизу

```

1 void build(int n, int a[]): // O(n)
2     t.resize(2 * n) // нам понадобится не более 2n ячеек
3     for (i = 0; i < n; i++)
4         t[i + n] = a[i] // листья дерева находятся в ячейках [n..2n)
5     for (i = n - 1; i > 0; i--)
6         t[i] = min(t[2 * i], t[2 * i + 1]) // давайте хранить минимум
7
8 int getMin(int l, int r) { // O(log(r-l+1))
9     int res = INT_MAX; // нейтральный элемент относительно операции
10    for (l += n, r += n; l <= r; l /= 2, r /= 2) {
11        // 1. Сперва спустимся к листьям: l += n, r += n
12        // 2. Вершины кроме, возможно, крайних, разбиваются на пары (общий отец)
13        // 3. Отрежем вершины без пары - нечётный l и чётный r, перейдём к отрезку отцов [l/2,
14            r/2]
15        if (l % 2 == 1) res = min(res, t[l++])
16        if (r % 2 == 0) res = min(res, t[r--])
17    }
18    return res;
19 }
```

Лм 15.2.1. *Время работы `get` на отрезке $[l, r]$ равно $\mathcal{O}(1 + \log(r - l + 1))$*

То есть, мало того, что `get` не рекурсивен, он ещё и на коротких отрезках работает за $\mathcal{O}(1)$.

¹ Можно писать на полуинтервалах. Я сам пишу и рекомендую именно полуинтервалы.

² Если vl и vr могут быть отрицательными или больше $\frac{1}{2} \text{INT_MAX}$, вычислять vm следует, как $vl + (vr - vl) / 2$.


```

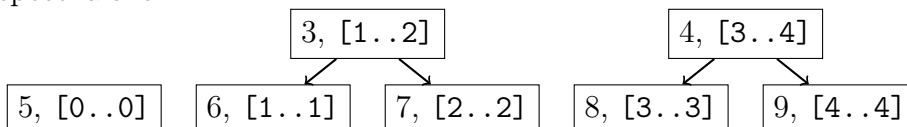
1 int change(int i, int x) {
2     t[i += n] = x; // обновили значение в листе
3     // пересчитали значения во всех ячейках на пути до корня (все отрезки, содержащие i)
4     for (i /= 2; i >= 1; i /= 2)
5         t[i] = min(t[2 * i], t[2 * i + 1]);
6 }

```

Время работы – $\mathcal{O}(\log n)$. И снова без рекурсии.

Казалось бы `get` и `change` корректно решают задачу «минимум на меняющемся массиве». В чём подвох? Пора обратить внимание на то, что «реализация дерева отрезков снизу» не является деревом, она является лесом (см. картинку). Этот лес состоит из слоёв. Для корректности нужно доказать, что слои не пересекаются.

Дерево отрезков для $n = 5$:



Lm 15.2.2. *Корректность.*

Доказательство. Нижний слой – исходный массив, он хранится в ячейках $[l_0, r_0) = [n, 2n)$. Сразу заметим $r_0 \leq 2l_i$ и докажем по индукции $r_i \leq 2l_i$. Слои отцов слоя $[l_i, r_i)$ обозначим за $[l_{i+1}, r_{i+1})$. Читая `get`, видим что $[l_{i+1}, r_{i+1}) = \left[\left\lfloor \frac{l_i}{2} \right\rfloor, \left\lfloor \frac{r_i}{2} \right\rfloor \right)$. Слои не пересекаются $\Leftrightarrow r_{i+1} \leq l_i \Leftrightarrow \left\lfloor \frac{r_i}{2} \right\rfloor \leq l_i \Rightarrow r_i \leq 2l_i$. Вторая часть: $r_{i+1} \leq 2l_{i+1} \Leftrightarrow \left\lfloor \frac{r_i}{2} \right\rfloor \leq 2 \left\lfloor \frac{l_i}{2} \right\rfloor \Rightarrow r_i \leq 2l_i$. ■

Функция `change` будет иногда ходить по бесполезным ячейкам. Это ничему не мешает: главное, что все ячейки с полезной информацией `change` корректно пересчитал.

15.3. Дерево отрезков с операциями сверху

Дерево отрезков с операциями сверху – гораздо более естественная структура.

Корень – отрезок $[0..n)$, далее дерево строится рекурсивно по определению (Section 15.1).

Все функции работы с деревом, включая построение, – рекурсивные функции спуска. Пример:

```

1 int getMin(int v, int vl, int vr, int l, int r) {
2     if (vr < l || r < vl) return INT_MAX; // не пересекаются
3     if (l <= vl && vr <= r) return t[v]; // вершина целиком внутри запроса
4     int vm = (vl + vr) / 2;
5     int fl = getMin(2 * v, vl, vm, l, r);
6     int fr = getMin(2 * v + 1, vm + 1, vr, l, r);
7     return min(fl, fr);
8 }
9 int result = getMin(1, 0, n - 1, l, r); // 1 = root

```

Здесь показана версия, в которой отрезок вершины $[vl, vr]$ не хранится, как свойство вершины, а вычисляется по ходу спуска сверху вниз. Версия с хранением ни чем не лучше – время пересчёта vl, vr сопоставимо с временем обращения к памяти для чтения уже посчитанной величине.

Операцию `change` можно сделать рекурсивно сверху.

А можно округлить n вверх до 2^k и сделать `get` сверху, `change` снизу. Правда преимуществ нет...

Чтобы пользоваться деревом “сверху” осталось его построить, а в построении – главное выделить массив нужной длины. При $n = 2^k$ получается полное бинарное дерево из $2^{k+1} = 2n$ вершин. Если не округлять n до 2^k , все индексы лишь уменьшатся.

Lm 15.3.1. $4n$ ячеек достаточно, более точно выделить $2^{1+\lceil \log_2 n \rceil}$ ячеек.

Теперь оценим время работы запроса `getMin`.

Lm 15.3.2. `getMin` посетит не более $4 \log n$ вершин дерева.

Доказательство. Уровней в дереве $\log n$. На каждом уровне мы посетим не более четырёх вершин, потому что только в двух вершинах предыдущего уровня мы ушли в рекурсию – в вершинах, которые содержали края отрезка $[l, r]$. ■

Lm 15.3.3. `getMin` разбивает любой отрезок $[l, r]$ на не более чем $2 \log n$ вершин дерева отрезков.

Доказательство. Уровней в дереве $\log n$, на каждом уровне мы выберем не более 2 вершин. ■

Последняя лемма верна и для реализации “снизу”. По тем же причинам.

Решая задачи, часто удобно думать про дерево отрезков так:

Мы даём дереву отрезков $[l, r]$, а оно разбивает отрезок $[l, r]$ на $\leq 2 \log n$ вершин, для которых уже посчитана полезная функция.

Минусы по сравнению с реализацией “снизу”:

- (a) Памяти нужно $4n$ вместо $2n$.
- (b) `get` почти всегда за $\mathcal{O}(\log n)$, даже для отрезков длины $\mathcal{O}(1)$.
- (c) Из-за рекурсии больше константа.

Зато есть много плюсов, главный из них: также, как и в BST, можно делать “модификацию на отрезке”. “Все элементы на отрезке $[l, r]$ увеличить на x ”. “Всем элементам присвоить x ”.

Модификация на отрезке делается *отложенными операциями*. Если в вершине v хранится отложенная операция, проходя через v сверху вниз, важно не забыть эту операцию протолкнуть вниз. Проталкивание вниз назовём `push`. Пример функции присваивания на отрезке:

```

1 void push(int v) {
2     if (value[v] == -1) return; // нет отложенной операции
3     value[2 * v] = value[2 * v + 1] = value[v];
4     value[v] = -1;
5 }
6 void setValue(int v, int vl, int vr, int l, int r, int x) {
7     if (vr < l || r < vl) return INT_MAX; // не пересекаются
8     if (l <= vl && vr <= r) { // вершина целиком внутри запроса
9         value[v] = t[v] = x; // не забываем пересчитать минимум в вершине
10        return;
11    }
12    push(v);
13    int vm = (vl + vr) / 2;
14    setValue(2 * v, vl, vm, l, r, x);
15    setValue(2 * v + 1, vm + 1, vr, l, r, x);
16    // 1. Сейчас в нашей вершине отложенной операции нет, мы её толкнули вниз
17    // 2. При написании кода важно заранее решить t[v] - минимум с учётом value[v] или без
18    t[v] = min(t[2 * v], t[2 * v + 1]); // у нас с учётом, т.к. под min ставится t[2 * v]
19 }

```

Ещё некоторые плюсы реализации сверху:

- (a) Дерево отрезков сверху – реально дерево!
- (b) Дерево отрезков сверху можно сделать динамическим (следующий раздел).
- (c) Дерево отрезков сверху можно сделать персистентным (следующий раздел).

15.4. (*) Хаки для памяти к дереву отрезков сверху

• Проблема

Если $n = 2^k$ – проблем нет. Памяти $2n$.

Проблема, это, например, $n = 2^k + 2$, тогда дети корня имеют размер $2^{k-1} + 1 \Rightarrow$ глубину $k \Rightarrow$ правый из детей будет использовать ячейку $2^{k+1} + 2^k \approx 3n$.

Бывает ещё хуже. Пусть $n = 2^k + 4 \Rightarrow$ дети $2^{k-1} + 2 \Rightarrow$ внуки $2^{k-2} + 1 \Rightarrow$ нужна ячейка для правого внука $2^{k+1} + 2^k + 2^{k-1} \approx 3.5n$.

Lm 15.4.1. Обычная реализация сверху при выборе $vl = (vl + vr) / 2$ и с округлением вверх, и с округлением вниз может привести к использованию ячеек с номером $4n - \Theta(\sqrt{n})$.

Доказательство. Рассмотрим $n = 2^k + 2^i \Rightarrow$ на нижнем уровне используем ячейку $2^{k+1} + 2^k + \dots + 2^{k-i+1} = 2^{k+2} - 2^{k-i+1}$. Возьмём $i = k/2$, получим $n \rightarrow 4n - \Theta(\sqrt{n})$. ■

• Решение

Есть способ чуть сэкономить память при реализации дерева отрезков сверху: делить отрезок длины w не пополам, а на максимальное $2^k < w$ и $w - 2^k$.

Lm 15.4.2. Метод деления $w \rightarrow \max(2^k < w \text{ и } w - 2^k)$ гарантирует $v < 3n$.

Доказательство. Дерево отрезков можно разбить на уровни: полное бинарное дерево из $n = 2^k$ вершин состоит из уровней $0, 1, \dots, k$. При $2^k \leq n < 2^{k+1}$ важно, какие вершины будут созданы на $(k+1)$ -м уровне. При $2^k \leq n \leq 2^k + 2^{k-1}$ на нижнем уровне будут только вершины левого поддерева \Rightarrow их номера лежат в $[0, 2^k)$. Иначе $2^k + 2^{k-1} < n < 2^{k+1} \Rightarrow v < 2^{k+2} \leq 3n$. ■

15.5. Динамическое дерево отрезков и сжатие координат

Пусть наш массив длины $M = 10^{18}$ и изначально заполнен нулями.

Есть два способа реализовать на таком массиве дерево отрезков.

Первый способ элегантный, не содержит лишнего кода, но резко увеличивает константу времени работы: давайте все массивы заменим на `unordered_map`-ы. При этом `t[]` заменим на `unordered_map<int, T>`, где `T` – специальный тип, у которого конструктор создаёт нейтральный относительно нашей операции элемент.

Замечание 15.5.1. Первый способ работает и для версии снизу, и для версии сверху.

Второй способ предлагает от хранения в массиве перейти к ссылочной структуре:

```
1 struct Node {
2     Node *l, *r;
3     int min;
4 };
```

Вершина дерева такая же, как и в BST. Вершины можно создавать лениво.

Изначально всё дерево отрезков состоит из `Node* root = NULL`.

Все запросы спускаются сверху вниз и, если попадают в `NULL`, создают на его месте `Node`.

Lm 15.5.2. Время работы любого запроса $\mathcal{O}(\log M)$.

Lm 15.5.3. После k запросов создано $\mathcal{O}(\min(M, k \log M))$ `Node`-ов.

Lm 15.5.4. После k запросов создано $\mathcal{O}(k \cdot \max(1, \log M - \log k))$ `Node`-ов.

Динамические деревья отрезков незаменимы при решении таких задач, как “ k -я статистика на отрезке за $\mathcal{O}(\log n)$ ”. Также их удобно применять в качестве внешнего дерева в 2D-деревьях.

• Сжатие координат

Динамическое дерево отрезков засчёт большей глубины и ссылочной структуры медленней обычного. Использует больше памяти \Rightarrow реже кэшируется. Поэтому, когда есть возможность, для решения исходной задачи применяют не его, а *сжатие координат*.

Задача: в offline обработать n запросов “изменение в точке i_j ” и “сумма на отрезке $[l_j, r_j]$ ” над массивом длины M .

Сжатие координат.

Offline \Rightarrow все запросы известны заранее, на $[0, M)$ не более $3n$ интересных точек: i_j, l_j, r_j .

Давайте, сложим их в массив, отсортируем и заменим на позиции в этом массиве.

То есть, перенумеруем числами из $[0, 3n)$. Свели исходную к задаче на массиве длины $3n$.

15.6. Персистентное дерево отрезков, сравнение с BST

Также, как динамическое дерево отрезков, пишется на указателях.

Также, как в персистентном BST, все запросы сверху. Также, как в персистентном BST, Node-ы нельзя менять, при желании менять всегда придётся создавать новую вершину.

Все операции, которые умело дерево отрезков, всё ещё работают. За то же время.

• Сравнение с персистентным BST

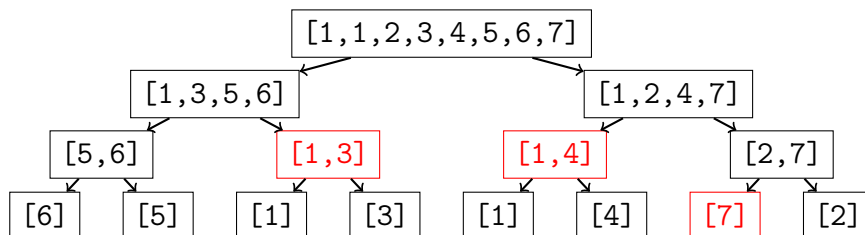
Персистентное BST по неявному ключу, конечно, умеет больше:

`insert`, `delete`, `split/merge/reverse`.

Если же всё, что мы хотим, – реализовать простейший персистентный массив ($a_i = x, x = a_i$), то у BST всё равно есть преимущество – оно хранит ровно n Node-ов, а дереву отрезков нужно от $2 \log n$ до $4 \log n$. Чтобы время работы у BST было не хуже, если мы не собираемся менять его структуру (нет `insert` и прочего), создадим его сразу идеально сбалансированным.

15.7. 2D-запросы, дерево сортированных массивов

Для начала попробуем сохранить в вершине дерева отрезков сортированную копию отрезка и посмотрим, что получится. На картинке дерево отрезков для массива $[6, 5, 1, 3, 1, 4, 7, 2]$.



Как мы помним (Lm 15.3.3), дерево отрезков разбивает любой отрезок на $\mathcal{O}(\log n)$ вершин дерева отрезков. Красным выделены вершины, на которые распадётся отрезок $[2, 7]$.

Задача #1: дан массив длины n , отвечать на запросы `get(L, R, D, U)`: число элементов на отрезке $[L, R]$, значения которых от D до U , то есть, $\#\{i: L \leq i \leq R \wedge D \leq a_i \leq U\}$.

Задача #2: даны n точек (x_i, y_i) на плоскости, отвечать на запросы “число точек в прямоуголь-

нике”, то есть, $\#\{i: X_1 \leq x_i \leq X_2 \wedge Y_1 \leq y_i \leq Y_2\}$.

Мы описали так называемые 2D-запросы на массиве и на плоскости.

Теорема 15.7.1. *Описанные выше задачи равносильны.*

Доказательство. Если есть массив a_i , можно обозначить $(x_i, y_i) = (i, a_i)$. В другую сторону: отсортируем точки по x_i , теперь двумя бинпоисками условие вида $X_1 \leq x_i \leq X_2$ можно превратить в равносильное $L \leq i \leq R$. ■

Обе задачи решаются деревом отрезков сортированных массивов за $\mathcal{O}(\log^2 n)$ на запрос³.

Решим задачу #1, вторая к ней сводится. Отрезок $[L, R]$ разделится на $\mathcal{O}(\log n)$ вершин, на каждой из них сделаем два бинпоиска, вернём \sum по вершинам “upper_bound(U) - lower_bound(D)”.

Время построения дерева отрезков сортированных массивов – $\mathcal{O}(n \log n)$, так как каждая вершина получается, как merge своих детей, который считается за линейное время.

• **k -я статистика на отрезке за $\mathcal{O}(\log^3 n)$.**

Запрос `get(L, R, k)` – вернуть `sorted(a[L..R])[k]`.

Сделаем бинпоиск по ответу. Внутри бинпоиска нам дают x и нужно узнать сколько чисел на $[L, R]$ не более x , что мы только что научились за $\mathcal{O}(\log^2 n)$. Заметим, что бинпоиск по ответу можно реализовать, как бинпоиск по `sorted(a)`, поэтому бинпоиск сделает $\mathcal{O}(\log n)$ итераций.

15.8. Многомерные деревья

• **Д.О. of Д.О.**

Задача. Даны x_i, y_i , запрос: на отрезке $[l, r]$ среди всех $i: \min_x \leq x_i \leq \max_x$ найти $\min y$.

Решение. В дереве отрезков сортированных массивов в каждой вершине на сортированном по x_i массиве пар $\langle x_i, y_i \rangle$ построим дерево отрезков с минимумом по y_i . Ответ на запрос: обход внешнего дерева и в $\mathcal{O}(\log n)$ вершинах, делящих $[l, r]$, сперва бинпоиск, затем запрос к внутреннему дереву отрезков (итого $\mathcal{O}(\log^2 n)$ на запрос). Память и время построения $\mathcal{O}(n \log n)$.

• **Д.О. of treap**

Дерево отрезков сортированных массивов — структура данных на статичном (не меняющемся) массиве. Если добавить запросы изменения массива, “`a[i] = x`”, то нам в каждой вершине дерева отрезков нужен динамический аналог сортированного массива. Например, treap.

Получили «дерево отрезков декартовых деревьев» и $\mathcal{O}(\log^2 n)$ на запрос.

• **3D-запросы**

В вершине дерева отрезков можно хранить вообще всё, что душе угодно. Например, 2D-дерево. Рассмотрим 3D-запрос на плоскости: даны n точек (x_i, y_i, z_i) , нужно отвечать на запросы $\#\{i: X_1 \leq x_i \leq X_2 \wedge Y_1 \leq y_i \leq Y_2 \wedge Z_1 \leq z_i \leq Z_2\}$. Будем решать задачу также, как уже решили задачу про 2D-запросы: отсортируем точки по x_i , на полученном массиве построим дерево отрезков. Пусть вершине v дерева отрезков соответствует отрезок $[v_l, v_r]$. Чтобы ответить на исходный запрос, в вершине v нашего дерева отрезков достаточно иметь структуру данных, которая умеет отвечать на 2D-запросы для множества 2D-точек (y_i, z_i) из отрезка $[v_l, v_r]$. Итого $\mathcal{O}(\log^3 n)$ на запрос и $\mathcal{O}(n \log^2 n)$ памяти.

³В Section 15.12 мы оптимизируем время до $\mathcal{O}(\log n)$

Аналогично можно на k -мерный запрос отвечать за $\mathcal{O}(\log^k n)$. Памяти будет использовано $\mathcal{O}(n \log^{k-1} n)$. При $k > 3$ это не эффективно.

• Динамичный аналог

Пусть исходные координаты точек — целые от 0 до $C \approx 10^9$. Сейчас для решения задачи нам нужно постоянно сортировать массивы, жать координаты. Есть более простой, но дорогой по памяти подход: все деревья отрезков на всех уровнях делать динамическими.

Плюсы: можно легко добавлять новые точки, проще пишется, не нужны бинпоиски.

Минусы: деревья на указателях, $\mathcal{O}(n \log^k C)$ вместо $\mathcal{O}(n \log^{k-1} n)$ памяти.

15.9. Сканирующая прямая

Идея сканирующей прямой (scanline, sweep line, заметающая прямая) пришла из вычислительной геометрии и в самом общем виде звучит так: пусть на плоскости есть какие-то объекты, нарисуем вертикальную прямую и будем двигать её слева направо, обрабатывая по ходу движения события вида “объект начался”, “объект закончился” и иногда “объект изменился”.

Нам уже встречалась одномерная версия той же идеи: на прямой даны n точек и m отрезков, для каждого отрезка нужно узнать число точек внутри, для каждой точки узнать, сколько отрезками она покрыта. Решение: идём слева направо, обрабатываем события “отрезок начался”, “отрезок закончился”, “точка”.

2D случай. Даны n точек и m прямоугольников со сторонами параллельными осям координат.

Задача #1: для каждой точки посчитать, сколько прямоугольниками она покрыта.

Решение: идём слева направо, встречаем и обрабатываем следующие события

- Прямоугольник начался: сделаем `count[y1..y2] += 1`;
- Прямоугольник закончился: сделаем `count[y1..y2] -= 1`;
- Встретили точку, тогда в `count[y]` хранится ровно число открытых ещё незакрытых прямоугольников, её покрывающих.

Дерево отрезков на массиве `count` справится с обеими операциями за $\mathcal{O}(\log n)$.

Итого время работы = сортировка + scanline = $\mathcal{O}((n + m) \log(n + m))$.

Задача #2: для каждого прямоугольника посчитать число точек внутри.

Сразу заметим, что прямоугольник можно разбить на два горизонтальных стакана:

количество точек в области $\{(x, y) : X_1 \leq x \leq X_2 \wedge Y_1 \leq y \leq Y_2\}$ равно разности количеств в областях $\{(x, y) : x \leq X_2 \wedge Y_1 \leq y \leq Y_2\}$ и $\{(x, y) : x \leq X_1 - 1 \wedge Y_1 \leq y \leq Y_2\}$.

Итого осталось решить задачу для n точек и $2m$ горизонтальных стаканов.

Решение: идём слева направо, встречаем и обрабатываем следующие события:

- Встретили точку, сделаем `count[y] += 1`;
- Встретили конец стакана, посчитали $\sum_{y \in [y_1..y_2]} \text{count}[y]$.

Дерево отрезков на массиве `count` справится с обеими операциями за $\mathcal{O}(\log n)$.

• Решение online версии.

Пусть теперь заранее известны только точки и в online приходят запросы-прямоугольники, для прямоугольника нужно посчитать число точек внутри. Возьмём решение задачи #2, дадим ему n точек и 0 стаканов. Теперь по ходу scanline-а мы хотим сохранить все промежуточные состояния дерева отрезков. Для этого достаточно сделать его персистентным. Асимптотика

времени работы не изменилась (константа, конечно, хуже). Памяти теперь нужно $\mathcal{O}(n \log n)$.

Пусть `root[i]` – версия дерева до обработки события с координатой `x[i]`, тогда запрос `get(x1, x2, y1, y2)` обрабатываем так:

```
1 return root[upper_bound(x, x + n, x2) - x].get(y1, y2) -
2    root[lower_bound(x, x + n, x1) - x].get(y1, y2);
```

Итого: используя предподсчёт за $\mathcal{O}(n \log n)$, мы умеем за $\mathcal{O}(\log n)$ отвечать на 2D-запрос на плоскости. Из [Thm 15.7.1](#) мы также умеем за $\mathcal{O}(\log n)$ обрабатывать 2D-запрос на массиве.

Важно запомнить, что любой “scanline с деревом отрезков” для решения offline задачи можно приспособить для решения online задачи, сделав дерево отрезков персистентным.

15.10. k -я порядковая статистика на отрезке

Мы уже умели бинпоиском по ответу искать k -ю статистику за $\mathcal{O}(\log^3 n)$.

Поскольку отвечаем на 2D-запросы мы теперь за $\mathcal{O}(\log n)$, это же решение работает за $\mathcal{O}(\log^2 n)$. Перед тем, как улучшить $\mathcal{O}(\log^2 n)$ до $\mathcal{O}(\log n)$ решим вспомогательную задачу:

- **Бинпоиск → спуск по дереву.**

Задача: дан массив из нулей и единиц, нужно обрабатывать запросы “`a[i]=x`” и “ k -я единица”.

Решение за $\mathcal{O}(\log^2 n)$.

На данном нам массиве будем поддерживать дерево отрезков с операцией сумма. Чтобы найти k -ю единицу, сделаем бинпоиск по ответу, внутри нужно найти число единиц на префиксе $[0, x)$. Это запрос к дереву отрезков. Например, спуск сверху вниз.

Решение за $\mathcal{O}(\log n)$.

Спускаемся по дереву отрезков: если слева сумма хотя бы k , идём налево, иначе направо.

Мораль.

Внутри бинпоиска есть спуск по дереву \Rightarrow скорее всего, от бинпоиска легко избавиться.

- **k -я статистика на отрезке за $\mathcal{O}(\log n)$.**

Сейчас у нас есть следующее решение за $\mathcal{O}(\log^2 n)$: возьмём точки (i, a_i) , сделаем scanline с персистентным деревом отрезков. Теперь для ответа на запрос `get(l, r, k)`, делаем бинпоиск по ответу, внутри считаем `tree[r+1].get(x) - tree[l].get(x)`, где `tree[i].get(x)` обращается к i -й версии дерева отрезков и возвращает количество чисел не больше x на префиксе $[0, i)$.

Вместо бинпоиска по ответу будем параллельно спускаться по деревьям `tree[r+1]` и `tree[l]`. Пусть мы сейчас стоим в вершинах `a` и `b`. Общим вершинам соответствует отрезок `[vl..vr]`, если $(a \rightarrow l \rightarrow \text{sum} - b \rightarrow l \rightarrow \text{sum}) \geq k$, есть хотя бы k чисел со значениями `[vl..vm]` и мы в обоих деревьях спустимся налево, иначе мы теперь хотим найти $(k - (a \rightarrow l \rightarrow \text{sum} - b \rightarrow l \rightarrow \text{sum}))$ -е число и в обоих деревьях спустимся направо.

15.11. (*) Фенвик**[e-maxx]**

Структура данных, умеющая делать изменения в точке и считать функцию на префиксе. Если функция обратима (плюс \rightarrow минус), можно считать и на отрезке.

```

1 struct Fenwick {
2     int n;
3     vector<int> a;
4
5     Fenwick(int n) : n(n), a(n) { }
6
7     // O(logn)
8     int get(int i) { // get = [0...i]
9         int res = 0;
10        // i&(i+1)-1 = 010101010100111111
11        // i         = 010101010101011111
12        // i+1       = 010101010101100000
13        // i&(i+1)   = 0101010101000000
14        // i&(i+1)-1 = 0101010100111111
15        for (; i >= 0; i &= i + 1, i--) // [x=i&(i+1)...i]
16            res += a[i];
17        return res;
18    }
19
20    // O(logn)
21    void change(int i, int d) { // += d
22        // i|(i+1)   = 010101010101111111
23        // i         = 010101010101011111
24        // i+1       = 010101010101100000
25        // i|(i+1)   = 010101010101111111
26        for (; i < n; i |= i + 1)
27            a[i] += d;
28
29        // a[k]
30        // k         = 010101010101011111
31        // i         = 010101010101011111
32        // k&(k+1)   = 0101010101000000
33    }
34 }

```


15.12. (*) Fractional Cascading

15.12.1. (*) Для дерева отрезков

У нас есть дерево отрезков сортированных массивов. Отвечать на запрос можно сверху или снизу. На скорость это не влияет, т.к. и там, и там мы делаем одни и те же $\mathcal{O}(\log n)$ бинпоисков.

Хотим отвечать на запрос $\#\{i: L \leq i \leq R \wedge a[i] < x\}$.

Будем отвечать на запрос сверху, но бинпоиск сделаем только в корне $root = 1$.

Вместо того, чтобы делать бинпоиски в детях, воспользуемся ответом для отца.

Для этого нам понадобится предподсчёт, который получается лёгкой модификацией `merge`:

```

1 void build( int n, int *a ) {
2     assert(n & (n - 1) == 0) // n = 2^k
3     vector<vector<int>> t(2 * n), l(2 * n);
4     for (int i = 0; i < n; i++)
5         t[n + i] = vector<int>(1, a[i]); // листья дерева
6     for (int i = n - 1; i > 0; i--) {
7         int A = t[2 * i].size(), B = t[2 * i + 1].size();
8         t[i].resize(A + B), l[i].resize(A + B);
9         for (size_t a = 0, b = 0, c = 0; a < L || b < R; ) { // собственно merge
10            l[i][c] = a; // среди первых c элементов t[i] ровно a ушли в левого сына
11            if (a == L || (b < R && t[2 * i][a] > t[2 * i + 1][b]))
12                t[i][c++] = t[2 * i + 1][b++];
13            else
14                t[i][c++] = t[2 * i][a++];
15        }
16    }
17 }
```

Если бинпоиск в вершине i вернул x , в детях нам вернут $l[i][x]$ и $t[i].size() - l[i][x]$.

Следствие 15.12.1. $\forall k \geq 2$ мы улучшили время k -мерного ортогонального запроса с $\log^k n$ до $\log^{k-1} n$, добавив во внутреннее $2D$ дерево fractional cascading.

На самом деле fractional cascading – более общая идея: за $\mathcal{O}(k + \log n)$ сделать бинпоиск сразу по k сортированным массивам. Мы сейчас решили частный случай этой задачи для $\log n$ массивов специального вида. [wiki](#) даёт общее описание и ссылки. Общая идея, как в skip list – половину элементов “нижнего списка” толкать вверх в следующий список.

15.12.2. (*) Для k массивов

Задача: даны k отсортированных массивов длины $\leq n$, поступает запрос x , сделать на каждом $lowerbound(x)$.

Тривиальное решение: на каждом бинпоиск итого $\mathcal{O}(k \log n)$.

Техникой fractional cascading мы получим $\mathcal{O}(k + \log(nk))$: обозначим массивы a_1, a_2, \dots, a_k , каждый второй элемент a_1 добавим в a_2 , после этого каждый второй элемент a_2 добавим в a_3 и так далее. Суммарное удлинение других массивов, которое вызовет a_1 равно $\frac{1}{2}|a_1| + \frac{1}{4}|a_1| + \frac{1}{8}|a_1| + \dots \leq |a_1|$ и так для всех a_i . Построение структуры – k раз вызвать `merge` \Rightarrow время работы $\mathcal{O}(\sum |a_i|)$. По ходу каждого `merge` запомним для каждой пары $\langle i, j \rangle$, сколько элементов

из $a_i[0:j)$ приехали из исходного a_i , а сколько добавлено из других массивов a_m .

Как ответить на запрос? Сделаем один бинпоиск, чтобы получить $i = a_k.lowerbound(x)$.

15.13. (*) КД-дерево

Даны n точек на плоскости, каждая задаётся тройкой (x_i, y_i, w_i) – координатами и весом. Построим за $\mathcal{O}(n \log n)$ структуру данных, которая будет занимать $\mathcal{O}(n)$ памяти и будет уметь для точек в прямоугольной области $(lx_j \leq x_i \leq rx_j, ly_j \leq y_i \leq ry_j)$ за $\mathcal{O}(\sqrt{n})$ делать все те же операции, что дерево отрезков на отрезке. Добавлять новые точки можно будет корневой по запросам. Удалять, если нет необратимых запросов типа «min на прямоугольнике», также.

Структура: бинарное дерево. В корне прямоугольник $[-\infty, +\infty] \times [-\infty, +\infty]$. На чётном уровне делим вертикальной прямой n точек на $\lfloor \frac{n}{2} \rfloor$ и $\lceil \frac{n}{2} \rceil$, на нечётном делим горизонтальной прямой.

Построение за $\mathcal{O}(n \log n)$. Отсортировали и по x , и по y , передали в рекурсию два порядка – и по x , и по y . Время работы кроме исходной сортировки $T(n) = \mathcal{O}(n) + 2T(n/2)$.

Запрос за $\mathcal{O}(\sqrt{n})$. Как и в дереве отрезков 3 варианта – прямоугольник запрос и прямоугольник вершина-дерева могут не пересекаться, вкладываться, нетривиально пересекаться. Время работы $2^{\frac{1}{2} \log n}$, так как на каждом втором уровне рекурсии мы не будем ветвиться.

Все точки в прямоугольнике за $\mathcal{O}(k + \log n)$. k – размер ответа. Давайте в каждой вершине КД-дерева в процессе построения сохраним список всех точек, в порядке и по x , и по y \Rightarrow если прямоугольник-запрос пересекает только одну сторону КД-вершины, мы можем сразу вывести ответ за $\mathcal{O}(k)$. Сколько будет КД-вершин, которые пересекает запрос иначе? На каждом уровне не более 4 (четыре угла запроса) \Rightarrow всего $\mathcal{O}(\log n)$.

Лекция #16: LCA & RMQ

25 мая – 1 июня 2022

Если речь идёт о структуре данных, у которой есть функция построения (предподсчёт) и умение online отвечать на запросы, обозначение $\langle f(n), g(n) \rangle$ означает, что предподсчёт работает за время $\mathcal{O}(f(n))$, а ответ на запрос за $\mathcal{O}(g(n))$.

16.1. RMQ & Sparse table

Def 16.1.1. *RMQ = Range Minimum Query = запросы минимумов на отрезке.*

Задачу RMQ можно решать на не меняющемся массиве (static) и на массиве, поддерживающем изменения в точке (dynamic). Запросы минимума на отрезке будем обозначать “get”, изменение в точке – “change”. Построение структуры – “build”.

Мы уже умеем решать задачу RMQ несколькими способами:

1. Дерево отрезков: build за $\mathcal{O}(n)$, get за $\mathcal{O}(\log n)$, change за $\mathcal{O}(\log n)$
2. Центроидная декомпозиция: build за $\mathcal{O}(n \log n)$, get за $\mathcal{O}(LCA)$, static.
3. Корневая (пока не умеем): build за $\mathcal{O}(n)$, get за $\mathcal{O}(\sqrt{n})$, change за $\mathcal{O}(1)$.
4. Корневая (пока не умеем): build за $\mathcal{O}(n)$, get за $\mathcal{O}(1)$, change за $\mathcal{O}(\sqrt{n})$.

Lm 16.1.2. \nexists структуры данных, поддерживающей build за $\mathcal{O}(n)$, change и get за $o(\log n)$.

Доказательство. Построим структуру от пар $\langle a_i, i \rangle$, чтобы вместе с минимумом получать и его позицию. После этого n раз достанем минимум, и на его место в массиве запишем $+\infty$. Получили сортировку за $o(n \log n)$. Противоречие. ■

А вот static версию задачи (только get-запросы) мы скоро решим за время $\langle n, 1 \rangle$.

• Sparse Table

Пусть $f[k, i]$ – минимум на отрезке $[i, i+2^k)$. Массив $f[]$ можно предподсчитать за $n \log n$: $f[0]$ – исходный массив, $f[k, i] = \min(f[k-1, i], f[k-1, i+2^{k-1}])$.

Теперь, get на $[l, r) = \min(f[k, l], f[k, r-2^k])$, где $2^k \leq r-l < 2^{k+1}$.

Чтобы за $\mathcal{O}(1)$ найти такое k , используем предподсчёт “log”, теперь $k = \log[r-l]$.

Итого получили решение static RMQ за $\langle n \log n, 1 \rangle$.

• Sparse Table++

$$| \underbrace{7 \ 5 \ 2 \ 4}_{b_1=2} | \underbrace{6 \ 1 \ 8 \ 4}_{b_2=1} | \underbrace{3 \ 7 \ 8 \ 3}_{b_3=3} | \underbrace{10 \ 5 \ 10 \ 6}_{b_4=5} |$$

Разобьём исходный массив a на куски длины $\log n$, минимум на i -м куске обозначим b_i .

\forall отрезок a , который содержит границу двух кусков, разбивается на отрезок b и два “хвоста”. Минимум на хвосте – это минимум на префиксе или суффиксе одного куска, все такие частичные минимумы предподсчитаем за $\mathcal{O}(n)$. Чтобы, находить минимум на отрезке массива b , построим на b Sparse Table, который весит $\frac{n}{\log n} \log \frac{n}{\log n} \leq n$. Получили $\langle n, 1 \rangle$ решение.

Осталось решить для отрезков, попадающих целиком в один из кусков.

Давайте на каждом куске построим структуру данных для решения RMQ.

Наивный цикл	$\langle n, \log n \rangle$	константа сверхмала
Sparse Table	$\langle n \log \log n, 1 \rangle$	$\frac{n}{\log n} \cdot (\log n \log \log n)$
Дерево отрезков	$\langle n, \log \log n \rangle$	
Рекурсивно построить себя	$\langle n \log^* n, \log^* n \rangle$	$T(n) = 1 + T(\log n), M(n) = n + (n/\log n)(\log n + \dots)$

16.2. RMQ за $\mathcal{O}(1)$ одним махом

Возьмём разбиение на куски длины 32, построим сверху Sparse Table. Для кусков длины 32 сделаем предподсчёт: пройдем слева направо, поддерживая стек элементов, которые являются **минимумами** на суффиксах и маску, какие элементы принадлежат этому стеку, для каждого префикса запоем маску.

Пример: кусок [7, 3, 5, 4, 6, 5, 2, 6].

Состояние стека: [7] → [3] → [3, 5] → [3, 4] → [3, 4, 6] → [3, 4, 5] → [2] → [2, 6]. Маски для соответствующих стеков: 1, 01, 011, 0101, 01011, 010101, 0000001, 00000011.

Предподсчёт стеков и масок работает линейное время. Суммарное время на предподсчёт: $n + \frac{n}{32} \log n + \frac{n}{32} \cdot 32 = \mathcal{O}(n)$. Тем, что куски длины именно 32, мы пользовались только, чтобы на любой машине все операции с масками происходили за $\mathcal{O}(1)$.

Как теперь ответить на запрос? Любому подотрезку $[L, R]$ куска длины 32 соответствует маска префикса $[0, R]$, у которой нужно отбросить первые L бит и найти крайнюю левую единицу. Нахождение крайней левой единицы, это или предподсчёт, или одна процессорная инструкция.

16.3. LCA & Двоичные подьёмы

В дереве с корнем для двух вершин можно определить отношение “ a – предок b ”.

Более того, на запрос `isAncestor(a, b)` легко отвечать за $\mathcal{O}(1)$.

Предподсчитаем времена входа выхода dfs-ом по дереву, тогда:

```
1 bool isAncestor(int a, int b):
2     return t_in[a] <= t_in[b] && t_out[b] <= t_out[a];
```

Def 16.3.1. $LCA(a, b)$ – общий предок вершин a и b максимальной глубины.

LCA = least common ancestor = наименьший общий предок.

Мы уже умеем искать LCA за $\Theta(\text{dist}(a, b))$: предподсчитаем глубины всех вершин, при подсчёте LCA сперва уравниваем глубины a и b , затем будем параллельно подниматься на 1 вверх.

Соптимизируем эту идею – научимся $\forall v, k$ из вершины v прыгать сразу на 2^k вверх.

$\text{up}[k, v] = \text{up}[k-1, \text{up}[k-1, v]]$ – прыжок на 2^k равен двум прыжкам на 2^{k-1} .

База: $\text{up}[0, v] = \text{parent}[v]$.

Переход: если уже построен слой динамики $\text{up}[k-1]$, мы за $\Theta(n)$ насчитаем слой $\text{up}[k]$.

Чтобы не было крайних случаев, сделаем $\text{up}[0, \text{root}] = \text{root}$.

LCA по-прежнему состоит из двух частей – уравнивать глубины и прыгать вверх:

```
1 int K = ⌊log2 N⌋, up[K][N];
2 int LCA(int a, int b) {
3     if (depth[a] < depth[b]) swap(a, b);
4     a = jump(a, depth[a] - depth[b]);
5     for (int k = K - 1; k >= 0; k--)
```

```

6     if (up[k][a] != up[k][b])
7         a = up[k][a], b = up[k][b];
8     return a == b ? a : up[0][a];
9 }

```

Здесь $\text{jump}(v, d)$ за $\mathcal{O}(\log n)$ прыгает вверх из v на d , для этого d нужно представить, как сумму степеней двойки. Время и память подсчёта – $\Theta(n \log n)$, время поиска LCA – $\Theta(\log n)$.

Можно уменьшить константу времени работы, используя `isAncestor`:

```

1 int LCA(int a, int b) {
2     for (int k = K - 1; k >= 0; k--)
3         if (!isAncestor(up[k][a], b))
4             a = up[k][a];
5     return isAncestor(a, b) ? a : up[0][a];
6 }

```

16.4. RMQ±1 за $\langle n, 1 \rangle$

Def 16.4.1. Говорят, что массив обладает ± 1 свойством, если $\forall i |a_i - a_{i+1}| = 1$

Наша цель – решить RMQ на ± 1 массиве.

Подкрутим уже имеющуюся у нас идею из Sparse Table ++.

“Разобьём исходный массив a на куски длины $k = \frac{1}{2} \log n$, минимум на i -м куске обозначим b_i .”

$\min(a_1, a_2, \dots, a_k) = a_1 + \min(0, a_2 - a_1, a_3 - a_1, \dots, a_k - a_1) = a_1 + X$.

Поскольку $\forall i |a_{i+1} - a_i| = 1$, X – минимум на одной из 2^{k-1} последовательностей.

$2^k = 2^{\log n / 2} = \sqrt{n} \Rightarrow$ можно за $o(n)$ предподсчитать ответы для всех последовательностей.

16.5. LCA \rightarrow RMQ±1 и эйлеров обход

Напомним, эйлеров обход графа – цикл, проходящий по каждому ребру ровно один раз.

Если у дерева каждое ребро заменить на два ориентированных, мы получим эйлеров оргграф.

Чтобы построить эйлеров обход дерева, пишем `dfs(v)`.

```

1 void dfs(int v) {
2     for (Edge e : graph[v]) { // пусть мы храним только рёбра, ориентированные вниз
3         answer.push_back(edge(v, x));
4         dfs(x);
5         answer.push_back(edge(x, v));
6     }
7 }

```

Такой обход назовём обычным или “эйлеровым обхода дерева первого типа”.

Иногда имеет смысл хранить другую информацию после обхода:

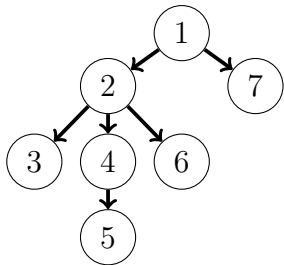
```

1 void dfs2(int v) {
2     index[v] = answer.size(); // сохранили для вершины v любое её вхождение в answer
3     answer.push_back(v);
4     for (Edge e : graph[v]) {
5         dfs2(x);
6         answer.push_back(v); // поднимаемся, проходим через вершину v
7     }
8 }
9 void dfs3(int v) {
10    L[v] = answer.size(); // аналог времени входа

```

```

11  answer.push_back(v); // просто сохраняем порядок обхода вершин dfs-ом
12  for (Edge e : graph[v])
13      dfs3(x);
14  R[v] = answer.size(); // аналог времени выхода
15  }
    
```



Эйлеров обход 1-го типа (обычный)	(1,2) (2,3) (3,2) (2,4) (4,5) (5,4) (4,2) (2,6) (6,2) (2,1) (1,7) (7,1)
Эйлеров обход второго типа (± 1)	Обход: 1 2 3 2 4 5 4 2 6 2 1 7 1 Высоты: 0 1 2 1 2 3 2 1 2 1 0 1 0
Обход третьего типа	Обход: 1 2 3 4 5 6 7

Второй обход по сути тоже эйлеров.

Мы сохраняем не рёбра, по которым проходим, а вершины, в которых оказываемся при обходе. Третий обход уже слабо напоминает эйлеров, но мы в контексте задач, где нужно выбирать между 2-м и 3-м обходами, будем иногда называть его эйлеровым.

Зачем нужны 2-й и 3-й обходы будет понятно уже сейчас, 1-й нам пригодится для ЕТТ.

Lm 16.5.1. После третьего обхода отрезок обхода $[L_v, R_v)$ задаёт ровно поддерево вершины v .

Следствие 16.5.2. Пусть у каждой вершины дерева есть вес w_v . Тогда мы теперь умеем за $\mathcal{O}(\log n)$ делать все операции на поддереве, которые ДО умело делал на отрезке.

Lm 16.5.3. Массив высот $h[i] = \text{height}[\text{answer}[i]]$ второго обхода обладает ± 1 свойством.

Lm 16.5.4. $LCA(a, b) = \text{answer}[h.RMQ[\text{index}[a], \text{index}[b]]]$

Доказательство. dfs по пути из a в b пройдёт через LCA. Это будет вершина минимальной высоты на пути, так как, чтобы попасть в ещё меньшие, dfs должен сперва выйти из LCA. ■

Замечание 16.5.5. Итого мы получили решение задачи LCA за $\langle n, 1 \rangle$. Этот относительно свежий результат был получен в 2000-м Фарах-Колтоном и Бендером (два человека). [Ссылка на статью.](#)

Замечание 16.5.6. На практике популярен способ решения LCA: LCA \rightarrow RMQ, а RMQ решим через Sparse Table. Это $\langle n \log n, 1 \rangle$, причём у $\mathcal{O}(1)$ относительно небольшая константа.

16.6. RMQ \rightarrow LCA

Чтобы свести задачу “RMQ на массиве a ” к LCA, построим декартово дерево на парах (i, a_i) . Пары уже отсортированы по x , поэтому построение – проход со стеком за $\mathcal{O}(n)$.

Lm 16.6.1. RMQ на $[l, r]$ в исходном массиве равно LCA(l, r) в полученном дереве.

Доказательство. Каждой вершине декартова дерева соответствует отрезок исходного массива (посмотрим на симметричный обход), корнем поддерева выбирается минимум по $y_i = a_i$ на этом отрезке.

Будем спускаться от корня дерева, пока не встретим вершину, которая разделяет l и r .

Ключ, записанный в найденной вершине, обозначим i , отрезок вершины $[L_i, R_i] \Rightarrow a_i = \min_{L_i \leq j \leq R_i} a_j$ и $L_i \leq l \leq i \leq r \leq R_i \Rightarrow a_i \geq \min_{l \leq j \leq r} a_j$ и $i \in [l, r]$.

Осталось заметить, что $i = LCA(l, r)$. ■

Следствие 16.6.2. Мы научились решать статичную версию RMQ за $\langle n, 1 \rangle$. Победа!

16.7. (*) LCA в offline, алгоритм Тарьяна

Для каждой вершины построим список запросов с ней связанных. Пусть i -й запрос – (a_i, b_i) .

```
1 q[a[i]].push_back(i), q[b[i]].push_back(i)
```

Будем обходить дерево dfs-ом, перебирать запросы, связанные с вершиной, и отвечать на все запросы, второй конец которых серый или чёрный.

```
1 void dfs( int v ) {
2     color[v] = GREY;
3     for (int i : q[v]) {
4         int u = a[i] + b[i] - v;
5         if (color[u] != WHITE)
6             answer[i] = DSU.get(u);
7     }
8     for (int x : graph[v])
9         dfs(x), DSU.parent[x] = v;
10    color[v] = BLACK;
11 }
```

Серые вершины образуют путь от v до корня. У каждой серой вершины есть чёрная часть поддерева, это и есть её множество в DSU. $LCA(v, u)$ – всегда серая вершина, то есть, нужно от u подниматься вверх до ближайшей серой вершины, что мы и делаем.

В коде для краткости используется DSU со сжатием путей, но без ранговой эвристики, поэтому время работы будет $\mathcal{O}((m+n)\log n)$. Если применить обе эвристики, получится $\mathcal{O}((m+n)\alpha)$, но нужно будет поддерживать в корне множества “самую высокую вершину множества”.

16.8. LA (level ancestor)

Запрос $LA(v, k)$ – подняться в дереве от вершины v на k шагов вверх.

Мы уже умеем решать эту задачу за $\langle n \log n, \log n \rangle$ двоичными подъёмами.

В offline на m запросов можно ответить dfs-ом за $\mathcal{O}(n+m)$: когда dfs зашёл в вершину v , у нас в стеке хранится весь путь до корня, и к любому элементу пути мы можем обратиться за $\mathcal{O}(1)$.

• Алгоритм Вишкина

Как и при сведении $LCA \rightarrow RMQ_{\pm 1}$, выпишем высоты Эйлера обхода второго типа.

$LA(v, k) = getNext(index[v], height[v]-k)$, где $index$ – позиция в Эйлера обходе, а $getNext(i, x)$ возвращает ближайший справа от i элемент $\leq x$.

Мы умеем отвечать на $getNext(i, x)$ за $\langle n, \log n \rangle$ одномерным ДО снизу или сверху.

16.9. (*) Euler-Tour-Tree

Задача: придумать структуру данных для хранения графа, поддерживающую операции

- $link(a, b)$ – добавить ребро между a и b .
- $cut(a, b)$ – удалить ребро между a и b .
- $isConnected(a, b)$ – проверить связность двух вершин.

При этом в каждый момент времени выполняется условие отсутствия циклов (граф – лес).

По сути мы решаем *Dynamic Connectivity Problem* с дополнительным условием “граф – лес”.

Решение: хранить обычный эйлеров обход дерева (ориентированные рёбра).

Эйлеров обход – массив. Заведём заведём на нём горе. Например, treap по неявному ключу.

```

1 map<pair<int,int>, Node*> node; // по ребру умеем получать вершину treap
2 vector<Node*> anyEdge; // для каждой вершины графа храним любое исходящее ребро
3 bool isConnected(int a, int b) {
4     // взяли у каждой вершины произвольное ребро, проверили, что два ребра живут в одном дереве
5     return getRoot(anyEdge[a]) == getRoot(anyEdge[b]);
6 }
7 void cut(int a, int b) {
8     // по ребру получаем Node*, находим его позицию в эйлеровом обходе
9     Node *node1 = node[make_pair(a, b)];
10    Node *node2 = node[make_pair(b, a)];
11    int i = getPosition(node1), j = getPosition(node2);
12    if (i > j) swap(i, j); // упорядочили (i,j)
13    Node *root = getRoot(node1), *a, *b, *c;
14    Split(root, a, b, i);
15    Split(b, b, c, j - i); // разделили дерево на три части: (a) (i b) (j c)
16    Delete(b, 0), Delete(c, 0); // собственно удаление лишнего ребра
17    Merge(a, c); // в итоге теперь есть два дерева: (a c) и (b)
18 }

```

С операцией link(a,b) чуть сложнее – нужно сделать циклические сдвиги обходов: если обходы называются X и Y, а добавляемые рёбра e_1 и e_2 , мы хотим представить ответ, как Xe_1Ye_2 .

```

1 void link(int a, int b) {
2     Node *pa = anyEdge[a], *pb = anyEdge[b];
3     Rotate(getRoot(pa), getPosition(pa)); // Rotate = Split + Merge
4     Rotate(getRoot(pb), getPosition(pb));
5     // теперь первые ребра эйлеровых обходов - исходящие из a и b соответственно
6     Node *e1 = createEdge(a, b), *e2 = createEdge(b, a);
7     Merge(Merge(getRoot(pa), e1), Merge(getRoot(pb), e2));
8 }

```

Можно реализовать pair<Node*,int> goUp(Node*), которая заменит и getRoot, и getPosition.

16.10. (*) LA, быстрые решения

16.10.1. (*) Вишкин за $\langle \mathcal{O}(n \log n), \mathcal{O}(1) \rangle$

В Section 16.8 у нас уже построен Эйлеров обход и мы знаем, что $LA(v,k) = getNext(index[v], height[v]-k) = getNext(i, a[i]-k)$.

Осталось научиться вычислять getNext за $\mathcal{O}(1)$.

Давайте $\forall i$ предподсчитаем getNext(i, a[i]-j) для всех j до $3 \cdot 2^k$, где 2^k – максимальная степень двойки, которая делит i. Суммарный размер предподсчёта = $\sum_k \frac{2n}{2^k} (3 \cdot 2^k) = \Theta(n \log n)$, где 2n – длина Эйлерова обхода. Предподсчёт делается через LA-offline за то же время.

Ответ на запрос: выберем $s: 2^s \leq k < 2^{s+1}$.

Поскольку $getNext(i, a[i]-k) \geq i + k \geq i + 2^s$, перейдём к $j = i - (i \bmod 2^s) + 2^s$.

Заметим, что getNext(j, a[i]-k) предподсчитан, так как $a[j] - a[i] \leq 2^s \wedge k \leq 2 \cdot 2^s$.

16.10.2. (*) Ladder decomposition + четыре русских

• Longest-Path-Decomposition и $\langle \mathcal{O}(n), \mathcal{O}(\sqrt{n}) \rangle$

Longest-Path-Decomposition – разбиение вершин дерева на пути, путь от вершины продолжается

вниз в сторону самого глубокого сына. Внутри одного пути LA считается за $\mathcal{O}(1)$. Иначе будем прыгать вверх по путям. Прыжков $\mathcal{O}(\sqrt{n})$, так как, на k -м подъёме у нас должен быть путь длины $\geq k \Rightarrow$ всего в дереве $1 + \dots + k = \Theta(k^2)$ вершин.

• **Ladder-Decomposition** и $\langle \mathcal{O}(n), \mathcal{O}(\log n) \rangle$

Возьмём пути из прошлого решения.

Пусть k_i – длина пути, насчитаем продолжение пути вверх на k_i .

Размер и время предподсчёта всё ещё линейны.

Прыгаем вверх. Пусть мы сейчас в вершине v и прошли снизу расстояние $x \Rightarrow$ из v есть путь вниз длины $\geq x \Rightarrow$ вверх тоже $\geq x \Rightarrow$ каждый раз расстояние удваивается.

• **Ladder-Decomposition** и $\langle \mathcal{O}(n \log n), \mathcal{O}(1) \rangle$

Добавим двоичных подъёмов. Отсюда и предподсчёт $\mathcal{O}(n \log n)$. С их помощью, желая прыгнуть на k , за $\mathcal{O}(1)$ прыгнем на 2^s : $2^s \leq k < 2^{s+1}$. Из вершины, в которой мы оказались, путь ladder decomposition позволяет подняться за $\mathcal{O}(1)$ на оставшиеся $k - 2^s \leq 2^s$.

• **Эйлеров обход, четыре русских** и $\langle \mathcal{O}(n \log n), \mathcal{O}(1) \rangle$

Нужно насчитать двоичные подъёмы не ото всех вершин, а лишь от $\Theta(\frac{n}{\log n})$ каких-то...

Предподсчитывать динамикой $f[i, v] = f[i-1, f[i-1, v]]$ не получится \Rightarrow используем LA-offline.

Алгоритм:

Берём ± 1 эйлеров обход, выбираем m и от каждой m -й вершины считаем подъёмы.

Чтобы ответить на запрос $LA(v, k)$ берём $i = pos[v]$ в эйлеровом обходе, $j = i - (i \bmod m) + m$.

Если $h[j] \leq h[i] + k$, то отвечаем за $\mathcal{O}(1)$ от j , иначе ответ лежит на $(j - m, j]$.

Ответ зависит только от битовой строки длины m , чисел $j - i \leq m$ и $k \leq m$.

Итого $2^m m^2$ различных задач, ответы на которые мы предподсчитаем. Пусть $m = \frac{1}{2} \log n$.