Э. Э. Гасанов, Оценки сложности одного метода решения задачи включающего поиска, *Дискрет. матем.*, 2000, том 12, выпуск 2, 118–139

DOI: 10.4213/dm329

Дискретная математика

том 12 выпуск 2 * 2000

УДК 519.7

Оценки сложности одного метода решения задачи включающего поиска

© 2000 г. Э. Э. Гасанов

В работе предлагается метод решения задачи включающего поиска, имеющий три модификации в зависимости от выбранного базового множества: множества монотонных булевых функций, множества элементарных монотонных конъюнкций и множества булевых переменных. Для каждой из модификаций оценивается функция Шеннона сложности метода и среднее значение сложности, причем для функций Шеннона сложности метода найдена асимптотика, совпадающая с асимптотикой функции Шеннона сложности включающего поиска, а для среднего значения — асимптотика логарифма.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, грант 98-01-00130.

1. Введение

В работе рассматривается задача включающего поиска [1, 2, 3, 4], которая состоит в поиске в конечном подмножестве единичного булева куба (это подмножество называется библиотекой) всех точек, которые по каждой из компонент не больше некоторой произвольно взятой точки-запроса. Исходя из имеющейся библиотеки и разрешенного для использования множества функций, называемого базовым множеством функций (в качестве которого могут быть выбраны множество монотонных булевых функций, множество элементарных монотонных конъюнкций и множество булевых переменных), предлагается некоторый метод организации данных и соответствующий этой организации алгоритм поиска, который позволяет решать задачу включающего поиска. Под сложностью алгоритма поиска при заданной библиотеке понимается среднее время поиска, измеряемое количеством вычисленных при поиске функций из базового множества. Для заданного натурального числа k исследуются четыре функции: функция Шеннона сложности метода (сложности включающего поиска), равная максимуму сложности предлагаемого алгоритма (оптимального алгоритма), достигаемому при вариации библиотеки по множеству k-элементных библиотек, и среднее значение сложности предлагаемого алгоритма (оптимального алгоритма), соответствующего библиотеке, при вариации библиотеки по множеству k-элементных библиотек. Для функции Шеннона сложности метода получены асимптотические оценки, а для средней сложности метода получена асимптотика ее логарифма. Показано, что асимптотика функции Шеннона сложности метода совпадает с функцией Шеннона сложности включающего поиска.

2. Основные понятия и формулировка результатов

Мы будем использовать терминологию и обозначения из работы [5], но с тем отличием, что понятие информационной сети с переключателями заменено в этой работе на понятие информационного графа, причем поскольку мы нигде в этой работе не будем использовать переключатели, здесь будет приведена несколько упрощенная версия понятия информационного графа.

Пусть X — множество запросов, причем на X определено вероятностное пространство $(X, \sigma, \mathbf{P})$, где σ — алгебра подмножеств множества X, $\mathbf{P}$ — вероятностная мера на σ; Y — множество записей (объектов поиска); ρ — бинарное отношение на $X \times Y$, называемое отношением поиска. Пятерку $S = \langle X, Y, \rho, \sigma, \mathbf{P}\rangle$ будем называть типом; тройку $I = \langle X, V, \rho\rangle$, где V — некоторое конечное подмножество множества Y, в дальнейшем называемое библиотекой, будем называть задачей информационного поиска (ЗИП) типа $S = \langle X, Y, \rho, \sigma, \mathbf{P}\rangle$ или ЗИП, принадлежащей типу S, и обозначать $I \in S$. Будем считать, что задача $I = \langle X, V, \rho\rangle$ содержательно состоит в перечислении для произвольно взятого запроса $x \in X$ всех тех и только тех записей из V, которые находятся в отношении ρ с запросом x, то есть удовлетворяют запросу x; $O(y, \rho) = \{x \in X : x\rho y\}$ — тень записи $y \in Y$; если f — одноместный предикат, определенный на X, то есть $f : X \to \{0, 1\}$, то $N_f = \{x \in X : f(x) = 1\}$ — характеристическое множество функции f; функция $\chi_{y,\rho} : X \to \{0, 1\}$ такая, что $N_{\chi_{y,\rho}} = O(y, \rho)$ — характеристическая функция записи y; F — множество символов одноместных предикатов, определенных на множестве X, называемое базовым множеством.

Понятие информационного графа (ИГ) над базовым множеством F, определяется следующим образом. Берется конечная многополюсная ориентированная сеть. В ней выбирается некоторый полюс, который называется корнем. Остальные полюса называются листьями и им приписываются записи из Y, причем разным листьям могут быть приписаны одинаковые записи. Ребрам приписываются предикаты из множества F. Таким образом нагруженную многополюсную ориентированную сеть называем ИГ над базовым множеством F.

Функционирование ИГ определяется следующим образом. Скажем, что ребро проводит запрос $x \in X$, если предикат, приписанный этому ребру, принимает значение 1 на запросе x; ориентированная цепочка ребер проводит запрос $x \in X$, если каждое ребро цепочки проводит запрос x; запрос $x \in X$ проходит в вершину β ИГ, если существует ориентированная цепочка, ведущая из корня в вершину β, которая проводит запрос x; запись y, приписанная листу α, попадает в ответ ИГ на запрос $x \in X$, если запрос x проходит в лист α. Ответом ИГ U на запрос x назовем множество записей, попавших в ответ ИГ на запрос x, и обозначим его $\mathcal{J}_U(x)$. Функцию $\mathcal{J}_U(x)$ будем считать результатом функционирования ИГ U.

Пусть нам дана ЗИП $I = \langle X, V, \rho\rangle$. Скажем, что ИГ U разрешает ЗИП $I = \langle X, V, \rho\rangle$, если

$$\mathcal{J}_U(x) = \{y \in V : x\rho y\}.$$

Введем понятие сложности ИГ. Пусть β — некоторая вершина ИГ. Предикат, определенный на множестве запросов, который принимает значение 1 на запросе x, если запрос проходит в вершину β, и 0 в противном случае, назовем функцией фильтра вершины β и обозначим $\varphi_\beta(x)$.

Сложностью ИГ U на запросе $x \in X$ назовем число

$$T(U,x) = \sum_{\beta \in \mathscr{R}} \psi_\beta \varphi_\beta(x),$$

где $\mathscr{R}$ — множество вершин ИГ U, ψ_β — количество ребер, исходящих из вершины β.

Скажем, что базовое множество F измеримо, если каждая функция из F измерима (относительно алгебры σ). Далее всюду будем предполагать, что базовое множество измеримо. В этом случае для любого ИГ U над F функция $T(U,x)$ как функция от x измерима.

Сложностью ИГ U назовем математическое ожидание величины $T(U,x)$, то есть число $T(U) = \mathbf{M}_x T(U,x)$. Сложностью задачи I при базовом множестве F назовем число

$$T(I,F) = \inf\{T(U)\colon U \in \mathscr{U}(I,F)\},$$

где $\mathscr{U}(I,F)$ — множество всех ИГ над базовым множеством F, разрешающих ЗИП I.

Скажем, что базовое множество F полно для типа S, если $\mathscr{U}(I,F) \neq \varnothing$ для любой ЗИП I типа S.

Пусть M — некоторое множество. Через $|M|$ обозначим число элементов в множестве M, называемое мощностью множества M.

Будем писать $\alpha(n) = o(1)$, если $\alpha(n) \to 0$ при $n \to \infty$, и $A(n) = o(B(n))$, если $A(n) = B(n)o(1)$.

Скажем, что $A(n)$ асимптотически не превосходит $B(n)$ при $n \to \infty$ и будем писать $A \lesssim B$, если существует $\alpha(n) = o(1)$ такое, что начиная с некоторого номера n_0, $A(n) \leqslant (1+\alpha(n))B(n)$. Если $A \lesssim B$ и $B \lesssim A$, то будем говорить, что A и B асимптотически равны при $n \to \infty$ и писать $A \sim B$.

Будем писать $A \overset{\frown}{<} B$, если существует такая положительная константа c, что $A(n) \leqslant cB(n)$, начиная с некоторого номера n_0. Если $A \overset{\frown}{<} B$ и $B \overset{\frown}{<} A$, то будем говорить, что A и B равны по порядку при $n \to \infty$ и писать $A \asymp B$ или $A = \underline{O}(B)$.

Через $\binom{n}{k}$ будем обозначать число сочетаний из n элементов по k, а через A_n^k число размещений из n элементов по k.

Рассмотрим тип задач поиска $S_b = \langle B^n, B^n, \succeq, \sigma, \mathsf{P}\rangle$, где B^n — единичный n-мерный куб, $\succeq$ — отношение поиска на $B^n \times B^n$, определяемое соотношением

$$(x_1,\dots,x_n) \succeq (y_1,\dots,y_n) \Longleftrightarrow x_i \geqslant y_i, \quad i = 1,\dots,n,$$

σ — алгебра подмножеств B^n, представляющая собой множество всех подмножеств B^n, P — равномерная вероятностная мера на σ, то есть такая мера, что $\mathsf{P}(x) = 1/2^n$ для любого $x \in B^n$ и $\mathsf{P}(A) = |A|/2^n$ для любого $A \subseteq B^n$.

Задачи поиска, принадлежащие данному типу, суть разновидность задач, именуемых в литературе задачами включающего поиска (см., например, [1]), поэтому тип S_b мы назовем типом включающего поиска, а задачи, принадлежащие этому типу, задачами включающего поиска.

Напомним, что в соответствии с терминологией [6] весом набора $(\alpha_1,\dots,\alpha_n) \in B^n$ называют число его координат, равных 1. Формула

$$x_{i_1}^{\sigma_1} \,\&\, x_{i_2}^{\sigma_2} \,\&\, \dots \,\&\, x_{i_r}^{\sigma_r},$$

где & — знак конъюнкции, $\sigma_k \in \{0,1\}$, $x_{i_k}^0 = \overline{x}_{i_k}$, $x_{i_k}^1 = x_{i_k}$, $i_k \in \{1,2,\dots,n\}$, $k = 1,2,\dots,r$, $r \geqslant 1$ и $n \geqslant 1$, называется конъюнкцией над множеством переменных $X^n = \{x_1,x_2,\dots,x_n\}$. Далее часто вместо символа & мы будем использовать символ $\cdot$. Если $x_{i_j} \neq x_{i_k}$ при $j \neq k$, то конъюнкция называется элементарной. Элементарная конъюнкция называется монотонной, если она не содержит отрицаний переменных. Длиной элементарной конъюнкции будем называть количество переменных в конъюнкции. Функцию тождественная 1 будем считать элементарной монотонной конъюнкцией длины 0. Множество элементарных монотонных конъюнкций от n переменных будем обозначать через $\mathscr{K}^n$. Будем считать, что X^n содержит тождественную 1.

Функция алгебры логики $f(x_1, \dots, x_n)$ называется монотонной, если для любых двух наборов α и β из B^n таких, что $\alpha \succeq \beta$, имеет место неравенство $f(\alpha) \geqslant f(\beta)$. Дизъюнкция элементарных монотонных конъюнкций есть монотонная функция. Множество монотонных булевых функций от n переменных будем обозначать через $\mathscr{M}^n$.

Рассмотрим произвольную запись $y \in B^n$. Пусть $\{i_1, \dots, i_k\}$ есть множество номеров координат вектора y, которые равны 1. Нетрудно заметить, что $\chi_{y,\succeq}(x_1, \dots, x_n) = x_{i_1} \cdot x_{i_2} \cdot \dots \cdot x_{i_k}$, то есть характеристическая функция записи y является элементарной монотонной конъюнкцией.

Отсюда, согласно теореме 1 из [5] ИГ, разрешающий некоторую задачу включающего поиска, представляет собой ИГ, реализующий как функции проводимости некоторую систему элементарных монотонных конъюнкций. Откуда согласно теореме 2 из [7] следует, что каждое из базовых множеств $\mathscr{M}^n$, $\mathscr{K}^n$ и X^n является полным для типа S_b.

Через $\mathscr{W}_n^k$ ($\mathscr{V}_n^k$) обозначим множество всех упорядоченных (неупорядоченных) k-элементных подмножеств B^n, то есть

$$|\mathscr{W}_n^k| = A_{2^n}^k, \qquad |\mathscr{V}_n^k| = \binom{2^n}{k}.$$

Далее нам удобно считать, что библиотека является упорядоченным множеством. Пусть n, k — натуральные числа и

$$\mathscr{I}(n,k) = \{I = \langle B^n, V, \succeq\rangle \in S_b \colon V \in \mathscr{W}_n^k\},$$
$$T(n,k,F) = \max_{I \in \mathscr{I}(n,k)} T(I,F),$$
$$\overline{T}(n,k,F) = |\mathscr{W}_n^k|^{-1} \sum_{I \in \mathscr{I}(n,k)} T(I,F).$$

В работе предлагается три алгоритма, которые позволяют для произвольно взятой библиотеки $V \subseteq B^n$ строить разрешающие ЗИП $I = \langle B^n, V, \succeq\rangle$ информационные графы соответственно над базовыми множествами $\mathscr{M}^n$, $\mathscr{K}^n$ и X^n. Эти алгоритмы будем обозначать соответственно P_1, P_2, P_3. Поскольку в задаче включающего поиска множество B^n и отношение $\succeq$ фиксированы, то будем считать, что алгоритм P_i, $i = 1, 2, 3$, применяется к библиотеке V и результатом его применения является разрешающий задачу $I = \langle B^n, V, \succeq\rangle$ ИГ, который обозначим через $P_i(V)$. Поскольку ИГ $P_i(V)$ существенно будет зависеть от порядка следования элементов в библиотеке V, будем считать, что библиотеки — упорядоченные множества. Каждый ИГ $P_i(V)$, $i = 1, 2, 3$, будет иметь вид дерева, и мы, используя в дальнейшем термин дерево, будем подразумевать некоторую его укладку, причем будем считать, что в каждой вершине среди ребер, исходящих из данной вершины, задан некоторый порядок следования, и будем называть первое по порядку ребро самым левым, а последнее самым правым, и соответственно будем определять понятия левее, правее.

Будем говорить, что функция f покрывает функцию g, если $N_g \subseteq N_f$.

Пусть f,g — элементарные монотонные конъюнкции. Пересечением конъюнкций f и g назовем конъюнкцию переменных, принадлежащих как f, так и g, и будем обозначать ее через $f \cap g$. Таким образом, пересечение f и g есть минимальная по числу элементов в характеристическом множестве элементарная монотонная конъюнкция, покрывающая и f, и g. В случае, когда все переменные f отличны от переменных g, будем говорить, что пересечение f и g пусто. Через $f \setminus g$ обозначим конъюнкцию переменных, принадлежащих f, но не принадлежащих g.

Будем говорить, что ИГ U допустим для библиотеки V, если ИГ U разрешает ЗИП $I = \langle B^n, V, \succeq \rangle$.

Обозначим через $\tilde{0}$ набор, состоящий из всех нулей, то есть $\tilde{0} = (0, \dots, 0)$.

Пусть дана упорядоченная библиотека V. Положим $V' = V \setminus \{\tilde{0}\} = \{y_1, y_2, \dots, y_k\}$. Индуктивно опишем процесс построения ИГ $P_2(V')$.

В качестве базиса индукции возьмем ИГ $P_2(\{y_1\})$, который состоит из одного ребра, начало которого является корнем ИГ, а конец — листом, которому приписана запись y_1, этому ребру приписана также функция $\chi_{y_1,\succeq}$.

Проведем индуктивный переход. Пусть построен ИГ

$$P_2(\{y_1, \dots, y_{i-1}\}), \qquad i = 2, 3, \dots, k.$$

Перестраивать его, чтобы он стал допустим для библиотеки $\{y_1, \dots, y_i\}$, будем следующим образом.

1. Объявим в качестве текущей вершины β корень ИГ $P_2(\{y_1, \dots, y_{i-1}\})$. Объявим $g = \chi_{y_i,\succeq}$.

2. Будем слева направо просматривать ребра, исходящие из вершины β до тех пор пока не найдем ребра, пересечение нагрузочной функции которого с функцией g не пусто.

2.1. Если для всех ребер пересечение их нагрузочных функций с g пусто, то из вершины β выпускаем новое ребро с нагрузочной функцией g, конец ребра объявляем листом и приписываем ему запись y_i, и на этом завершаем работу, ИГ $P_2(\{y_1, \dots, y_i\})$ построен. Здесь, как и везде далее, новое выпускаемое ребро становится самым правым в пучке на данный момент.

2.2. Если же нашлось ребро с нагрузочной функцией f такой, что пересечение f и g не пусто, то поступаем следующим образом.

2.2.1. Если $f = g$, то конец этого ребра объявляем листом и приписываем ему запись y_i, и на этом завершаем работу, ИГ $P_2(\{y_1, \dots, y_i\})$ построен.

2.2.2. Если $f \neq g$ и $f = f \cap g$, то конец этого ребра объявляем текущей вершиной β, объявляем $g = g \setminus f$ и переходим к шагу 2.

2.2.3. Если же $f \neq f \cap g$, то изменим нагрузку рассматриваемого ребра на $f \cap g$, пусть конец этого ребра обозначен через β', отцепим ветвь, растущую из β', выпустим из β' новое ребро и прицепим отцепленную ветвь к концу нового ребра (если вершина β' была листом, то листом вместе с его нагрузкой становится конец нового ребра) и приписываем новому ребру функцию $f \setminus (f \cap g)$, после чего из вершины β' выпустим еще одно ребро с нагрузкой $g \setminus (f \cap g)$, конец последнего ребра объявляем листом, приписываем ему запись y_i и на этом завершаем работу, ИГ $P_2(\{y_1, \dots, y_i\})$ построен.

Теперь, если $V \neq V'$, то выпустим из корня $P_2(V')$ еще одно ребро и пусть оно будет самым правым, исходящим из корня. Припишем данному ребру функцию тождественная 1, а конец ребра объявим листом и припишем ему запись $\tilde{0}$. Полученный ИГ и будет $P_2(V)$.

Неформально процесс построения $P_2(V)$ можно описать следующим образом. Нам надо в виде дерева реализовать некоторую систему элементарных монотонных конъюнкций, и пусть часть этой системы уже реализована. Берем очередную конъюнкцию, и просматривая ребра, исходящие из корня дерева, ищем такое, чья нагрузка пересекается по переменным с нашей конъюнкцией. Когда находим такое ребро, то мы проходим к вершине, являющейся концом этого ребра, и в этой вершине процедуру повторяем. Если же такое ребро не находится, то выпускаем новое ребро.

Заметим, что $P_2(V)$ есть ИГ над базовым множеством $\mathscr{K}^n$.

Введем обозначение

$$\chi_{V,\succeq} = \bigvee_{y \in V} \chi_{y,\succeq}.$$

Теперь легко описать ИГ $P_1(V)$. Возьмем вершину и объявим ее корнем, выпустим из нее ребро, которому припишем функцию $\chi_{V,\succeq}$, а из конца ребра выпустим ветвь, идентичную $P_2(V)$. ИГ $P_1(V)$ построен. Очевидно, что $P_1(V)$ есть ИГ над базовым множеством $\mathscr{M}^n$.

По сути первое ребро ИГ $P_1(V)$ позволяет на первом же шаге отсечь все запросы, для которых в библиотеке V ответ пуст.

Пусть, как и ранее, V — упорядоченная библиотека и

$$V' = V \setminus \{\tilde{0}\} = \{y_1, y_2, \ldots, y_k\}.$$

Индуктивно опишем процесс построения ИГ $P_3(V')$.

В качестве базы индукции $P_3(\{y_1\})$. Если

$$\chi_{y_1,\succeq} = x_{i_1} \cdot \ldots \cdot x_{i_t}, \qquad 1 \leqslant i_1 < \ldots < i_t \leqslant n,$$

то $P_3(\{y_1\})$ представляет собой ориентированную цепочку из t ребер, причем начало цепочки есть корень ИГ, конец цепочки лист с записью y_1, а j-му ребру цепочки, $j = 1, 2, \ldots, t$, приписана функция x_{i_j}.

Проведем индуктивный переход. Пусть построен ИГ

$$P_3(\{y_1, \ldots, y_{i-1}\}), \qquad i = 2, 3, \ldots, k.$$

Перестраивать его, чтобы он стал допустим для библиотеки $\{y_1, \ldots, y_i\}$, будем следующим образом.

1. Корень ИГ $P_3(\{y_1, \ldots, y_{i-1}\})$ объявим текущей вершиной β. Объявим $g = \chi_{y_i,\succeq}$.

2. Будем слева направо просматривать ребра, исходящие из вершины β, до тех пор пока не найдем ребро, которому соответствует переменная, покрывающая g.

2.1. Если для всех ребер, исходящих из β, переменная, соответствующая ребру, не покрывает g, то из вершины β выпускаем новую цепочку ребер, аналогичную цепочке, описанной в базисе индукции, такую, чтобы конъюнкция переменных, приписанных ребрам цепочки, была равна g, конец цепочки объявляем листом, приписываем ему запись y_i и завершаем работу, ИГ $P_3(\{y_1, \ldots, y_i\})$ построен. Здесь первое ребро новой цепочки становится самым правым, исходящим из β.

2.2. Если же нашлось ребро с нагрузочной функцией x_j, покрывающей g, то в случае, когда $g \neq x_i$, конец этого ребра объявляем текущей вершиной β, объявляем

$g = g \setminus x_j$ и переходим к шагу 2, иначе конец рассматриваемого ребра объявляем листом, приписываем ему запись y_i и завершаем работу, ИГ $P_3(\{y_1, \dots, y_i\})$ построен.

Теперь, если $V \neq V'$, то выпустим из корня $P_3(V')$ еще одно ребро и пусть оно будет самым правым, исходящим из корня. Припишем данному ребру функцию тождественная 1, а конец ребра объявим листом и припишем ему запись $\tilde{0}$. Полученный ИГ и будет $P_3(V)$.

Легко заметить, что по построению каждого из $P_1(V)$, $P_2(V)$, $P_3(V)$ к произвольной записи $y \in V$ ведет цепочка ребер, проводимость которой равна $\chi_{y,\succeq}$. Отсюда, согласно теореме 1 из [5] каждый из ИГ $P_1(V)$, $P_2(V)$, $P_3(V)$ допустим для библиотеки V.

Пусть

$$T(P_i, k, n) = \max_{V \in \mathscr{W}_n^k} T(P_i(V)),$$
$$\overline{T}(P_i, k, n) = (A_{2^n}^k)^{-1} \sum_{V \in \mathscr{W}_n^k} T(P_i(V)),$$

где $i = 1, 2, 3$.

Справедливы следующие теоремы.

Теорема 1. *Если $n \to \infty$ и $k(n)$ и $m(n)$ — такие числа, зависящие от n, что $k \to \infty$, $m(n) = o(n)$ при $n \to \infty$, то при $\binom{n}{m-1} = o(k)$, $k \leqslant \binom{n}{m}$ и $i = 1, 2$*

$$T(n, k, \mathscr{M}^n) \sim T(n, k, \mathscr{K}^n) \sim T(P_i, k, n) \sim 2^{1-m}k,$$

при $\binom{n}{m} \leqslant k \leqslant 2\binom{n}{m}$

$$2^{-m}\left(k + \binom{n}{m}\right) \leqslant T(n, k, \mathscr{M}^n) \leqslant T(P_1, k, n) \lesssim 2^{1-m}k,$$
$$2^{-m}\left(k + \binom{n}{m}\right) \leqslant T(n, k, \mathscr{K}^n) \leqslant T(P_2, k, n) \lesssim 2^{1-m}k,$$

при $k \asymp \binom{n}{m}$ и $k \geqslant 2\binom{n}{m}$

$$2^{-m}\left(k + \binom{n}{m}\right) \leqslant T(n, k, \mathscr{M}^n) \leqslant T(P_1, k, n) \lesssim 2^{-m}\left(k + 2\binom{n}{m}\right),$$
$$2^{-m}\left(k + \binom{n}{m}\right) \leqslant T(n, k, \mathscr{K}^n) \leqslant T(P_2, k, n) \lesssim 2^{-m}\left(k + 2\binom{n}{m}\right).$$

Как видно из теоремы 1, в худшем случае задачи включающего поиска решаются сложно по времени. Так например, когда $k \asymp n^m$, где m — константа, то в худшем случае среднее время поиска пропорционально перебору.

Теорема 2. *Если $n \to \infty$ и $k(n)$, $m(n)$ и $\alpha(n)$ — такие числа, зависящие от n, что*

$$k(n) \sim \binom{n}{m-1}\alpha(n), \quad m(n) \geqslant 1, \quad m(n) = o(\log_2 n / \log_2 \log_2 n),$$
$$\alpha(n) \to \infty, \qquad (\log_2 n - \log_2 \alpha(n))/(m(n)\log_2(m(n)+1)) \to \infty$$

при $n \to \infty$, то для любых таких k, n

$$\max_{I \in \mathscr{I}(n,k)} \inf_{U \in \mathscr{U}'(I,X^n)} T(U) \sim T(P_3, k, n) \sim 2^{2-m} k,$$

где $\mathscr{U}'(I, X^n)$ — множество древовидных ИГ над X^n, разрешающих ЗИП I.

Тем самым, при использовании переменных время поиска ухудшается вдвое по сравнению с использованием конъюнкций.

Теорема 3. *Если $n \to \infty$ и $k(n) \to \infty$ и $\log_2 k(n) = o(n)$ при $n \to \infty$, то*

$$\overline{T}(n, k, \mathscr{M}^n) \sim \overline{T}(P_1, k, n) \sim 1,$$

причем $T(I, \mathscr{M}^n) \sim 1$ для почти всех ЗИП $I \in \mathscr{I}(n, k)$.

Смысл этой теоремы в том, что для почти всех задач включающего поиска ответ на почти все запросы пуст.

Теорема 4. *Если $n \to \infty$ и $\log_2 \log_2 n = o(\log_2 k(n))$ и $\log_2 k(n) \lesssim 2n/3$ при $n \to \infty$, то при $i = 2, 3$*

$$\log_2 \overline{T}(P_i, k, n) \sim (\log_2 3 - 1) \log_2 k.$$

Эта теорема показывает, что хотя согласно теореме 1 для худшей библиотеки перебор неизбежен, предлагаемые алгоритмы позволяют для случайно взятой библиотеки существенно уходить от перебора.

3. Доказательство теоремы 1

Поскольку все $P_i(V)$, $i = 1, 2, 3$, являются деревьями, далее часто вместо термина ИГ будем использовать термин дерево.

Следом вершины β в дереве D назовем поддерево дерева D, содержащее все ребра, которые исходят из вершин, принадлежащих пути, соединяющему корень дерева D с вершиной β.

Левым следом вершины β в дереве D назовем поддерево дерева D, которое состоит из принадлежащих следу вершины β ребер, которые в дереве D находятся не правее ребра, принадлежащего пути, соединяющему корень дерева D с вершиной β.

Библиотекой вершины β в дереве D назовем множество записей, приписанных листьям ветви, растущей из вершины β.

Вершина дерева называется проходной, если из нее исходит одно ребро.

Лемма 1. *Пусть V — произвольная библиотека, $D = P_i(V)$, $i = 2, 3$, тогда это дерево обладает следующими свойствами.*

(1) *Для любой вершины β дерева D пересечение любых двух нагрузочных функций, соответствующих ребрам, принадлежащих левому следу вершины β, всегда пусто.*

(2) *Пусть β — произвольная вершина дерева D. Пусть f_1 есть конъюнкция всех нагрузочных функций, соответствующих ребрам из левого следа вершины β. Пусть $f_2 = x_{i_1} \cdot \ldots \cdot x_{i_t}$ — функция фильтра вершины β. Пусть $f_1 \setminus f_2 = x_{i_{t+1}} \cdot \ldots \cdot x_{i_{t+r}}$. Тогда библиотека вершины β содержит все те и только те записи V, которые в разрядах i_j, $j = 1, 2, \ldots, t$, имеют 1, а в разрядах i_j, $j = t+1, t+2, \ldots, t+r$, имеют 0.*

(3) *Функции фильтров любых двух различных вершин в D различны.*

(4) *В дереве $P_2(V)$ каждая проходная вершина является полюсом.*

Доказательство. Свойства 1, 2 и 4 достаточно очевидны и следуют из определения $P_2(V)$ и $P_3(V)$, а свойство 3 является следствием свойства 2.

Заметим, что нагрузка ребер в $P_2(V)$ однозначно определяет нагрузку листьев. Поэтому далее мы часто не будем упоминать о нагрузке листьев.

Высотой вершины в дереве назовем длину пути, ведущего из корня в данную вершину.

Сложностью вершины β в ИГ называется вероятность множества запросов, проходящих в вершину β, или иными словами, сложность вершины β равна $\mathsf{P}(N_{\varphi_\beta})$.

Сложностью ребра ИГ назовем сложность начала ребра. Понятно, что сложность ИГ равна сумме сложностей его ребер.

Пусть

$$\mathscr{D}_i = \{P_i(V)\colon V \in \mathscr{W}_n^k\}, \qquad i = 1, 2, 3.$$

Лемма 2. *Для любой вершины, находящейся на высоте h в любом дереве $D \in \mathscr{D}_i$, $i = 2, 3$, ее сложность не более, чем 2^{-h}.*

Доказательство. Возьмем произвольное дерево $D \in \mathscr{D}_i$, $i = 2, 3$, и рассмотрим произвольную вершину β, находящуюся на высоте h в D. Поскольку у нас равномерная вероятностная мера, $\mathsf{P}(N_{\varphi_\beta}) = 2^{-n}|N_{\varphi_\beta}|$. Длина пути от корня до вершины β равна h, поэтому φ_β есть конъюнкция h элементарных монотонных конъюнкций, которые согласно свойству 1 из леммы 1 не пересекаются по переменным. Следовательно, $|N_{\varphi_\beta}|$ достигает минимума, когда каждый из h сомножителей φ_β является переменной, и тогда $|N_{\varphi_\beta}| = 2^{n-h}$ откуда и следует утверждение леммы.

Тем самым лемма доказана.

Пусть D — некоторое дерево, h — натуральное число. Через $r(D, h)$ обозначим число вершин дерева D, находящихся на высоте h.

Лемма 3. *Для любого дерева $D \in \mathscr{D}_i$, $i = 2, 3$,*

$$r(D, 1) \leqslant \binom{n}{1} + 1$$

и для любого $h \in \{2, \ldots, n\}$

$$r(D, h) \leqslant \binom{n}{h}.$$

Доказательство. Отметим, что согласно свойству 1 из леммы 1 в любом дереве $D \in \mathscr{D}_i$, $i = 2, 3$, не может быть вершин, имеющих высоту более n. Обозначим через $\mathscr{D}_0$ множество всех деревьев, нагрузка ребер которых берется из $\mathscr{K}^n$ таким образом, чтобы удовлетворялось свойство 1 из леммы 1. Пусть $D' \in \mathscr{D}_0$ — такое дерево, на котором достигается максимум числа вершин, находящихся на высоте h. Покажем, что в D' нагрузка любого ребра, исходящего из вершины, находящейся на высоте, меньшей h, есть переменная. Предположим, что это не так и имеется ребро, исходящее из вершины, находящейся на высоте, меньшей h, нагрузка которого содержит по крайней мере 2 переменные x_l и x_q. Заменим нагрузку данного ребра на x_l и в ветви, растущей из начала данного ребра, пройдем в самую правую вершину, находящуюся на высоте $h-1$, и выпустим из нее ребро с нагрузкой x_q. Понятно, что полученное дерево будет удовлетворять свойству 1 из леммы 1 и иметь большее число вершин, находящихся на высоте h, чем в D'. Получаем противоречие.

Таким образом, в дереве D' функции фильтров всех вершин, находящихся на высоте h, являются элементарными монотонными конъюнкциями длины h. Откуда, согласно свойству 3 из леммы 1 число вершин, находящихся на высоте h в дереве D', не больше, чем число элементарных монотонных конъюнкций длины h, то есть

$$r(D', h) \leqslant \binom{n}{h}.$$

Дополнительная единица при $h = 1$ появляется из-за ребра с тождественной единицей в качестве нагрузки. А поскольку любое дерево D, принадлежащее $\mathscr{D}_i$, $i = 2, 3$, принадлежит и $\mathscr{D}_0$, утверждение леммы доказано.

Вершину ИГ назовем внутренней, если она не является полюсом.

Лемма 4. *Пусть $D \in \mathscr{D}_2$, t_j — число листьев в дереве D, находящихся на высоте j, $j = 1, 2, \ldots, n$. Тогда для любого $h \in \{0, 1, \ldots, n-1\}$*

$$r(D, n-h) \leqslant \sum_{j=0}^{h} 2^{j-h} t_{n-j}.$$

Доказательство. Доказательство будем вести индукцией по h. В качестве базы индукции возьмем $h = 0$. Тогда

$$r(D, n-0) \leqslant \sum_{j=0}^{0} 2^{j-h} t_{n-j} = t_n.$$

Поскольку других вершин кроме листьев на высоте n быть не может, базис индукции доказан.

Проведем индуктивный переход. Пусть

$$r(D, n-h+1) \leqslant \sum_{j=0}^{h-1} 2^{j-h+1} t_{n-j}.$$

Так как в каждую вершину высоты $n-h+1$ ведет ребро из вершины высоты $n-h$, согласно свойству 4 из леммы 1 на каждую пару вершин высоты $n-h+1$ придется самое большее одна внутренняя вершина высоты $n-h$, то есть число внутренних вершин высоты $n-h$ не превышает

$$2^{-1} \sum_{j=0}^{h-1} 2^{j-h+1} t_{n-j} \leqslant \sum_{j=0}^{h-1} 2^{j-h} t_{n-j}.$$

Учитывая, что у нас есть еще t_{n-h} листьев, получаем, что

$$r(D, n-h) \leqslant \sum_{j=0}^{h-1} 2^{j-h} t_{n-j} + t_{n-h} = \sum_{j=0}^{h} 2^{j-h} t_{n-j}.$$

Тем самым лемма доказана.

Доказательство теоремы 1. Получим верхнюю оценку для функции $T(P_2, k, n)$.

Возьмем произвольную библиотеку $V \in \mathscr{W}_n^k$. Пусть t_i — число листьев в дереве $P_2(V)$, находящихся на высоте i, $i = 1, 2, \dots, n$.

Обозначим через $d_1(h)$ сумму сложностей ребер дерева $P_2(V)$, ведущих в вершины высоты не менее, чем h (напомним, что поскольку у нас все ребра предикатные, сложность ребра есть сложность начала ребра). Поскольку в вершину высоты i ведет ребро из вершины высоты $i-1$, используя 2 и 4, находим, что

$$\begin{aligned} d_1(h) &\leqslant \sum_{i=h}^{n} 2^{1-i} \sum_{j=0}^{n-i} 2^{j+i-n} t_{n-j} = 2 \sum_{i=h}^{n} \sum_{j=0}^{n-i} 2^{j-n} t_{n-j} \\ &= 2 \sum_{i=h}^{n} \sum_{j=i}^{n} t_j 2^{-j} = 2 \sum_{i=h}^{n} t_i 2^{-i}(i-h+1) \leqslant 2^{1-h} \sum_{i=h}^{n} t_i. \end{aligned}$$

Обозначим через $d_2(h)$ сумму сложностей ребер дерева $P_2(V)$, ведущих в вершины высоты, меньшей h. Тогда согласно леммам 2 и 3

$$d_2(h) \leqslant 1 + \sum_{i=1}^{h-1} 2^{1-i} \binom{n}{i} \sim 2^{2-r} \binom{n}{h-1}$$

при условии, что $h = o(n)$ при $n \to \infty$.

Если m — такая последовательность, что

$$\binom{n}{m-1} = o(k), \qquad k \lesssim 2\binom{n}{m}$$

при $n \to \infty$, то возьмем $h = m$ и тогда

$$T(P_2(V)) \leqslant d_1(m) + d_2(m) \lesssim 2^{2-m} \binom{n}{m-1} + 2^{1-m} \sum_{i=m}^{n} t_i \lesssim 2^{1-m} k.$$

Если m — такая последовательность, что

$$k \asymp \binom{n}{m}, \qquad 2\binom{n}{m} \lesssim k$$

при $n \to \infty$, то возьмем $h = m + 1$ и тогда

$$\begin{aligned} T(P_2(V)) &\leqslant d_1(m+1) + d_2(m+1) \\ &\lesssim 2^{1-m} \binom{n}{m} + 2^{-m} \sum_{i=m+1}^{n} t_i \lesssim 2^{-m} \left(k + 2\binom{n}{m}\right). \end{aligned}$$

Из этих соотношений в силу произвольности V получаем требуемую верхнюю оценку.

Верхняя оценка для $T(P_1, k, n)$ следует из соотношения

$$T(P_1, k, n) \leqslant T(P_2, k, n) + 1,$$

а неравенства

$$T(n, k, \mathscr{M}^n) \leqslant T(n, k, \mathscr{K}^n) \leqslant T(P_2, k, n)$$

очевидны.

Найдем нижнюю оценку. Согласно теореме 2 из [2] для любой ЗИП $I = \langle B^n, V, \succeq \rangle$ типа S_b

$$T(I, \mathscr{M}^n) \geqslant 2 \sum_{y \in V} \mathsf{P}(O(y, \succeq)) - r_0,$$

где r_0 — число записей веса 0 в библиотеке V.

Если m — такая последовательность, что при $n \to \infty$

$$\binom{n}{m-1} = o(k), \qquad k \leqslant \binom{n}{m},$$

то рассмотрим библиотеку V, состоящую только из записей, принадлежащих m-му слою B^n, и тогда для $i = 1, 2$

$$T(P_i, k, n) \geqslant T(n, k, \mathscr{M}^n) \geqslant T(\langle B^n, V, \succeq \rangle, \mathscr{M}^n) \geqslant 2^{1-m} k,$$

поскольку для каждой записи y из m-го слоя $\mathsf{P}(O(y, \succeq)) = 2^{-m}$.

Если m — такая последовательность, что при $n \to \infty$

$$k \asymp \binom{n}{m}, \qquad k \geqslant \binom{n}{m},$$

то рассмотрим библиотеку V, состоящую из всех записей принадлежащих m-му слою B^n, в которой остальные $k - \binom{n}{m}$ записей принадлежат $(m+1)$-му слою B^n, и тогда для $i = 1, 2$

$$T(P_i, k, n) \geqslant T(n, k, \mathscr{M}^n) \geqslant T(\langle B^n, V, \succeq \rangle, \mathscr{M}^n)$$
$$\geqslant 2^{1-m} \binom{n}{m} + 2^{-m} \left(k - \binom{n}{m} \right) = 2^{-m} \left(k + \binom{n}{m} \right).$$

Тем самым нижняя оценка и теорема 1 доказаны.

4. Доказательство теоремы 2

Найдем нижнюю оценку. В доказательстве теоремы 1 из [3] показано, что в условиях теоремы 2 существует такая задача включающего поиска $I' = \langle B^n, V', \succeq \rangle$, что $|V'| = k$ и

$$\inf_{U \in \mathscr{U}'(I', X^n)} T(U) \gtrsim 2^{2-m} k.$$

Следовательно,

$$T(P_3, k, n) \geqslant \max_{I \in \mathscr{I}(n,k)} \inf_{U \in \mathscr{U}'(I, X^n)} T(U) \geqslant \inf_{U \in \mathscr{U}'(I', X^n)} T(U) \gtrsim 2^{2-m} k.$$

Тем самым нижняя оценка доказана.

Найдем верхнюю оценку. Возьмем произвольную библиотеку $V \in \mathscr{W}_n^k$. Пусть t_i — число листьев в дереве $P_3(V)$, находящихся на высоте i, $i = 1, 2, \ldots, n$.

Нетрудно заметить, что число вершин высоты i в дереве $P_3(V)$ не более, чем $\sum_{j=i}^{n} t_j$. Теперь, оценивая число вершин высоты, не меньшей, чем $m(n)$, этим числом, а число вершин высоты, меньшей, чем $m(n)$, по лемме 3, и вспоминая, что согласно лемме 2 ребро, ведущее в вершину высоты i, будет иметь сложность, не большую, чем 2^{1-i}, получим, что

$$\begin{aligned} T(P_3(V)) &\leqslant \sum_{i=1}^{m-1} 2^{1-i}\binom{n}{i} + \sum_{i=m}^{n} 2^{1-i}\sum_{j=i}^{n} t_j \\ &\lesssim 2^{2-m}\binom{n}{m-1} + k\sum_{i=m}^{n} 2^{1-i} \\ &\lesssim 2^{2-m}\binom{n}{m-1} + k2^{2-m} \sim 2^{2-m}k. \end{aligned}$$

Отсюда в силу произвольности V вытекает справедливость верхней оценки.

Тем самым теорема 2 доказана.

5. Доказательство теоремы 3

Пусть $s(n) =]3\log_2 k(n)[$ и $\mathscr{W}(n,k,s)$ — множество всех библиотек из $\mathscr{W}_n^k$, которые содержат только записи веса, не меньшего, чем s. Тенью библиотеки V назовем множество

$$O(V,\succeq) = \bigcup_{y\in V} O(y,\succeq).$$

Нетрудно заметить, что для любой библиотеки $V \in \mathscr{W}(n,k,s)$

$$|O(V,\succeq)| \leqslant k2^{n-s}.$$

Возьмем произвольную библиотеку $V \in \mathscr{W}(n,k,s)$. По определению алгоритма P_1

$$T(P_1(V)) \leqslant 1 + 2^{-n}|O(V,\succeq)|l(P_2(V)) \leqslant 1 + 2^{-s}k(n)l(P_2(V)),$$

где $l(P_2(V))$ — число ребер дерева $P_2(V)$. Теперь воспользуемся следующим очевидным фактом: если дерево имеет m вершин, степень инцидентности которых не превышает 2, то в этом дереве не более $2m-2$ ребер. Отсюда, привлекая свойство 4 из леммы 1, находим, что

$$T(P_1(V)) \leqslant 1 + 2^{-s}2k \sim 1$$

при $n \to \infty$.

Легко заметить, что при $n \to \infty$ справедлива цепочка неравенств

$$A_{2^n}^k \gtrsim |\mathscr{W}(n,k,s)| = A_{\Sigma}\sum_{j=s}^{n}\binom{n}{j}^k = A^k_{2^n-\sum_{j=0}^{s-1}\binom{n}{j}} \gtrsim A^k_{2^n-s(n\cdot e/s)^s}.$$

Теперь, оценивая сложность библиотек, не принадлежащих $\mathscr{W}(n,k,s)$ через $2k$, находим, что при $n \to \infty$

$$\begin{aligned} \overline{T}(P_1,k,n) &\lesssim (|\mathscr{W}(n,k,s)|\cdot 1 + (A_{2^n}^k - |\mathscr{W}(n,k,s)|)2k)/A_{2^n}^k \\ &\lesssim 1 + 2k\left(1 - A^k_{2^n-s(ne/s)^s}/A^k_{2^n}\right) \sim 1, \end{aligned}$$

так как если положить $\log_2 k = n/\alpha_n$, где $\alpha_n \to \infty$ при $n \to \infty$, то при $n \to \infty$

$$k(1 - A^k_{2^n - s(ne/s)^s}/A^k_{2^n}) = k\left(1 - \binom{2^n - s(ne/s)^s}{k}\binom{2^n}{k}^{-1}\right)$$
$$\sim k\left(1 - \left(\frac{(2^n - 2^{\log_2 s + s\log_2(n/s) + s\log_2 e})e}{k}\right)^k \frac{k^k}{e^k 2^{nk}}\right)$$
$$\lesssim k(1 - (1 - 2^{2s\log_2(n/s) - n})^k)$$
$$= k\left(1 - \left(1 + \sum_{i=1}^{k}\binom{k}{i}(-2^{2s\log_2(n/s)-n})^i\right)\right)$$
$$= -k\sum_{i=1}^{k}\binom{k}{i}(-2^{2s\log_2(n/s)-n})^i \leqslant k\sum_{i=1}^{k}(k2^{2s\log_2(n/s)-n})^i$$
$$\sim k\sum_{i=1}^{k}(2^{\log_2 k + 6\log_2 k\cdot\log_2(n/(3\log_2 k)) - n})^i$$
$$= k\sum_{i=1}^{k}(2^{n/\alpha_n + (6n\log_2\alpha_n)/\alpha_n - n})^i \lesssim 2^{3n/\alpha_n + (6n\log_2\alpha_n)/\alpha_n - n} = o(1).$$

Учитывая, что

$$\overline{T}(P_1, k, n) \geqslant \overline{T}(n, k, \mathscr{M}^n) \geqslant 1,$$

получаем утверждение теоремы.

6. Доказательство теоремы 4

Исследуем функцию

$$L = (m, n, t, k) = \binom{t}{m}\binom{n-t}{k-m}\Big/\binom{n}{k},$$

где m, n, t, k — целые числа и $0 \leqslant m \leqslant k$, $0 < k < t$, $k < n - t$, $t < n$.

Лемма 5. *Функция $L(m, n, t, k)$ при фиксированных n, t, k относительно переменной m до точки $[tk/n]$ монотонно возрастает, а после точки $[tk/n] + 1$ монотонно убывает.*

Доказательство. Рассмотрим отношение

$$\frac{L(m+1, n, t, k)}{L(m, n, t, k)} = \frac{m!(t-m)!(k-m)!(n-t-k+m)!}{(m+1)!(t-m-1)!(k-m-1)!(n-t-k+m+1)!}$$
$$= \frac{(t-m)(k-m)}{(m+1)(n-t-k+m+1)}$$

и сравним это отношение с 1. Это равносильно сравнению $(t-m)(k-m)$ с $(m+1)(n-t-k+m+1)$, что в свою очередь равносильно сравнению $(tk-(n-t-k))/(n+2)$ с m. Так как

$$[tk/n] + 1 > tk/n > (tk - (n-t-k) - 1)/(n+2),$$

при $m \geqslant [tk/n]+1$
$$L(m+1,n,t,k) < L(m,n,t,k),$$
то есть функция монотонно убывает.

Так как
$$[tk/n]-1 < (tk-n)/n < (tk-n+t+k-1)/(n+2),$$
при $m \leqslant [tk/n]-1$
$$L(m+1,n,t,k) > L(m,n,t,k)$$
или, что то же самое, при $m \leqslant [tk/n]$
$$L(m,n,t,k) > L(m-1,n,t,k),$$
то есть функция монотонно возрастает.

Тем самым лемма доказана.

Лемма 6. *Пусть $m,k,n,t \to \infty$ и $m=o(t)$, $m=o(k)$, $k=o(n)$, $t=o(n)$, тогда*
$$L(m,n,t,k) \sim \sqrt{\frac{1}{2\pi m}}\left(\frac{tk}{mn}\right)^m\left(\frac{n-t}{n}\frac{k}{k-m}\right)^{k-m} \times \exp\left(-\frac{(mn-kt)^2}{2(t-m)(n-t+m-k)(n-k)}+\underline{O}\left(\frac{m^3}{t^2}+\frac{k^3}{n^2}\right)\right).$$

Доказательство. Так как
$$\binom{n}{k} \sim \sqrt{\frac{1}{2\pi k}}\left(\frac{n}{k}\right)^k\exp\left(k-\frac{k^2}{2(n-k)}+\underline{O}\left(\frac{k^3}{n^2}\right)\right)$$
при $k,n\to\infty$ и $k/n\to 0$, справедлива оценка
$$L(m,n,t,k)=\binom{t}{m}\binom{n-t}{k-m}\Big/\binom{n}{k} \sim \sqrt{\frac{k}{2\pi m(k-m)}}\left(\frac{t}{m}\right)^m\left(\frac{n-t}{k-m}\right)^{k-m}$$
$$\times\left(\frac{k}{n}\right)^k\exp\left(m-\frac{m^2}{2(t-m)}+\underline{O}\left(\frac{m^3}{t^2}\right)+k-m+\right.$$
$$\left.-\frac{(k-m)^2}{2(n-t-k+m)}+\underline{O}\left(\frac{k^3}{n^2}\right)-k+\frac{k^2}{2(n-k)}+\underline{O}\left(\frac{k^3}{n^2}\right)\right)$$
$$\sim\sqrt{\frac{1}{2\pi m}}\left(\frac{tk}{mn}\right)^m\left(\frac{n-t}{n}\frac{k}{k-m}\right)^{k-m}$$
$$\times\exp\left(-\frac{(mn-kt)^2}{2(t-m)(n-t-k+m)(n-k)}+\underline{O}\left(\frac{m^3}{t^2}+\frac{k^3}{n^2}\right)\right).$$

Тем самым лемма доказана.

Лемма 7. *Пусть $n,k,t,tk/n\to\infty$ и $k=o(n)$, $t=o(n)$, $\alpha=o(tk/n)$, тогда*
$$L([tk/n]\pm\alpha,n,t,k)\sim\sqrt{\frac{n}{2\pi tk}}\exp\left(-\frac{\alpha^2 n}{2tk}+\bar{o}\left(\frac{\alpha^2 n}{2tk}\right)+\underline{O}\left(\frac{k^3}{n^2}\right)\right).$$

Доказательство. Так как $tk/n \pm \alpha \sim tk/n = o(t)$ и $tk/n = o(k)$ при $n \to \infty$, мы находимся в условиях леммы 6, откуда следует, что

$$
\begin{aligned}
L([tk/n]-\alpha,n,t,k) &\sim \sqrt{\frac{1}{2\pi(tk/n-\alpha)}}\left(\frac{tk}{n(tk/n-\alpha)}\right)^{tk/n-\alpha}\\
&\quad\times\left(\frac{n-t}{n((n-t)k/n+\alpha)}\right)^{(n-t)k/n+\alpha}\\
&\quad\times\exp\left(-\frac{\alpha^2n^2}{2(t-tk/n)(n-t-k+tk/n)(n-k)}+\underline{O}\left(\frac{k^3}{n^2}\right)\right)\\
&\sim\sqrt{\frac{n}{2\pi tk}}\left(1+\frac{\alpha}{tk/n-\alpha}\right)^{tk/n-\alpha}\left(1-\frac{\alpha}{(n-t)k/n+\alpha}\right)^{(n-t)k/n+\alpha}\\
&\quad\times\exp\left(-\frac{\alpha^2}{2t}+\frac{\alpha^2}{t}(\underline{O}(k/n)+\underline{O}(t/n))+o(k^3/n^2)\right).
\end{aligned}
$$

Так как

$$
\begin{aligned}
\ln&\left(\left(1+\frac{\alpha}{tk/n-\alpha}\right)^{tk/n-\alpha}\left(1-\frac{\alpha}{(n-t)k/n+\alpha}\right)^{(n-t)k/n+\alpha}\right)\\
&=\left(\frac{tk}{n}-\alpha\right)\ln\left(1+\frac{\alpha}{tk/n-\alpha}\right)\\
&\quad+\left(\frac{(n-t)k}{n}+\alpha\right)\ln\left(1-\frac{\alpha}{(n-t)k/n+\alpha}\right)\\
&=-\left(\frac{tk}{n}-\alpha\right)\sum_{i=1}^{\infty}\frac{1}{i}\left(\frac{\alpha}{tk/n-\alpha}\right)^i\\
&\quad+\left(\frac{(n-t)k}{n}+\alpha\right)\sum_{i=1}^{\infty}\frac{(-1)^{i-1}}{i}\left(\frac{\alpha}{(n-t)k/n+\alpha}\right)^i\\
&=-\frac{\alpha^2k}{2(tk/n-\alpha)((n-t)k/n+\alpha)}\\
&\quad\times\sum_{i=3}^{\infty}\left(\frac{(-1)^{i-1}\alpha^i}{i((n-t)k/n+\alpha)^{i-1}}-\frac{\alpha^i}{i(tk/n-\alpha)^{i-1}}\right)\\
&=-\frac{\alpha^2n^2}{2tk(n-t)}+\underline{O}\left(\frac{\alpha^3}{(tk/n)^2}\right),
\end{aligned}
$$

то справедливо соотношение

$$
L([tk/n]-\alpha,n,t,k)\sim\sqrt{\frac{n}{2\pi tk}}\exp\left(-\frac{\alpha^2n}{2tk}+o\left(\frac{\alpha^2n}{2tk}\right)+\underline{O}\left(\frac{k^3}{n^2}\right)\right).
$$

Асимптотика функции $L([tk/n]+\alpha,n,t,k)$ находится аналогично и имеет точно такой же вид.

Тем самым лемма доказана.

Пусть V_n — последовательность библиотек и $s(n)$ — неубывающая последовательность натуральных чисел такие, что $|V_n|2^{-s(n)} \to \infty$ при $n \to \infty$. Скажем, что последовательность библиотек V_n является равномерно распределенной порядка $s(n)$,

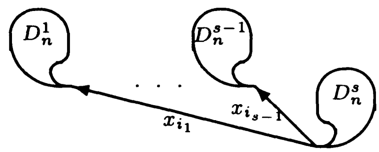

Рис. 1. Дерево равномерно распределенной библиотеки

если в любом $(n-s(n))$-мерном подкубе куба B^n число записей из V_n асимптотически равно $|V_n|2^{-s(n)}$ при $n \to \infty$.

Лемма 8. *Пусть V_n — равномерно распределенная порядка $s(n)$ последовательность библиотек, $t(n)$ — неубывающая последовательность натуральных чисел такая, что $t(n) < s(n)$ и $s(n) - t(n)$ — неубывающая последовательность, $B_{t(n)}$ — некоторая последовательность $(n-t(n))$-мерных подкубов куба B^n. Тогда $V_n \cap B_{t(n)}$ — равномерно распределенная порядка $s(n) - t(n)$ последовательность библиотек на подкубах $B_{t(n)}$.*

Доказательство. В $B_{t(n)}$ можно выделить $2^{n-t}/2^{n-s} = 2^s/2^t$ непересекающихся подкубов размерности $n - s(n)$. Число записей из V_n, содержащихся в каждом из этих подкубов, асимптотически равно $|V_n|\ 2^{-s(n)}$, поэтому при $n \to \infty$

$$|V_n \cap B_t| \sim |V_n|2^{-s}2^s/2^t = |V_n|2^{-t}.$$

Так как V_n — равномерно распределенная порядка $s(n)$ последовательность, в любом подкубе размерности $n - t(n) - (s(n) - t(n)) = n - s(n)$ куба B_t содержится асимптотически при $n \to \infty$

$$|V_n|2^{-s} = |V_n|2^{-t}2^{-(s-t)} \sim |V_n \cap B_t|2^{-(s-t)}$$

записей. Тем самым лемма доказана.

Библиотекой ветви дерева назовем библиотеку корня ветви.

Лемма 9. *Пусть V_n — равномерно распределенная порядка $s(n) \geqslant 2$ последовательность библиотек. Тогда $P_i(V_n)$, $i = 2, 3$, имеет вид, изображенный на рис. 1, и если обозначить через V_n^t библиотеку дерева D_n^t, $t = \overline{1, s}$, то*

$$\begin{aligned}
V_n^1 &= \{(y_1, \dots, y_n) \in V_n \colon y_{i_1} = 1\},\\
V_n^t &= \{(y_1, \dots, y_n) \in V_n \colon y_{i_j} = 0, j = 1, \dots, t-1, y_{i_t} = 1\}, \quad t = 2, \dots, s-1,\\
V_n^s &= \{(y_1, \dots, y_n) \in V_n \colon y_{i_j} = 0, j = 1, \dots, s-1\}.
\end{aligned}$$

Доказательство. Докажем сначала, что если $s \geqslant 2$, то $P_i(V_n)$, $i = 2, 3$, представимо в виде, изображенном на рисунке 2, где библиотека дерева D_n'' есть библиотека

$$V_n'' = \{(y_1, \dots, y_n) \in V_n \colon y_{i_1} = 0\}.$$

Из корня дерева $P_i(V_n)$, $i = 2, 3$, должно исходить хотя бы одно ребро. Возьмем первое ребро, исходящее из корня. Пусть из этого ребра растет дерево D_n^1. Оставшееся поддерево, растущее из корня, обозначим D_n''. Предположим, что нагрузочная

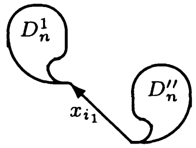

Рис. 2. Базис индукции

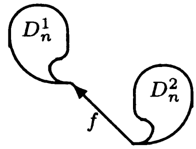

Рис. 3. Дерево равномерно распределенной библиотеки порядка 1

функция первого ребра равна $f = x_{j_1} \cdot \ldots \cdot x_{j_r}$. По определению алгоритма P_3 справедливо равенство $r = 1$. Покажем, что $r = 1$ и в случае алгоритма P_2. Предположим обратное, то есть что $r \geqslant 2$. Так как $s \geqslant 2$, подкуб $x_{j_1}\overline{x}_{j_2}$ содержит записи из V_n. Любая запись из этого подкуба в j_1 разряде содержит единицу, поэтому согласно свойству 2 из леммы 1 она не может принадлежать библиотеке ветви D_n''. С другой стороны, любая запись из подкуба $x_{j_1}\overline{x}_{j_2}$ в j_1 разряде содержит 0, и значит не может пройти через фильтр f и соответственно не может принадлежать библиотеке ветви D_n^1. Пришли к противоречию, то есть первому ребру, исходящему из корня приписана переменная. Обозначим ее x_{i_1}. Согласно свойству 2 из леммы 1 все записи из V_n, содержащие единицу в i_1-м разряде, пройдут по ребру x_{i_1}, то есть библиотека дерева D_n^1 есть V_n^1. Остается заметить, что ни одна запись из V_n'' не пройдет по ребру x_{i_1}. Тем самым мы доказали, что дерево $P_i(V_n)$, $i = 2, 3$, представимо в виде, изображенном на рисунке 2.

Заметим теперь, что согласно лемме 8 последовательность библиотек V_n'' на подкубе $\overline{x}_{i_1}$ является равномерно распределенной порядка $s - 1$, и если $s - 1 \geqslant 2$, то к V_n'' можно применить это же утверждение. И вообще, применив к V_n вышеприведенное утверждение $s - 1$ раз, мы получим утверждение леммы. Тем самым лемма 9 доказана.

Если дерево D — ИГ, то через $|D|$ обозначим мощность библиотеки дерева D.

Лемма 10. *Пусть V_n — равномерно распределенная порядка* 1 *последовательность библиотек. Тогда $P_i(V_n)$, $i = 2, 3$, имеет вид, изображенный на рис.* 3, *где при $n \to \infty$*

$$|D_n^1| \sim |D_n^2| \sim |V_n|/2.$$

Доказательство. Из корня дерева $P_i(V_n)$, $i = 2, 3$, должно исходить хотя бы одно ребро. Возьмем первое ребро, исходящее из корня. Пусть из этого ребра растет дерево D_n^1. Оставшееся поддерево, растущее из корня обозначим D_n^2. Предположим, что нагрузочная функция первого ребра равна $f = x_{j_1} \cdot \ldots \cdot x_{j_r}$. Так как V_n — равномерно распределенная порядка 1 последовательность, число записей из V_n, содержащих

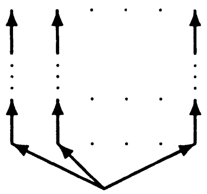

Рис. 4. Дерево простейшего вида

единицу в $(j-1)$-м разряде (то есть принадлежащих подкубу x_{j_1}) будет асимптотически равно $|V_n|/2$. А так как все записи с единицей в j_1-м разряде согласно свойству 2 из леммы 1 пройдут по первому ребру, ветвь, растущая из этого ребра, имеет мощность, асимптотически не меньшую, чем $|V_n|/2$. Но ни одна запись, содержащая 0 в j_1-м разряде, не пройдет через фильтр f, а число таких записей (как содержащихся в подкубе $\overline{x}_{j_1}$) асимптотически равно $|V_n|/2$, то есть дерево $P_i(V_n)$, $i=2,3$, в самом деле имеет вид, изображенный на рис. 3, где при $n\to\infty$

$$|D_n^1| \sim |D_n^2| \sim |V_n|/2.$$

Тем самым лемма 10 доказана.

Лемма 11. *Для любой библиотеки* $V \subset B^n$

$$T(P_i(V)) < 2|V|, \qquad i=2,3.$$

Доказательство. Рассмотрим дерево D, имеющее вид, изображенный на рисунке 4, в котором число ребер, исходящих из корня, равно $|V|$. Вершины, из которых не исходят ребра, являются листьями, и им приписаны взаимно однозначным образом записи из V. Нагрузка ребер дерева D осуществляется функциями из $X^n = \{x_1,\dots,x_n\}$ так, что конъюнкция переменных, приписанных каждой из $|V|$ цепочек, совпадает с характеристической функцией тени записи, соответствующей листу, в который данная цепочка ведет.

Сложность каждой цепочки не больше, чем $\sum_{i=0}^{n-1} 2^{-i} < 2$, поэтому $T(D) < 2|V|$. Теперь остается заметить, что дерево $P_i(V)$, $i=2,3$, можно получить некоторой склейкой ребер дерева D, а значит, $T(P_i(V)) < 2|V|$, $i=2,3$.

Тем самым лемма 11 доказана.

Лемма 12. *Пусть* V_n *— равномерно распределенная порядка* $s(n)$ *последовательность библиотек, где* $|V_n|2^{-s(n)}\to\infty$ *при* $n\to\infty$. *Тогда для* $i=2,3$ *при* $n\to\infty$

$$4(3/2)^{s(n)-1}-2 \leqslant T(P_i(V_n)) \lesssim 2|V_n|(3/4)^{s(n)}.$$

Доказательство. Доказательство будем вести индукцией по порядку последовательности библиотек. При $s(n)\equiv 1$ согласно лемме 10 дерево $P_i(V_n)$, $i=2,3$, имеет вид, изображенный на рисунке 3. Откуда следует, что

$$T(P_i(V_n)) = T(D_n^1)2^{-r} + T(D_n^2) + 1,$$

где r — длина элементарной конъюнкции f. Применяя леммы 10 и 11, находим, что

$$T(P_i(V_n)) \geqslant 1\cdot 2^{-r}+1+1 \geqslant 2, \qquad T(P_i(V_n)) \lesssim 2|D_n^1|2^{-r} + 2|D_n^2| + 1 \lesssim 3|V_n|/2,$$

при $n \to \infty$, то есть при $s \equiv 1$ утверждение леммы справедливо.

Проведем индуктивный переход. Пусть утверждение леммы справедливо для любой равномерно распределенной последовательности библиотек порядка, меньшего, чем $s(n)$. Согласно лемме 9 дерево $P_i(V_n)$, $i = 1, 2$, имеет вид изображенный на рисунке 1, и деревья D_n^t суть деревья для библиотек V_n^t, где V_n^t описаны в условии леммы 9, $t = 1, \dots, s$. Так как согласно лемме 8 для любого $t(n) \in \{1, \dots, s-1\}$ последовательность библиотек V_n^t является в своем подкубе $\overline{x}_{i_1} \cdot \ldots \cdot \overline{x}_{i_{t-1}} \cdot \overline{x}_{i_t}$ равномерно распределенной порядка $s(n) - t(n)$ и $|V_n^t| \sim |V_n| 2^{-t}$ при $n \to \infty$, а V_n^s — равномерно распределенная порядка 1 в подкубе $\overline{x}_{i_1} \cdot \ldots \cdot \overline{x}_{i_{s-1}}$ и $|V_n^s| \sim |V_n| 2^{1-s}$ при $n \to \infty$, согласно предположению индукции при $i = 1, 2$ и $n \to \infty$

$$T(P_i(V_n)) = s - 1 + T(D_n^s) + \frac{1}{2} \sum_{i=1}^{s-1} T(D_n^i) \geqslant$$
$$\geqslant s - 1 + 2 + \frac{1}{2} \sum_{i=1}^{s-1} \left(4(3/2)^{s-i-1} - 2\right) = 4(3/2)^{s-1} - 2,$$
$$T(P_i(V_n)) \lesssim s(n) - 1 + 3|V_n^s|/2 + \frac{1}{2} \sum_{i=1}^{s-1} 2|V_n^i|(3/4)^{s-i} \lesssim$$
$$\lesssim 3\, 2^{-s}|V_n| + |V_n|(3/4)^s \sum_{i=1}^{s-1} (2/3)^i = 2|V_n|(3/4)^{s(n)},$$

что соответствует предположению индукции.

Тем самым лемма 12 доказана.

Обозначим через $\mathscr{V}'(n, k, s, \alpha)$ — множество неупорядоченных библиотек мощности k таких, что в любом подкубе размерности $n - s$ содержится больше, чем $[k2^{-s}] - \alpha$, и меньше, чем $[k \ 2^{-s}] + \alpha$ записей.

Лемма 13. *Пусть $s(n) \to \infty$, $k(n)2^{-s(n)} \to \infty$, $\alpha(n) = o(k2^{-s})$, $k(n) = o(2^n)$ при $n \to \infty$. Тогда*

$$|\mathscr{V}'(n, k(n), s(n), \alpha(n))|$$
$$\gtrsim \binom{2^n}{k} \left(1 - \exp\left(2s \ln n + \ln k - \frac{\alpha^2 2^s}{2k} + o\left(\frac{\alpha^2 s}{k}\right) + \underline{O}\left(\frac{k^3}{4^n}\right)\right)\right).$$

Доказательство. Пусть $L(m) = L(m, 2^n, 2^{n-s}, k)$.

Выделим некоторый подкуб размерности $n - s(n)$. Число библиотек, которые в этом подкубе содержат либо не больше, чем $[k2^{-s}] - \alpha$, либо не меньшее, чем $[k2^{-s}] + \alpha$ записей равно

$$r(n) = \left(\sum_{m=0}^{[k2^{-s}]-\alpha} L(m) + \sum_{m=[k2^{-s}]+\alpha}^{k} L(m) \right) \binom{2^n}{k}.$$

Если выбросить все эти библиотеки из рассмотрения, то останутся $\binom{2^n}{k} - r$ библиотек, и все они в рассматриваемом подкубе будут содержать больше, чем $[k2^{-s}] - \alpha$, и меньшее, чем $[k2^{-s}] + \alpha$ записей.

Эту процедуру можно проделать для всех $2^s\binom{n}{s}$ подкубов размерности $n-s$. Откуда получим, что

$$|\mathscr{V}'(n,k(n),s(n),\alpha(n)| \geqslant \binom{2^n}{k} - 2^s r\binom{n}{s}.$$

Так как согласно лемме 5 $L(m)$ возрастает до точки $[k2^{-s}]$ и убывает после $[k2^{-s}]+1$ и так как выполняются условия леммы 7, то

$$r(n) \lesssim \binom{2^n}{k} L([k2^{-s}]-\alpha)k \sim \binom{2^n}{k}\sqrt{\frac{2^s}{2\pi k}}\exp\left(\ln k - \frac{\alpha^2 2^s}{2k} + o\left(\frac{\alpha^2 s}{k}\right) + \underline{O}\left(\frac{k^3}{4^n}\right)\right)$$

при $n\to\infty$. Учитывая, что $2^s\binom{n}{s} \leqslant n^{2s}$, находим, что при $n\to\infty$

$$|\mathscr{V}'(n,k(n),s(n),\alpha(n))| \gtrsim \binom{2^n}{k}\left(1-\exp\left(2s\ln n + \ln k - \frac{\alpha^2 2^s}{2k} + o\left(\frac{\alpha^2 s}{k}\right) + \underline{O}\left(\frac{k^3}{4^n}\right)\right)\right).$$

Тем самым лемма 13 доказана.

Доказательство теоремы 4. Пусть $\mathscr{W}'(n,k,s,\alpha)$ — множество упорядоченных библиотек мощности k таких, что в любом подкубе размерности $n-s$ содержится больше, чем $[k2^{-s}]-\alpha$, и меньше, чем $[k2^{-s}]+\alpha$ записей. Так как любую упорядоченную библиотеку можно получить некоторой перестановкой неупорядоченной библиотеки, справедливо равенство

$$|\mathscr{W}'(n,k,s,\alpha)| = k!|\mathscr{V}'(n,k,s,\alpha)|. \tag{1}$$

Выберем параметры следующим образом:

$$s(n) = \log_2 k(n) - (1+3\varepsilon)\log_2(\ln k(n)\ln n + k^3(n)/4^n),$$
$$\alpha(n) = (\ln k(n)\ln n + k^3(n)/4^n)^{1+2\varepsilon},$$

где ε — положительная постоянная такая, что

$$(\ln k(n)\ln n + k^3(n)/4^n)^{1+3\varepsilon} = o(k(n))$$

при $n\to\infty$. Тогда при $n\to\infty$

$$k\exp\left(2s\ln n + \ln k - \frac{\alpha^2 2^s}{2k} + o\left(\frac{\alpha^2 s}{k}\right) + \underline{O}\left(\frac{k^3}{4^n}\right)\right) \lesssim \exp\left(2\ln k + 2\log_2 k\ln n - \frac{1}{2}\left(\ln k\ln n + \frac{k^3}{4^n}\right)^{1+\varepsilon} + o\left(\left(\ln k\ln n + \frac{k^3}{4^n}\right)^{1+\varepsilon}\right) + \underline{O}\left(\frac{k^3}{4^n}\right)\right) = o(1). \tag{2}$$

Так как $\ln k\ln n + k^3/4^n = o(k)$, то при $n\to\infty$

$s(n)\to\infty$, $k\ 2^{-s} = (\ln k\ln n + k^3/4^n)^{1+3\varepsilon}\to\infty$, $\alpha\ 2^s/k = (\ln k\ln n + k^3/4^n)^{-\varepsilon}\to 0$, то есть мы находимся в условиях леммы 13, и любая последовательность библиотек $V_n \in \mathscr{W}'(n, k(n), s(n), \alpha(n))$ является равномерно распределенной порядка $s(n)$.

Теперь, оценивая сложность библиотек из $\mathscr{W}'(n,k,s,\alpha)$ согласно лемме 12, а библиотек из $\mathscr{W}_n^k \setminus \mathscr{W}'(n,k,s,\alpha)$ — согласно лемме 11 через $2k(n)$ и учитывая лемму 13 и соотношения (1), (2), находим, что для $i = 2, 3$ при $n \to \infty$

$$\begin{aligned}\overline{T}(P_i,k,n) &\lesssim (A_{2^n}^k)^{-1}(|\mathscr{W}'(n,k,s,\alpha)|2k(3/4)^s + (A_{2^n}^k - |\mathscr{W}'(n,k,s,\alpha)|)2k)\\ &\lesssim k\exp\left(\ln\frac{3}{4}(\log_2 k - \log_2(\ln k \ln n + k^3/4^n)^{1+3\varepsilon})\right)\\ &= \exp(\ln 2((\log_2 3 - 1)\log_2 k + (2-\log_2 3)\log_2(\ln k\ln n + k^3/4^n)^{1+3\varepsilon})). \quad (3)\end{aligned}$$

С другой стороны, оценивая сложность библиотек из $\mathscr{W}_n^k \setminus \mathscr{W}'(n,k,s,\alpha)$ через 1, для $i = 1, 2$ при $n \to \infty$ получим, что

$$\begin{aligned}&\overline{T}(P_i,k,n) \gtrsim (A_{2^n}^k)^{-1}(|\mathscr{W}'(n,k,s,\alpha)|(4(\tfrac{3}{2})^{s-1}-2)+(A_{2^n}^k - |\mathscr{W}'(n,k,s,\alpha)|)) \gtrsim 4(\tfrac{3}{2})^{s-1}\\ &\sim \exp(\ln 2((\log_2 3 - 1)\log_2 k - (\log_2 3 - 1)\log_2(\ln k \ln n + k^3/4^n)^{1+3\varepsilon} + \log_2(\tfrac{3}{8})). \quad (4)\end{aligned}$$

При заданных в условии теоремы ограничениях на k

$$\log_2(\ln k \ln n + k^3/4^n) = o(\log_2 k),$$

поэтому из (3) и (4) для $i = 1, 2$ при $n \to \infty$ следует, что

$$\log_2 \overline{T}(P_i,k,n) \sim (\log_2 3 - 1)\log_2 k.$$

Тем самым теорема 4 доказана.

Автор выражает благодарность В. Б. Кудрявцеву и А. С. Подколзину за поддержку в работе.

Список литературы

1. Селтон Г., *Автоматическая обработка, хранение и поиск информации*. Сов. радио, Москва, 1973.

2. Гасанов Э. Э., Нижняя оценка сложности информационных сетей для одного отношения частичного порядка. *Дискретная математика* (1996) **8**, №4, 108–122.

3. Гасанов Э. Э., Нижняя оценка сложности включающего поиска в классе древовидных схем. *Дискретная математика* (1998) **10**, № 1, 63–72.

4. Гасанов Э. Э., Косолапов А. В., К вопросу о древовидности оптимальных информационных сетей включающего поиска. *Интеллектуальные системы* (1998) **3**, № 1–2, 167–192.

5. Гасанов Э. Э., Об одномерной задаче интервального поиска. *Дискретная математика* (1995) **7**, №2, 40–60.

6. Гаврилов Г. П., Сапоженко А. А., *Задачи и упражнения по курсу дискретной математики*. Наука, Москва, 1992.

7. Гасанов Э. Э., Об одной математической модели информационного поиска. *Дискретная математика* (1991) **3**, №2, 69–76.

Статья поступила 13.09.1999.
Переработанный вариант поступил 10.04.2000.