

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИМЕНИ М. В. ЛОМОНОСОВА

МЕХАНИКО–МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ

На правах рукописи  
УДК 519.71

Пивоваров Александр Павлович

НЕПЕРЕЧИСЛИТЕЛЬНЫЕ ЗАДАЧИ  
ИНФОРМАЦИОННОГО ПОИСКА

01.01.09 – дискретная математика и математическая кибернетика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени  
кандидата физико-математических наук



МОСКВА — 2012

Работа выполнена на кафедре Математической теории интеллектуальных систем Механико-математического факультета Московского государственного университета имени М.В. Ломоносова.

Научный руководитель — доктор физико-математических наук, профессор  
**Гасанов Эльяр Эльдарович**

Официальные оппоненты:

**Ложкин Сергей Андреевич**

доктор физико-математических наук, профессор  
факультет Вычислительной Математики  
и Кибернетики, МГУ имени М. В. Ломоносова

**Майлыбаева Гульнара Абаевна**

кандидат физико-математических наук  
Каргилл Энтерпрайзис Инк.  
аналитик по инфраструктурной безопасности

Ведущая организация — Московский энергетический институт  
(Национальный исследовательский университет)

Защита диссертации состоится 8 июня 2012 г. в 16 ч. 45 м. на заседании диссертационного совета Д.501.001.84 при Московском государственном университете имени М.В.Ломоносова по адресу: Российская Федерация, 119991, Москва, ГСП-1, Ленинские горы, д.1, Московский государственный университет имени М.В. Ломоносова, Механико-математический факультет, аудитория 14-08.

С диссертацией можно ознакомиться в библиотеке Механико-математического факультета МГУ имени М.В. Ломоносова (Главное здание, 14 этаж).

Автореферат разослан 25 апреля 2012 г.

Ученый секретарь диссертационного  
совета Д.501.001.84 при МГУ  
доктор физико-математических наук,  
профессор



А.О. Иванов

# Общая характеристика работы

**Актуальность темы.** Диссертация посвящена исследованию задач информационного поиска и принадлежит важнейшему разделу дискретной математики и математической кибернетики — теории управляющих систем. Задачи информационного поиска (ЗИП) — один из наиболее часто встречающихся на практике видов задач. При написании практически любой программы для ЭВМ требуется решать те или иные задачи поиска. Постоянная необходимость решения таких задач привела к появлению множества специализированных структур данных и алгоритмов поиска. В книге Дональда Кнута “Искусство программирования для ЭВМ”<sup>1</sup>, рассчитанной, в том числе, на инженеров, собраны некоторые из наиболее часто используемых на практике алгоритмов поиска и структур данных.

ЗИП могут иметь самую разную природу. Распространённый вид ЗИП — геометрические задачи информационного поиска. В таких задачах в качестве объектов, хранящихся в базах данных, а также в качестве запросов выступают геометрические фигуры. В качестве примера таких задач можно привести задачу о метрической близости<sup>2</sup>, задачу о доминировании<sup>3</sup>, задачу интервального поиска<sup>4,5</sup>, задачу двумерного интервального поиска с фиксированной стороной<sup>6</sup>. В работе Шеймоса и Препараты<sup>7</sup> собрано воедино множество разнообразных геометрических задач информационного поиска и подходов к их решению. Для задач информационного поиска, в которых возникает потребность в последовательном поиске одного и того же ключа в нескольких линейно упорядоченных множествах, оказывается полезным использование техники частичного каскадирования<sup>8</sup>, позволяющей снизить время поиска за счёт увеличения объёма требуемой памяти.

Важным параметром задач информационного поиска является возможность добавлять и удалять элементы из базы данных. По этому параметру задачи можно разделить на статические и динамические. Динамическим

---

<sup>1</sup>Кнут Д. Искусство программирования для ЭВМ. Т. 3. Сортировка и поиск. — М.: Мир, 1978, 846 с.

<sup>2</sup>Gasanov E. E. On Functional Complexity of Two-dimensional Manhattan Metrics Closeness Problem, *Emerging Database Research In East Europe. Proceedings of the pre-conference workshop of VLDB 2003*, 51-56.

<sup>3</sup>JaJa J., Mortensen C. W., Shi Q. Space-Efficient and Fast Algorithms for Multi-dimensional Dominance Reporting and Counting, *Proc. ISAAC (2004)*, 558-568.

<sup>4</sup>Chazelle B. Lower Bounds for Orthogonal Range Searching: II. The Arithmetic Model. *Journal of the ACM (1990)*, 200-212.

<sup>5</sup>Willard D. E. Applications of the fusion tree method to computational geometry and searching. In *Proceedings of the 3rd Annual ACM - SIAM Symposium on Discrete Algorithms (1992)*, 286-295.

<sup>6</sup>McCreight E. M. Priority search trees. *SIAM Journal of Computing (1985)*, **14(2)**, 257-276.

<sup>7</sup>Препарата Ф., Шеймос М. Вычислительная Геометрия: Введение, Москва «Мир», 1989, 478 с.

<sup>8</sup>Chazelle B., Guibas L. J. Fractional cascading: I. A Data Structuring Technique, *Algorithmica 1 (1986)*, 133-162.

задачам информационного поиска посвящены работы Овермарса<sup>9</sup>, Мортенсена<sup>10</sup> и других авторов.

В монографии Гасанова и Кудрявцева<sup>11</sup> рассматриваются статические задачи поиска, в которых в ответ на запрос надо перечислить элементы базы данных, удовлетворяющие запросу — так называемые перечислительные задачи поиска. Для формального описания и оценки алгоритмов поиска в вышеупомянутой работе вводится понятие информационного графа (ИГ), использование которого показывает свою эффективность для оценки таких характеристик алгоритма поиска, как время поиска в среднем, время поиска в худшем случае и затраты памяти, необходимые для реализации алгоритма.

В диссертации автором исследуются другие варианты ЗИП, а именно ЗИП в смысле поиска представителя и вычислительные ЗИП (ВЗИП). Поиск представителя отличается от перечислительной ЗИП тем, что в такой задаче не требуется перечислять все записи базы данных, удовлетворяющие запросу, а требуется найти хотя бы одну такую запись, которая и называется представителем. Решение ВЗИП подразумевает вычисление некоторой интегральной характеристики множества записей БД, удовлетворяющих запросу. Это может быть, например, количество таких записей. Также возможен вариант, когда каждой записи БД приписан некоторый ранг и требуется вычислить, к примеру, максимальный ранг записей, удовлетворяющих запросу.

Неперечислительные задачи поиска возникают в прикладных задачах и требуют особого рассмотрения, так как алгоритмы решения соответствующих им перечислительных задач могут оказаться для них крайне неэффективными. Для проведения такого рода исследований требуются специальные модели для неперечислительных ЗИП и алгоритмов их решения. В диссертации вводятся такие модели для задач поиска представителя и вычислительных задач информационного поиска, и приводятся эффективные алгоритмы решения таких задач.

**Цель работы.** В диссертации ставится задача разработки моделей для неперечислительных задач информационного поиска, таких как задачи поиска представителя и вычислительные задачи информационного поиска.

Целью работы является исследование неперечислительных модификаций некоторых известных задач информационного поиска в рамках пред-

---

<sup>9</sup>Overmars M. H. The design of dynamic data structures, Ph. D. thesis, University of Utrecht, The Netherlands, 1983.

<sup>10</sup>Mortensen C. W. Fully Dynamic Orthogonal Range Reporting on RAM. *SIAM Journal of Computing* (2006), **35(6)**, 1494-1525.

<sup>11</sup>Гасанов Э. Э., Кудрявцев В. Б. *Теория хранения и поиска информации*. ФИЗМАТЛИТ, Москва, 2002, 288 с.

ложенных моделей. Полученные результаты представляют не только самостоятельный интерес, но также подтверждают эффективность используемых моделей информационного поиска.

Ставится задача оценить явно объём структур, получаемых с помощью техники частичного каскадирования в зависимости от различных параметров.

Для некоторых геометрических вычислительных задач, в частности, для задачи суммирования по полугрупповой операции в двумерной задаче о доминировании ставится цель получить верхнюю оценку для функции зависимости времени поиска в худшем случае от объёма требуемой памяти.

**Научная новизна.** В диссертации проводится исследование именно неперечисляемых задач информационного поиска и применимости к ним информационно-графовой модели данных. Ранее в рамках этой модели широко исследовались только перечислительные задачи информационного поиска, поэтому возможность её использования для задач поиска представителя и вычислительных задач информационного поиска, вообще говоря, не очевидна.

В работе получены следующие основные результаты.

1. Разработаны модели для неперечисляемых задач информационного поиска, таких как задачи поиска представителя и вычислительные задачи информационного поиска. В понятии информационного графа произведено отделение базисной части, так называемого информационного графа общего вида, отвечающей за процедуру обхода графа, и для этих графов предложен один общий метод получения мощностных нижних оценок. Впервые автоматная идеология была внедрена в информационно-графовую модель данных и была показана эффективность ее использования для моделирования вычислительных задач поиска.

2. Для задачи поиска представителя в задаче о метрической близости предложен алгоритм поиска с линейным от размера базы данных объемом памяти, логарифмическим временем поиска в худшем случае и константным временем поиска в среднем.

3. Развита техника частичного каскадирования путем введения варьируемого параметра, отвечающего за максимальную ширину пробелов между мостами, соединяющими соседние каталоги, и впервые произведена оценка размеров получаемых структур данных.

4. Для суммирования по полугрупповой операции в двумерной задаче о доминировании и в двумерной задаче интервального поиска с фиксированной стороной разработано несколько параметризованных семейств алгоритмов поиска, с помощью которых получены рекордные верхние оценки для функции зависимости времени поиска в худшем случае от объёма тре-

буемой памяти. Причем впервые такие оценки имеют вид асимптотических (и обычных) неравенств, а не оценок роста сложности по порядку.

**Основные методы исследования.** В работе использованы методы теории информационного поиска, вычислительной геометрии и теории сложности управляющих систем.

**Теоретическая и практическая ценность работы.** Работа имеет теоретический характер. Введённые модели могут использоваться для описания алгоритмов решения прикладных задач информационного поиска с целью последующего сравнения их характеристик.

Полученные в работе алгоритмы для решения задач поиска представителя в двумерной задаче о доминировании и двумерной задаче о метрической близости, а также суммирования по полугрупповой операции в двумерной задаче о доминировании и двумерной задаче интервального поиска с фиксированной стороной могут быть использованы в прикладных задачах.

**Апробация работы.** Результаты настоящей работы докладывались

- на научно-исследовательском семинаре кафедры Математической теории интеллектуальных систем механико-математического факультета МГУ «Теория автоматов» под руководством академика В. Б. Кудрявцева, 2009–2011 гг. (неоднократно);
- на семинаре механико-математического факультета МГУ «Вопросы сложности алгоритмов поиска» под руководством профессора Э. Э. Гасанова, 2008–2011 гг. (неоднократно);
- на семинаре механико-математического факультета МГУ «Кибернетика и информатика» под руководством академика В. Б. Кудрявцева, 2011 г.;
- на семинаре факультета ВМиК МГУ «Дискретная математика и математическая кибернетика» под руководством проф., д.ф-м.н. В.Б.Алексеева, проф., д.ф-м.н. А.А.Сапоженко, проф., д.ф-м.н. С.А.Ложкина, 2012 г.;
- на конференции «Ломоносовские чтения» (2008, 2011, 2012 гг.);
- на Международной конференции «Современные проблемы математики, механики и их приложений», посвященной 70-летию ректора МГУ академика В.А.Садовниченко (2009 г.);
- на X Международном семинаре «Дискретная математика и ее приложения» (2010 г.);

- на X Международной конференции «Интеллектуальные системы и компьютерные науки» (2011 г.);
- на Международной конференции студентов, аспирантов и молодых ученых «Ломоносов-2009», «Ломоносов-2010» (2009, 2010 гг.).

**Публикации по теме диссертации.** Основные результаты диссертации опубликованы в пяти статьях [1]–[5] и материалах конференций [6]–[10], список которых приведен в конце автореферата.

**Структура и объем диссертации.** Диссертация состоит из введения и трех глав. Объем диссертации 103 стр. Список литературы содержит 23 наименования.

## Краткое содержание работы

Первая глава диссертации посвящена модификации и исследованию информационно-графовой модели данных.

В разделе 1.1 даётся определение информационного графа общего вида (ИГОВ), используемого далее как основа для определения информационных графов (ИГ), вычислительных информационных графов (ВИГ) и упрощённых вычислительных информационных графов (УВИГ). Для ИГОВ доказывается теорема 1, предоставляющая один способ получения мощностных нижних оценок для ИГОВ.

### Определение информационного графа общего вида

Определим понятие информационного графа общего вида (ИГОВ). Для этого нам потребуются следующие множества: множество запросов  $X$ , множество  $F$  одноместных предикатов, заданных на множестве  $X$ , множество  $G$  одноместных переключателей, заданных на множестве  $X$  (переключатели — это функции, область значений которых является начальным отрезком натурального ряда).

Пару  $\mathcal{F} = \langle F, G \rangle$  будем называть *базовым множеством*.

### Определение ИГОВ с точки зрения его структуры.

Пусть нам дан ориентированный граф с выделенной вершиной, которую мы назовём *корнем*.

Выделим в графе некоторые вершины и назовём их *точками переключения* (корень может быть точкой переключения).

Если  $\beta$  — вершина графа, то через  $\psi_\beta$  обозначим *полустепень исхода* вершины  $\beta$ .

Каждой точке переключения  $\beta$  сопоставим некий символ из  $G$ . Это соответствие назовём *нагрузкой точек переключения*.

Для каждой точки переключения  $\beta$  рёбрам, из неё исходящим, поставим во взаимно однозначное соответствие числа из множества  $\{\overline{1}, \overline{\psi_\beta}\}$ . Эти ребра назовём *переключательными*, а это соответствие — *нагрузкой переключательных рёбер*.

Ребра, не являющиеся переключательными, назовем *предикатными*.

Каждому предикатному ребру сети сопоставим некоторый символ из множества  $F$ . Это соответствие назовем *нагрузкой предикатных ребер*.

Полученный нагруженный граф назовем *информационным графом общего вида* над базовым множеством  $\mathcal{F} = \langle F, G \rangle$ .

### **Определение функционирования ИГОВ.**

Скажем, что предикатное ребро проводит запрос  $x \in X$ , если предикат, приписанный этому ребру, принимает значение 1 на запросе  $x$ ; переключательное ребро, которому приписан номер  $n$ , проводит запрос  $x \in X$ , если переключатель, приписанный началу этого ребра, принимает значение  $n$  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  $\beta$ , если существует ориентированная цепочка, ведущая из корня в вершину  $\beta$ , которая проводит запрос  $x$ .

Результатом функционирования ИГОВ на запросе  $x \in X$  считается набор вершин, в которые проходит запрос  $x$ .

Понятие ИГОВ определено.

Пусть  $\beta$  — некоторая вершина ИГОВ. Тогда обозначим  $\varphi_\beta(x)$  предикат на  $X$ , такой что  $\varphi_\beta(x) = 1$  для тех и только тех запросов  $x$ , которые проходят в вершину  $\beta$ . Предикат  $\varphi_\beta(x)$  называется *функцией фильтра* вершины  $\beta$ .

Переключатель  $g$ , приписанный переключательной вершине  $\beta$  считаем *вычисленным* на запросе  $x$ , если запрос  $x$  проходит в вершину  $\beta$ . Аналогично предикат  $f$ , приписанный некоторому ребру, выходящему из предикатной вершины  $\beta$ , считаем *вычисленным* на запросе  $x$ , если запрос  $x$  проходит в вершину  $\beta$ .

*Сложностью ИГОВ  $U$  на запросе  $x$*  назовем число  $T(U, x)$ , равное сумме числа переключателей и предикатов, вычисленных в процессе обработки запроса  $x$ . Эту величину также называют *временем работы  $U$  на запросе  $x$* . *Верхней сложностью* (или *временем работы в худшем случае*)



называют величину  $\hat{T}(U) = \max_{x \in X} T(U, x)$ . В случае, когда на  $X$  введено вероятностное пространство, а переключатели и предикаты являются измеримыми функциями, с помощью действий, дословно совпадающих с описанными в монографии Гасанова и Кудрявцева, можно показать, что  $T(U, x)$  является случайной величиной относительно  $x$  и можно рассматривать  $T(U) = \mathbf{M}_x T(U, x)$  — сложность ИГОВ  $U$  в среднем (или время работы  $U$  в среднем).

Объёмом  $Q(U)$  ИГОВ  $U$  назовем число ребер в ИГОВ  $U$ .

Пусть  $X$  — некоторое множество запросов,  $Z$  — множество, которое мы будем называть множеством ответов и задана функция  $A: X \rightarrow Z$ . Нам будет интересно, когда область значений функции  $A$  конечна.

Пусть дан ИГОВ  $U$ . Обозначим множество его вершин как  $W$ . Для любого запроса  $x \in X$  обозначим множество достижимых вершин  $U$  на этом запросе как  $W(x) = \{\beta \in W: \varphi_\beta(x) = 1\}$ . Будем говорить, что  $U$  является допустимым для функции  $A: X \rightarrow Z$ , если существует такая функция  $B: 2^W \rightarrow Z$ , что для любого  $x \in X$  выполнено  $B(W(x)) = A(x)$ .

**Теорема 1.** Пусть  $X$  — множество запросов,  $Z$  — множество ответов и задана функция  $A: X \rightarrow Z$ . Пусть также задано некоторое базовое множество  $\mathcal{F} = \langle F, G \rangle$  и  $U$  — некоторый ИГОВ над  $\mathcal{F}$ , допустимый для функции  $A$ .

1. Тогда для любого непустого множества  $X_0 \subseteq X$  выполнено:  $\max_{x \in X_0} T(U, x) \geq \lceil \log_m |A(X_0)| \rceil$ , где  $m$  — максимальная полустепень исхода среди переключательных вершин  $U$  (но не менее 2), либо 2, если таких вершин нет.
2. Пусть относительно  $X$  задано вероятностное пространство и функция  $A$  и  $\mathcal{F}$  измеримы относительно этого пространства,  $A(X) = \{z_1, \dots, z_t\}$  и  $p_i = P(A^{-1}(z_i))$ , причём  $p_1 \geq \dots \geq p_t$ . В этом случае имеют место следующие оценки на сложность в среднем:

$$T(U) \geq \max \left( \sum_{i=1}^{|A(X)|} u_i p_i, \max_{1 \leq i \leq |A(X)|} i \cdot p_i \cdot \left( \lceil \log_m i \rceil - \frac{m}{m-1} \right) \right),$$

где  $u_1 = 1$ , а при  $i \geq 2$  выполнено  $u_i = \lceil \log_m i \rceil$ .

В разделе 1.2 формализуются задача поиска представителя и понятие решения для таких задач. В дальнейшем эта формализация используется для формулировки результатов в главе 2, посвящённой как раз задачам поиска представителя.

С точки зрения определения, задача поиска представителя не отличается от определения задачи информационного поиска в перечислительном смысле, вводимого в монографии Гасанова и Кудрявцева. Для полноты описания, мы повторим здесь это определение.

Пусть  $X$  — множество запросов,  $Y$  — множество записей,  $\rho$  — бинарное отношение на  $X \times Y$  называемое *отношением поиска*. Тройку  $S = \langle X, Y, \rho \rangle$  будем называть *типом задач информационного поиска*. Тройка  $I = \langle X, V, \rho \rangle$ , где  $V$  — конечное подмножество  $Y$  (называемое в дальнейшем библиотекой) называется *задачей информационного поиска* (ЗИП) типа  $S$  (пишем  $I \in S$ ). Пусть задана некоторая ЗИП  $I = \langle X, V, \rho \rangle$ . Введём функцию  $\mathcal{J}_I(x): X \rightarrow 2^V$ , заданную соотношением  $\mathcal{J}_I(x) = \{y \in V \mid x \rho y\}$ . Эту функцию будем называть *функцией ответа на запрос для задачи  $I$* .

Описывать алгоритмы решения задач поиска представителя предлагается также с помощью обычных информационных графов (ИГ). Здесь мы приведём определение понятия ИГ, однако сделаем это с использованием информационных графов общего вида.

Пусть  $F$  — некоторое множество предикатов, определённых на  $X$ , а  $G$  — некоторое множество переключателей, определённых на  $X$ . Будем называть пару  $\mathcal{F} = \langle F, G \rangle$  *базовым множеством для ИГ*.

Определим ИГ над базовым множеством  $\mathcal{F} = \langle F, G \rangle$  следующим образом. Будем называть *информационным графом (ИГ)* такие графы  $U$ , что  $U$ , во-первых, является ИГОВ над базовым множеством  $\langle F, G \rangle$ , во-вторых, выделено некоторое множество вершин  $U$ , называемых *листьями* и каждому листу приписана некоторая запись из  $Y$ . Объём и сложность в среднем и худшем случаях для ИГ считаются так же, как и для ИГОВ.

Определим функционирование ИГ. Будем считать ответом графа  $U$  на запрос  $x$  множество записей  $\mathcal{J}_U(x)$ , состоящее из тех и только тех записей  $y \in Y$ , что в  $U$  есть лист, которому приписана запись  $y$  и при этом запрос  $x$  проходит в этот лист.

При этом для перечислительных задач требовалось равенство  $\mathcal{J}_I(x) = \mathcal{J}_U(x)$  для всех  $x \in X$ . Если же рассматривать задачу  $I$  как задачу поиска представителя, то определение графа решающего задачу меняется: *граф  $U$  решает задачу поиска представителя для  $I$* , если выполнены следующие условия:  $\mathcal{J}_U(x) \subseteq \mathcal{J}_I(x)$  и  $\mathcal{J}_I(x) \neq \emptyset \Rightarrow \mathcal{J}_U(x) \neq \emptyset$ .

Раздел 1.3 содержит определение вычислительной задачи информационного поиска (ВЗИП), определение УВИГ и ВИГ. Для модельной задачи одномерного интервального поиска в смысле подсчёта доказываются теоремы 2 и 3, с помощью которых обосновывается необходимость введения ВИГ.

## Формализация вычислительных задач информационного поиска

Дадим строго определение вычислительной задаче информационного поиска (ВЗИП). Пусть  $X$  — множество запросов,  $Y$  — множество записей,  $\rho$  — бинарное отношение на  $X \times Y$  называемое *отношением поиска*. Пусть также задано множество  $Z$ , которое мы будем называть *множеством ответов*. Наконец, задана функция  $\xi : 2^Y \rightarrow Z$  — *функция ответа*, определённая на  $2^Y$  — множестве подмножеств  $Y$  (достаточно, чтобы  $\xi$  была определена для всех конечных подмножеств  $Y$ ). Будем называть пятерку  $S = \langle X, Y, Z, \rho, \xi \rangle$  *типом вычислительных задач информационного поиска*. Пятерка  $I = \langle X, V, Z, \rho, \xi \rangle$ , где  $V$  — конечное подмножество  $Y$  (называемое в дальнейшем библиотекой) называется *вычислительной задачей информационного поиска (ВЗИП) типа  $S$*  (пишем  $I \in S$ ). Фактически задача  $I = \langle X, V, Z, \rho, \xi \rangle$  состоит в том, чтобы для заданного запроса  $x \in X$  найти такой ответ  $z \in Z$ , что  $z = \xi(\{y \in V : x \rho y\})$ . Будем обозначать этот ответ как  $z_I(x) = \xi(\{y \in V : x \rho y\})$  и называть *правильным ответом для ВЗИП  $I$  на запрос  $x$* .

## Определение упрощённого вычислительного информационного графа (УВИГ).

Пусть  $F$  — некоторое множество предикатов, определённых на  $X$ , а  $G$  — некоторое множество переключателей, определённых на  $X$ . Пусть  $Z$  — некоторое множество. Будем называть тройку  $\mathcal{F} = \langle F, G, Z \rangle$  *базовым множеством для УВИГ*.

Определим УВИГ над базовым множеством  $\mathcal{F} = \langle F, G, Z \rangle$  следующим образом. Будем называть *упрощённым вычислительным информационным графом (УВИГ)* такие графы  $U$ , что  $U$ , во-первых, является ИГОВ над базовым множеством  $\langle F, G \rangle$ , во-вторых, выделено некоторое множество вершин  $U$ , называемых *листьями* и каждому листу приписан некоторый элемент множества  $Z$ . Объём и сложность в среднем и худшем случаях для УВИГ считаются так же, как и для ИГОВ.

Определим функционирование УВИГ. Элемент  $z \in Z$ , приписанный листу  $\alpha$  некоторого УВИГ  $U$  будем считать ответом  $U$  на запрос  $x \in X$ , если запрос  $x$  проходит в лист  $\alpha$ . Этот ответ  $z$  обозначим как  $z_U(x)$  и будем называть *функцией ответа УВИГ  $U$* . В случае, если запрос  $x$  проходит в два листа  $\alpha_1$  и  $\alpha_2$ , которым приписаны ответы  $z_1$  и  $z_2$  соответственно и при этом  $z_1 \neq z_2$ , то считаем, что  $z_U(x)$  не определено. Также  $z_U(x)$  не определено в случае, если  $x$  не проходит ни в один лист УВИГ  $U$ .

Будем говорить, что некоторый УВИГ  $U$  над базовым множеством  $\mathcal{F} = \langle F, G, Z \rangle$  *решает ВЗИП  $I = \langle X, V, Z, \rho, \xi \rangle$* , если для любого запроса  $x \in X$

функция ответа УВИГ  $z_U(x)$  определена и равна значению  $z_I(x)$ .

### Определение вычислительного информационного графа (ВИГ).

Пусть, как и ранее,  $F$  — некоторое множество предикатов, определённых на  $X$ , а  $G$  — некоторое множество переключателей, определённых на  $X$ . Пусть  $M$  — некоторое множество, которое мы будем называть *множеством состояний ВИГ* и  $m_0$  — некоторый выделенный элемент  $M$ , который мы будем называть *начальным состоянием*. Пусть дополнительно определено множество  $H$  *функций изменения состояния* вида  $h : M \rightarrow M$ , такое, что любые две функции из  $H$  коммутируют относительно операции суперпозиции. Пусть, наконец, определено *множество ответов*  $Z$  и *функцию вычисления ответа по состоянию*  $\sigma : M \rightarrow Z$ . Будем называть шестерку  $\mathcal{F} = \langle F, G, H, M, m_0, \sigma \rangle$  *базовым множеством для ВИГ*.

Определим ВИГ над базовым множеством  $\mathcal{F} = \langle F, G, H, M, m_0, \sigma \rangle$  следующим образом. Будем называть *вычислительным информационным графом (ВИГ)* такие графы  $U$ , что  $U$ , во-первых, является ИГОВ над базовым множеством  $\langle F, G \rangle$ , во-вторых, выделено некоторое множество вершин  $U$ , называемых *листьями* и каждому листу приписан некоторый элемент множества  $H$ . Объём и сложность в среднем и худшем случаях для УВИГ считаются так же, как и для ИГОВ.

Определим функционирование ВИГ. Пусть есть некоторый ВИГ  $U$  над базовым множеством  $\mathcal{F} = \langle F, G, H, M, m_0, \sigma \rangle$ . Рассмотрим запрос  $x \in X$ . Пусть  $\alpha_1, \dots, \alpha_p$  — это все листья, в которые проходит запрос  $x$  и этим листьям приписаны функции  $h_1, \dots, h_p \in H$  соответственно. Назовем состояние  $m_U(x) = h_1 \circ \dots \circ h_p(m_0)$  *финальным состоянием* ВИГ  $U$  на запросе  $x$ . Если запрос  $x$  не проходит ни в один лист, положим  $m_U(x) = m_0$ . Будем называть  $z_U(x) = \sigma(m_U(x))$  *функцией ответа* ВИГ  $U$ .

Как и для УВИГ, будем говорить, что некоторый ВИГ  $U$  над базовым множеством  $\mathcal{F} = \langle F, G, H, M, m_0, \sigma \rangle$  *решает* ВЗИП  $I = \langle X, V, Z, \rho, \xi \rangle$ , если для любого запроса  $x \in X$  функция ответа ВИГ  $z_U(x)$  равна значению  $z_I(x)$ .

### Объёмная сложность вычислительных задач поиска

Пусть задана некоторая ВЗИП  $I = \langle X, V, Z, \rho, \xi \rangle$ . Пусть задано некоторое базовое множество для УВИГ  $\mathcal{F} = \langle F, G, Z \rangle$ . Обозначим  $\mathcal{U}^{wig}(I, \mathcal{F})$  множество всех УВИГ над базовым множеством  $\mathcal{F}$ , которые решают задачу  $I$ . Обозначим  $\mathcal{U}_D^{wig}(I, \mathcal{F})$  множество всех древовидных УВИГ над базовым множеством  $\mathcal{F}$ , которые решают задачу  $I$ . *Объёмной сложностью задачи*

$I$  при базовом множестве  $\mathcal{F}$  назовем число

$$Q^{uvig}(I, \mathcal{F}) = \min\{Q(U) : U \in \mathcal{U}^{uvig}(I, \mathcal{F})\}.$$

Число

$$Q_{\mathcal{D}}^{uvig}(I, \mathcal{F}) = \min\{Q(U) : U \in \mathcal{U}_{\mathcal{D}}^{uvig}(I, \mathcal{F})\}$$

назовем *древовидной объёмной сложностью задачи  $I$  при базовом множестве  $\mathcal{F}$* .

Если  $k$  — натуральное число,  $S$  — тип вычислительных задач информационного поиска, то обозначим

$$\mathcal{I}(k, S) = \{I = \langle X, V, Z, \rho, \xi \rangle \in S : |V| = k\}.$$

Будем исследовать функцию Шеннона, характеризующую объёмную сложность (либо древовидную объёмную сложность) класса ВЗИП  $\mathcal{I}(k, S)$ :

$$Q^{uvig}(k, S, \mathcal{F}) = \sup_{I \in \mathcal{I}(k, S)} Q^{uvig}(I, \mathcal{F}),$$

$$Q_{\mathcal{D}}^{uvig}(k, S, \mathcal{F}) = \sup_{I \in \mathcal{I}(k, S)} Q_{\mathcal{D}}^{uvig}(I, \mathcal{F}).$$

Аналогичные функции  $Q^{vig}(k, S, \mathcal{F})$  и  $Q_{\mathcal{D}}^{vig}(k, S, \mathcal{F})$  введем и для ВИГ — с той лишь разницей, что использоваться должны базовые множества для ВИГ  $\mathcal{F} = \langle F, G, H, M, m_0, \sigma \rangle$ .

Введем тип вычислительных задач информационного поиска  $S_{int1}^{card}$ , заданный следующими соотношениями:

$$S_{int1}^{card} = \langle X_{int1}, Y_{int1}, \mathbb{Z}, \rho_{int1}, \xi_{int1} \rangle, \text{ где} \quad (1)$$

$$X_{int1} = \{(x_1, x_2) \in [0, 1]^2 : x_1 \leq x_2\}, \quad Y_{int1} = [0, 1], \quad (2)$$

$$\forall (x_1, x_2) \in X_{int1}, \forall y \in Y_{int1} : (x_1, x_2) \rho_{int1} y \Leftrightarrow x_1 \leq y \leq x_2, \quad (3)$$

$$\forall W \subseteq Y_{int1} : \xi_{int1}(W) = |W|. \quad (4)$$

Рассмотрим множество предикатов  $F_{int1}$  и переключателей  $G_{int1}$ , заданные следующим образом:  $F_{int1} = \{f^{id}(x) \equiv 1\}$ ,  $G_{int1} = G_1 \cup G_2$ ,

$$G_1 = \left\{ g_{\leq, a}^1(x) = \begin{cases} 1, & \text{если } x_1 \leq a, \\ 2, & \text{если } x_1 > a \end{cases}, \text{ где } a \in \mathbb{R} \right\},$$

$$G_2 = \left\{ g_{<, a}^2(x) = \begin{cases} 1, & \text{если } x_2 < a, \\ 2, & \text{если } x_2 \geq a \end{cases}, \text{ где } a \in \mathbb{R} \right\}.$$

Рассмотрим также множество состояний  $M_{int1} = \mathbb{Z}$ , выделенный его элемент  $m_0 = 0$ , набор функций, меняющих состояние,  $H_{int1} =$

$\{h_a : \mathbb{Z} \rightarrow \mathbb{Z} : a \in \mathbb{Z}\}$ , где  $\forall a \in \mathbb{Z}, \forall m \in \mathbb{Z} : h_a(m) = m + a$ ; и функцию  $\sigma_{int1} : M_{int1} \rightarrow \mathbb{Z}$ , такую что  $\forall m \in \mathbb{Z} : \sigma_{int1}(m) = m$ .

Определим два базовых множества  $\mathcal{F}_{int1}^1$  и  $\mathcal{F}_{int1}^2$ , над которыми мы будем строить УВИГ и ВИГ соответственно:

$$\mathcal{F}_{int1}^1 = \langle F_{int1}, G_{int1}, \mathbb{Z} \rangle, \quad (5)$$

$$\mathcal{F}_{int1}^2 = \langle F_{int1}, G_{int1}, H_{int1}, M_{int1}, m_0, \sigma_{int1} \rangle. \quad (6)$$

**Теорема 2.** *Рассмотрим тип вычислительных задач информационного поиска  $S_{int1}^{card}$ , определяемый соотношениями (1)–(4). Рассмотрим базовое множество  $\mathcal{F}_{int1}^1$ , определённое соотношением (5). Для любого натурального числа  $k$  справедливо:*

$$\frac{k(k+1)}{2} \leq \mathcal{Q}_{\mathcal{D}}^{wig}(k, S_{int1}, \mathcal{F}_{int1}^1) \leq k^2 + 3k,$$

$$4k - 4 \leq \mathcal{Q}^{wig}(k, S_{int1}, \mathcal{F}_{int1}^1) \leq 6k \log_2 k + 4k + 2 \log_2 k + 2.$$

**Следствие 2.1.** *В условиях теоремы 2 выполнено:*

$$\mathcal{Q}_{\mathcal{D}}^{wig}(k, S_{int1}, \mathcal{F}_{int1}^1) \asymp k^2 \quad (\text{при } k \rightarrow \infty),$$

$$k \lesssim \mathcal{Q}^{wig}(k, S_{int1}, \mathcal{F}_{int1}^1) \lesssim k \log_2 k \quad (\text{при } k \rightarrow \infty).$$

**Теорема 3.** *Рассмотрим тип вычислительных задач информационного поиска  $S_{int1}^{card}$ , определяемый соотношениями (1)–(4). Рассмотрим базовое множество  $\mathcal{F}_{int1}^2$ , определённое соотношением (6). Для любого натурального числа  $k$  справедливо:*

$$4k - 4 \leq \mathcal{Q}^{vig}(k, S_{int1}, \mathcal{F}_{int1}^2) \leq \mathcal{Q}_{\mathcal{D}}^{vig}(k, S_{int1}, \mathcal{F}_{int1}^2) \leq 4k + 2.$$

**Следствие 3.1.** *В условиях теоремы 3 выполнено:*

$$\mathcal{Q}^{vig}(k, S_{int1}, \mathcal{F}_{int1}^2) \sim \mathcal{Q}_{\mathcal{D}}^{vig}(k, S_{int1}, \mathcal{F}_{int1}^2) \sim 4k \quad (\text{при } k \rightarrow \infty).$$

Вторая глава диссертации посвящена задачам поиска представителя.

В разделе 2.1 рассматривается поиск представителя для двумерной задачи о доминировании.

Рассмотрим тип задач информационного поиска  $S_{dom2} = \langle X_{dom2}, Y_{dom2}, \geq \rangle$ , где  $X_{dom2} = Y_{dom2} = [0, 1]^2$ , а отношение поиска  $\geq$  задаётся следующим соотношением:  $(a, b) \geq (c, d) \Leftrightarrow a \geq c$  и  $b \geq d$ . Это так называемая *двумерная задача о доминировании*.

В качестве базового множества рассмотрим  $\mathcal{F}_{dom2} = \langle F_{dom2}, G_{dom2} \rangle$ , где  $F_{dom2} = \{f_{\geq a}^2 \mid a \in [0, 1]\}$ ;  $f_{\geq a}^2(x) = 1 \Leftrightarrow x_2 \geq a$  для любых  $x = (x_1, x_2) \in [0, 1]^2$ ;  $G_{dom2} = \{g_{\geq a}^1 \mid a \in [0, 1]\}$ ;  $g_{\geq a}^1(x) = \begin{cases} 1, & \text{если } x_1 \geq a, \\ 2 & \text{иначе.} \end{cases}$  для любых  $x = (x_1, x_2) \in [0, 1]^2$ .

Пусть  $P$  – конечное множество точек из единичного квадрата  $[0, 1]^2$ . Определим множеств  $N(P)$  соотношением  $N(P) = \{p \in P : \forall p' \in P \setminus \{p\} \text{ неверно, что } p \geq p'\}$ . Это множество будем называть *нижним слоем множества  $P$* . По определению  $N(P)$  все его элементы несравнимы (две точки  $a, b \in [0, 1]^2$  называются *сравнимыми*, если  $a \geq b$  или  $b \geq a$ ).

**Теорема 4.** Пусть задана некоторая задача  $I = \langle X_{dom2}, V, \geq \rangle \in S_{dom2}$ , где  $V$  – некоторое конечное подмножество  $Y_{dom2}$ . Тогда существует ИГ  $U$  над базовым множеством  $\mathcal{F}_{dom2}$ , который решает задачу поиска представителя для  $I$  и имеет следующие характеристики:

$$Q(U) = 3k, \quad \lfloor \log_2(k+1) \rfloor + 1 \leq \hat{T}(U) \leq \lceil \log_2(k+1) \rceil + 1,$$

где  $k = |N(V)|$  – размер нижнего слоя множества  $V$ ,  $Q(U)$  – объём  $U$ , а  $\hat{T}(U)$  – сложность  $U$  в худшем случае.

В разделе 2.2 рассматривается поиск представителя для задачи о метрической близости.

Рассмотрим тип задач поиска, описывающий двумерную задачу о метрической близости:  $S_{\square}^R = \langle [0, 1]^2, [0, 1]^2, \rho_{\square}^R \rangle$ , где  $x \rho_{\square}^R y \Leftrightarrow \rho_{\square}(x, y) = \max(|x_1 - y_1|, |x_2 - y_2|) \leq R/2$ ,  $R \in (0, 2)$ .

Будем рассматривать базовое множество  $\mathcal{F}_{\square}^R = \langle F, G \rangle$ , где  $F = F_0 \cup F_1$ ,  $F_0 = \{f^{id}(x) \equiv 1\}$ ,  $F_1 = \{\chi_{y, \rho_{\square}^R}(x) \mid y \in Y\}$ ,  $G = G_1 \cup G_2$ ,  $G_1 = \{g_m(x) = m(i(x_1) - 1) + i(x_2) \mid m \in \mathbb{N}\}$ , где  $i(x) = \max(1, \lceil mx \rceil)$ ,  $G_2 = \left\{ g_{\leq, a}^i(x) = \begin{cases} 1, & x_i \leq a, \\ 2, & x_i > a \end{cases}, g_{\geq, a}^i(x) = \begin{cases} 1, & x_i \geq a, \\ 2, & x_i < a \end{cases} \mid i \in \{1, 2\}, a \in \mathbb{R} \right\}$ .

**Теорема 5.** Пусть дана задача  $I = \langle [0, 1]^2, V, \rho_{\square}^R \rangle \in S_{\square}^R$ . Если вероятностная мера на  $X$  задаётся ограниченной функцией плотности вероятности  $p(x) \leq c$ , то для любого  $t \in \mathbb{N}$ :  $1/t < R/2$  существует ИГ  $U$  над базовым множеством  $\mathcal{F}_{\square}^R$ , решающий задачу поиска представителя для  $I$ , удовлетворяющий условиям:

$$T(U) \leq 1 + 4c \frac{(mR - 1)(2 + R)^2}{(mR - 2)^2} (2 + \lceil \log_2 |V| \rceil), \quad \hat{T}(U) \leq 9 + 4 \lceil \log_2 |V| \rceil,$$

$$Q(U) \leq 4k + m^2 \left( 1 + 60 \frac{(mR - 1)(2 + R)^2}{(mR - 2)^2} \right).$$

**Следствие 5.1.** Если последовательность  $R_k$  удовлетворяет условию  $\frac{1}{R_k} = o\left(\frac{\sqrt{k}}{\log_2 k}\right)$ , то для последовательности задач  $I_k = \langle [0, 1]^2, V, \rho_{\square}^{R_k} \rangle$ , где  $|V| = k$ , существует последовательность ИГ  $U_k$ , решающих соответствующие задачи поиска представителя, такие, что выполнено:

$$Q(U_k) \lesssim 4k, \quad T(U_k) \sim 1, \quad \hat{T}(U_k) \leq 9 + 4\lceil \log_2 k \rceil.$$

Третья глава диссертации посвящена вычислительным задачам информационного поиска.

В разделе 3.1 рассматривается итеративный поиск и техника частичного каскадирования. Техника частичного каскадирования впервые упоминается в работах Чазелле и Гуибаса. В диссертации рассматривается только статический вариант этой техники, однако, в отличие от предыдущих работ по этой теме, частичное каскадирование рассматривается отдельно от каких бы то ни было моделей поиска. Важным отличием от предыдущих результатов на эту тему является то, что полученные результаты имеют вид обычных оценок, а не оценок по порядку. Кроме того, оценки получены при различных значениях дополнительного параметра, ограничивающего максимальную ширину моста. О возможности варьирования этого параметра в работе Чазелле и Гуибаса упоминается, однако все вычисления производятся при фиксированном его значении.

Под *расширенными действительными числами* будем понимать элементы множества  $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ . *Каталогами* будем называть неубывающие последовательности расширенных действительных чисел. Например, последовательность  $-\infty, -5, 0, 0, 7, +\infty, +\infty$  – каталог из семи элементов. Один каталог будем называть *подкаталогом* другого, если он как последовательность является подпоследовательностью второго. Каталог  $-5, 7, +\infty, +\infty$  – подкаталог указанного выше каталога. Пусть  $C$  – каталог. Тогда каталог, получающийся из него удалением дубликатов будем обозначать  $U(C)$ . Например, если  $C = -\infty, -5, 0, 0, 7, +\infty, +\infty$ , то  $U(C) = -\infty, -5, 0, 7, +\infty$ .

Пусть есть два каталога  $C = \{c_i\}_{i=1}^n$  и  $D = \{d_j\}_{j=1}^m$ . Пару индексов  $p = (i, j)$ , такую что  $1 \leq i \leq n$ ,  $1 \leq j \leq m$  будем называть *переходом* между каталогами  $C$  и  $D$ , если  $c_i = d_j$ . Про элементы  $c_i$  и  $d_j$  будем говорить, что они *принадлежат переходу*  $p$ . Для удобства будем называть некоторый переход  $x$ -*переходом*, если значение соединяемых этим переходом элементов равно  $x$ . Последовательность переходов  $B = \{p_k = (i_k, j_k)\}_{k=1}^s$



где  $s \geq 2$  будем называть *мостом* если номера элементов и в первом и во втором каталоге строго возрастают.

Будем говорить, что элемент  $c_i$  каталога  $C$  принадлежит  $k$ -му ярусу моста  $B$  ( $1 \leq k \leq s-1$ ) в том случае, если  $i_k < i < i_{k+1}$ . Очевидно, каждый элемент каталога  $C$  принадлежит не более чем одному ярусу моста  $B$ , но может не принадлежать ни одному ярусу если для него выполнено одно из условий: элемент принадлежит одному из переходов моста, либо элемент лежит выше верхнего перехода моста, либо элемент лежит ниже нижнего перехода моста. Аналогично, будем говорить, что элемент  $d_j$  каталога  $D$  принадлежит  $k$ -му ярусу моста  $B$ , если  $j_k < j < j_{k+1}$ .

*Шириной  $k$ -го яруса* моста  $B$  будем называть суммарное количество элементов в каталогах  $C$  и  $D$ , принадлежащих  $k$ -му ярусу моста  $B$ . *Шириной моста  $B$*  будем называть максимальную ширину некоторого яруса данного моста. Отрезок  $[c_{i_1}, c_{i_s}]$  ( $= [d_{j_1}, d_{j_s}]$ ) будем называть *отрезком покрытия моста  $B$* .

*Графом каталогов* будем называть тройку  $L = (G, \mathcal{C}, \mathcal{R})$ , где  $G = (V, E)$  – граф без петель и кратных ребер ( $V$  – множество вершин графа  $G$ , а  $E$  множество рёбер  $G$ ),  $\mathcal{C} = \{C_v\}_{v \in V}$  – набор каталогов, поставленных в соответствие вершинам  $G$ ,  $\mathcal{R} = \{R_e\}_{e \in E}$  – набор отрезков  $R_e = [a_e, b_e]$  ( $a_e, b_e \in \overline{\mathbb{R}}$  и  $a_e < b_e$ ), поставленных в соответствие рёбрам  $G$ . *Локальной степенью вершины  $v$  графа каталогов  $L = (G, \mathcal{C}, \mathcal{R})$*  будем называть величину  $d(v) = \max_{x \in \overline{\mathbb{R}}} |\{e \in E : v \in e, x \in [a_e, b_e]\}|$ . *Локальной степенью графа каталогов  $L = (G, \mathcal{C}, \mathcal{R})$*  будем называть величину  $d(L) = \max_{v \in V} d(v)$ . Как видно из определения, локальная степень графа каталогов зависит только от  $G$  и  $\mathcal{R}$ , но не зависит от  $\mathcal{C}$ .

*Системой мостов* будем называть четвёрку  $S = (G, \mathcal{A}, \mathcal{R}, \mathcal{B})$ , где  $(G, \mathcal{A}, \mathcal{R})$  образуют граф каталогов ( $\mathcal{A} = \{A_v\}_{v \in V}$  – некоторый набор каталогов), а  $\mathcal{B} = \{B_e\}_{e \in E}$  – набор мостов, такой что выполнены условия: для любого ребра  $\{v, w\} = e \in E$  мост  $B_e$  является мостом между каталогами  $A_v$  и  $A_w$ , либо между каталогами  $A_w$  и  $A_v$ ; для любого ребра  $e \in E$  отрезок покрытия моста  $B_e$  равен отрезку  $R_e$ ; для любого  $x \in \overline{\mathbb{R}}$  каждый мост из  $\mathcal{B}$  содержит не более одного  $x$ -перехода.

Систему мостов  $S = (G, \mathcal{A}, \mathcal{R}, \mathcal{B})$  будем называть *строгой*, если каждый элемент любого каталога  $A_v$ , приписанного любой вершине  $v \in V$ , либо не принадлежит ни одному переходу ни одного моста из  $\mathcal{B}$ , либо принадлежит только одному переходу одного моста из  $\mathcal{B}$ ; Некоторый элемент каталога  $A_v$  будем называть *свободным* в том случае, если он не принадлежит ни одному переходу ни одного моста из  $\mathcal{B}$ . *Шириной системы мостов  $S = (G, \mathcal{A}, \mathcal{R}, \mathcal{B})$*  будем называть максимальную ширину моста среди  $\mathcal{B}$ . Система мостов  $S = (G, \mathcal{A}, \mathcal{R}, \mathcal{B})$  называется *согласованной* с графом

каталогов  $L = (G, \mathcal{C}, \mathcal{R})$  если для любого  $v \in V$  каталог  $U(C_v)$  является подкаталогом  $A_v$ .

**Теорема 6.** Пусть имеется граф каталогов  $L = (G, \mathcal{C}, \mathcal{R})$ , где  $G = (V, E)$  и  $|V| = n$ ,  $|E| = m$ . Пусть  $d = d(L)$  – локальная степень графа каталогов  $L$ . Пусть зафиксирован нечётный натуральный параметр  $g_0 \geq 4d + 1$ .

Тогда существует строгая система мостов  $S = (G, \mathcal{A}, \mathcal{R}, \mathcal{B})$ , согласованная с графом каталогов  $L$  и при этом выполнены следующие условия:

- (а) ширина системы мостов  $S$  не превосходит параметра  $g_0$ ;
- (б) имеет место неравенство  $\sum_{v \in V} |A_v| \leq (1 + c)(\sum_{v \in V} |U(C_v)| + 4m)$ , где  $c = c(g_0) = \frac{4d}{g_0 + 1 - 4d}$ ;
- (в) для любого  $v \in V$  значение каждого из элементов  $A_v$  либо равно значению некоторого элемента из  $C_w$  для некоторого  $w \in V$ , либо равно нижнему или верхнему концу некоторого отрезка  $R \in \mathcal{R}$ ;
- (г) суммарное количество свободных элементов во всех каталогах  $\mathcal{A}$  в точности равно  $\sum_{v \in V} |U(C_v)|$ .

**Следствие 6.1.** Пусть выполнены условия теоремы 6. Тогда существует система мостов  $S = (G, \mathcal{A}, \mathcal{R}, \mathcal{B})$  (уже не обязательно строгая), согласованная с графом каталогов  $L$  и при этом выполнены следующие условия:

- (а) ширина системы мостов  $S$  не превосходит параметра  $g_0$ ;
- (б) имеет место неравенство  $\sum_{v \in V} |A_v| \leq (1 + c)(\sum_{v \in V} |U(C_v)| + 4m)$ , где  $c = c(g_0) = \frac{4d}{g_0 + 1 - 4d}$ ;
- (в) для любого  $v \in V$  значение каждого из элементов  $A_v$  либо равно значению некоторого элемента из  $C_w$  для некоторого  $w \in V$ , либо равно нижнему или верхнему концу некоторого отрезка  $R \in \mathcal{R}$ ;
- (г) любой каталог  $A_v$  содержит не более одной записи с некоторым значением  $x \in \overline{\mathbb{R}}$ .

Если к тому же все отрезки  $R_e \in \mathcal{R}$  равны между собой, то имеет место слегка изменённое ограничение на суммарное количество записей во всех каталогах:  $\sum_{v \in V} |A_v| \leq (1 + c) \sum_{v \in V} |U(C_v)| + 2n + 4mc$ .

В разделе 3.2 рассматривается суммирование по полугрупповой операции для двумерной задачи о доминировании. В работе строится три семейства алгоритмов, решающих задачу суммирования по полугрупповой операции для двумерной задачи о доминировании. Теорема 7 содержит описание семейства алгоритмов, получаемых несколько изменённым методом дерева регионов. Теорема 8 посвящена модификации этого метода с

помощью многократного локального применения техники частичного каскадирования. В теореме 9 производится модификация, подразумевающая одно глобальное применение техники частичного каскадирования.

Пусть  $I$  — некоторая вычислительная задача информационного поиска,  $\mathcal{F}$  — некоторое базовое множество для ВИГ. Будем обозначать как  $\mathcal{U}(I, \mathcal{F})$  множество ВИГ над базовым множеством  $\mathcal{F}$ , решающих задачу  $I$ . Сложностью задачи  $I$  при базовом множестве  $\mathcal{F}$  и объёме  $q$  назовём число  $\hat{T}(I, \mathcal{F}, q) = \min\{\hat{T}(U) : U \in \mathcal{U}(I, \mathcal{F}) \text{ и } Q(U) \leq q\}$ , где  $\hat{T}(U)$  — сложность  $U$  в худшем случае, а  $Q(U)$  — объём  $U$ . Если нет таких ВИГ  $U \in \mathcal{U}(I, \mathcal{F})$ , у которых  $Q(U) \leq q$ , то считаем  $\hat{T}(I, \mathcal{F}, q) = \infty$ . Пусть  $k$  — натуральное число,  $S = \langle X, Y, Z, \rho, \xi \rangle$  — тип вычислительных задач информационного поиска. Введём следующее обозначение:  $\mathcal{I}(k, S) = \{I = \langle X, V, Z, \rho, \xi \rangle \in S : |V| = k\}$ . Таким образом,  $\mathcal{I}(k, S)$  — множество всех ВЗИП типа  $S$ , база данных которых содержит ровно  $k$  элементов.

Наконец, введём величину  $\hat{\mathcal{T}}(k, S, \mathcal{F}, q) = \max_{I \in \mathcal{I}(k, S)} \hat{T}(I, \mathcal{F}, q)$ , характеризующую максимальную сложность задач типа  $S$  с базой данных из  $k$  элементов при решении её вычислительными информационными графами над  $\mathcal{F}$  и объём которых не превышает  $q$ .

Пусть  $\mathcal{Z} = (Z, \oplus)$  — коммутативная полугруппа с нулём. Операцию  $\oplus$  будем для простоты именовать *суммой*. Рассмотрим тип вычислительных задач информационного поиска  $S_{dom2}^{\mathcal{Z}} = \langle X_{dom2}, Y_{dom2}^{\mathcal{Z}}, Z, \rho_{dom2}, \xi_{dom2}^{\mathcal{Z}} \rangle$ , где  $X_{dom2} = [0, 1]^2$ ,  $Y_{dom2}^{\mathcal{Z}} = [0, 1]^2 \times Z$ , отношение поиска  $\rho_{dom2}$  определено таким образом, что для любых  $x = (x_1, x_2) \in X_{dom2}$  и  $y = (y_1, y_2, y_*) \in Y_{dom2}^{\mathcal{Z}}$  соотношение  $x \rho_{dom2} y$  справедливо тогда и только тогда, когда одновременно выполнены условия  $x_1 \geq y_1$  и  $x_2 \geq y_2$ . Элементы множества записей  $Y$  можно рассматривать как точки на плоскости, которым дополнительно приписаны веса из  $Z$ . Функция ответа  $\xi_{dom2}^{\mathcal{Z}} : 2^Y \rightarrow Z$  определена на конечных подмножествах множества  $Y$  следующим образом (значение на бесконечных подмножествах  $Y$  определять не требуется): пусть  $V = \{y^1, \dots, y^k\} \subset Y$  и  $y^i = (y_1^i, y_2^i, y_z^i)$ , тогда  $\xi_{dom2}^{\mathcal{Z}}(V) = y_z^1 \oplus \dots \oplus y_z^k$ . Отдельно определим  $\xi_{dom2}^{\mathcal{Z}}(\emptyset)$  равным нейтральному элементу  $\mathcal{Z}$ .

Вычислительный информационный граф (ВИГ) для решения задач типа  $S_{dom2}^{\mathcal{Z}}$  будем строить над базовым множеством  $\mathcal{F}_{dom2}^{\mathcal{Z}} = \langle F_{dom2}, G_{dom2}, H_{\mathcal{Z}}, M_{\mathcal{Z}}, m_0^{\mathcal{Z}}, \sigma_{id} \rangle$ . Здесь  $F_{dom2} = \{f_{\geq a}^i : a \in [0, 1], i \in \{1, 2\}\}$ , где для любого  $x = (x_1, x_2) \in X$  значение  $f_{\geq a}^i(x)$  равно 1 тогда и только тогда, когда  $x_i \geq a$ . Предикат  $f_{\geq 0}^1$ , значение которого равно 1 на любом запросе  $x$ , будем называть *тождественно единичным предикатом*. Множество переключателей  $G_{dom2} = \{g_{< a}^i : a \in [0, 1], i \in \{1, 2\}\}$ , где для любого

запроса  $x = (x_1, x_2) \in X$  имеет место  $g_{<a}^i(x) = \begin{cases} 1, & \text{если } x_i < a; \\ 2, & \text{если } x_i \geq a \end{cases}$ .

Множество состояний ВИГ  $M_Z$  совпадает со множеством элементов рассматриваемой полугруппы  $Z$ , и начальное состояние  $m_0^Z$  – нейтральный элемент  $Z$ , множество функций, меняющих состояние  $H_Z = \{h_z: h_z(b) = b \oplus z \forall b \in Z\}_{z \in Z}$  и функция выдачи ответа  $\sigma_{id}(z) = z$  для любых состояний  $z \in Z$ .

**Теорема 7.** *Существуют такие  $\gamma(k), \beta(k) = o(1)$  при  $k \rightarrow \infty$ , что для любой коммутативной полугруппы с нулём  $Z = (Z, \oplus)$ , для любых натуральных  $k \geq 3$  и  $s$  таких, что  $2 \leq s < k$ , имеет место:*

$$\hat{\mathcal{T}} \left( k, S_{dom2}^Z, \mathcal{F}_{dom2}^Z, \left( \frac{2k \log_2 k}{\log_2 s} + 10k \right) (1 + \gamma(k)) \right) \leq \frac{s}{2 \log_2 s} \log_2^2 k \cdot (1 + \beta(k)).$$

**Теорема 8.** *Существуют такие  $\gamma(k), \beta(k) = o(1)$  при  $k \rightarrow \infty$ , что для любой коммутативной полугруппы с нулём  $Z = (Z, \oplus)$ , для любых натуральных  $k \geq 3$ ,  $s$  таких, что  $2 \leq s < k$  и нечётного  $g'$ , такого что  $g' \geq 5$  если  $s = 2$  и  $g' \geq 9$  если  $s \geq 3$ , имеет место:*

$$\begin{aligned} \hat{\mathcal{T}} \left( k, S_{dom2}^Z, \mathcal{F}_{dom2}^Z, \left( \frac{5(1+c')k \log_2 k}{\log_2 s} + (9 + 20c')k \right) (1 + \gamma(k)) \right) &\leq \\ &\leq \left( \frac{\log_2^2 k}{2 \log_2 s} + (1 + \lceil \log_2(g' + 1) \rceil) \frac{s-1}{\log_2 s} \log_2 k + 2s \right) (1 + \beta(k)), \end{aligned}$$

где  $c' = \frac{4}{g'-3}$  если  $s = 2$  и  $c' = \frac{8}{g'-7}$  если  $s \geq 3$ .

**Теорема 9.** *Существуют такие  $\gamma(k), \beta(k) = o(1)$  при  $k \rightarrow \infty$ , что для любой коммутативной полугруппы с нулём  $Z = (Z, \oplus)$  выполнено:*

1. *Для любых натуральных  $k \geq 4$  и  $s$  таких, что  $3 \leq s < k$  и нечётного  $g'' \geq 17$ , имеет место:*

$$\begin{aligned} \hat{\mathcal{T}} \left( k, S_{dom2}^Z, \mathcal{F}_{dom2}^Z, \left( \frac{7(1+c'')k \log_2 k}{\log_2 s} + (14 + 196c'')k \right) \cdot (1 + \gamma(k)) \right) &\leq \\ &\leq \frac{\log_2 k}{\log_2 s} (s + \lceil \log_2(2s + 1) \rceil) (1 + \lceil \log_2(g'' + 1) \rceil) + 2s, \end{aligned}$$

где  $c'' = \frac{16}{g''-15}$ .

2. Для любых натуральных  $k \geq 3$  и любого нечётного  $g'' \geq 13$ , имеет место:

$$\begin{aligned} \hat{\mathcal{T}}(k, S_{dom2}^{\mathcal{Z}}, \mathcal{F}_{dom2}^{\mathcal{Z}}, 6(1+c'')k \log_2 k \cdot (1+\gamma(k))) &\leq \\ &\leq 5 \log_2 k (1 + \lceil \log_2(g'' + 1) \rceil) \cdot (1 + \beta(k)), \end{aligned}$$

$$\text{где } c'' = \frac{12}{g''-11}.$$

В разделе 3.3 рассматривается суммирование по полугрупповой операции для двумерной задачи интервального поиска с фиксированной стороной.

Пусть  $\mathcal{Z} = (Z, \oplus)$  – коммутативная полугруппа с нулём. Рассмотрим тип вычислительных задач информационного поиска  $S_{int2^*}^{\mathcal{Z}} = \langle X_{int2^*}, Y_{int2^*}^{\mathcal{Z}}, Z, \rho_{int2^*}, \xi_{int2^*}^{\mathcal{Z}} \rangle$ , где  $X_{int2^*} = \{(x'_1, x''_1, x_2) \in [0, 1]^2 : x'_1 \leq x''_1\}$ ,  $Y_{int2^*}^{\mathcal{Z}} = [0, 1]^2 \times Z$ , отношение поиска  $\rho_{int2^*}$  определено таким образом, что для любых  $x = (x'_1, x''_1, x_2) \in X_{int2^*}$  и  $y = (y_1, y_2, y_*) \in Y_{int2^*}^{\mathcal{Z}}$  соотношение  $x \rho_{int2^*} y$  справедливо тогда и только тогда, когда одновременно выполнены условия  $x'_1 \leq y_1 \leq x''_1$  и  $y_2 \leq x_2$ . Элементы множества записей  $Y_{int2^*}^{\mathcal{Z}}$  можно рассматривать как точки на плоскости, которым дополнительно приписаны веса из  $Z$ . Функция ответа  $\xi_{int2^*}^{\mathcal{Z}} : 2^Y \rightarrow Z$  определена на конечных подмножествах множества  $Y_{int2^*}^{\mathcal{Z}}$  следующим образом: пусть  $V = \{y^1, \dots, y^k\} \subset Y_{int2^*}^{\mathcal{Z}}$  и  $y^i = (y_1^i, y_2^i, y_*^i)$ , тогда  $\xi_{int2^*}^{\mathcal{Z}}(V) = y_*^1 \oplus \dots \oplus y_*^k$ . Отдельно определим  $\xi_{int2^*}^{\mathcal{Z}}(\emptyset)$  равным нейтральному элементу  $\mathcal{Z}$ .

Вычислительный информационный граф (ВИГ) для решения задач введённого выше типа  $S_{int2^*}^{\mathcal{Z}}$  будем строить над базовым множеством  $\mathcal{F}_{int2^*}^{\mathcal{Z}} = \langle F_{int2^*}, G_{int2^*}, H_{\mathcal{Z}}, M_{\mathcal{Z}}, m_0^{\mathcal{Z}}, \sigma_{id} \rangle$ . Положим  $F_{int2^*} = \{f_a^i : a \in [0, 1], i \in \{1', 1'', 2\}\}$ , где для любого  $x = (x'_1, x''_1, x_2) \in X_{int2^*}$  значение  $f_a^{1'}(x)$  равно 1 тогда и только тогда, когда  $x'_1 \leq a$ ; значение  $f_a^{1''}(x)$  равно 1 тогда и только тогда, когда  $x''_1 \geq a$ ; значение  $f_a^2(x)$  равно 1 тогда и только тогда, когда  $x_2 \geq a$ . Предикат  $f_0^2$ , значение которого равно 1 на любом запросе  $x$ , будем называть *тождественно единичным предикатом*. Множество переключателей  $G_{int2^*} = \{g_a^i : a \in [0, 1], i \in \{1', 1'', 2\}\}$ , где для любого запроса  $x = (x'_1, x''_1, x_2) \in X_{int2^*}$  имеет место  $g_a^{1'}(x) = \begin{cases} 1, & \text{если } x'_1 > a; \\ 2, & \text{если } x''_1 \leq a \end{cases}$  ;

$$g_a^{1''}(x) = \begin{cases} 1, & \text{если } x''_1 < a; \\ 2, & \text{если } x''_1 \geq a \end{cases} ; g_a^2(x) = \begin{cases} 1, & \text{если } x_2 < a; \\ 2, & \text{если } x_2 \geq a \end{cases} .$$

Множество состояний ВИГ  $M_{\mathcal{Z}}$  совпадает со множеством элементов рассматриваемой полугруппы  $Z$ , и начальное состояние  $m_0^{\mathcal{Z}}$  – нейтральный элемент  $\mathcal{Z}$ , множе-

ство функций, меняющих состояние  $H_Z = \{h_z: h_z(b) = b \oplus z \forall b \in Z\}_{z \in Z}$  и функция выдачи ответа  $\sigma_{id}(z) = z$  для любых состояний  $z \in Z$ .

**Теорема 10.** *Для любой коммутативной полугруппы с нулём  $Z = (Z, \oplus)$  и любых натуральных  $k \geq 3$  и  $2 \leq s < k$  имеет место:*

$$\hat{T}(k, S_{int2^*}^Z, \mathcal{F}_{int2^*}^Z, 3635 k \log_s k) \leq 59 \frac{s}{\log_2 s} \log_2 k.$$

## Благодарности

Я выражаю глубокую благодарность своему научному руководителю профессору Гасанову Эльяру Эльдаровичу за постановку задач и помощь в работе. Выражаю благодарность заведующему кафедрой математической теории интеллектуальных систем академику, профессору Валерию Борисовичу Кудрявцеву и всему коллективу кафедры за творческую атмосферу, способствующую исследовательской работе.

# Публикации автора по теме диссертации

## Из списка ВАК:

1. Пивоваров А. П. *Поиск представителя в задаче о метрической близости*, Интеллектуальные системы, том 12, вып. 1-4, стр. 333-350 (2008).
2. Пивоваров А. П. *Моделирование вычислительных задач информационного поиска*, Интеллектуальные системы, том 14, вып. 1-4, стр. 237-260 (2010).
3. Пивоваров А. П. *Об одном способе получения нижних оценок сложности информационных графов*, Интеллектуальные системы, том 15, вып. 1-4, стр. 571-580 (2011).
4. Пивоваров А. П. *Техника частичного каскадирования для итеративного поиска в линейно упорядоченных множествах*, Интеллектуальные системы, том 15, вып. 1-4, стр. 249-264(2011).
5. Пивоваров А. П. *Функциональная сложность задачи подсчёта для двумерной задачи о доминировании*, Интеллектуальные системы, том 15, вып. 1-4, стр. 581-610 (2011).

## Не из списка ВАК:

6. Пивоваров А. П. *Суммирование по полугрупповой операции для двумерной задачи интервального поиска с фиксированной стороной*, Материалы X Международной конференции «Интеллектуальные системы и компьютерные науки» (5 - 10 декабря 2011 года). М.: 2011. С. 68 - 71.
7. Пивоваров А. П. *Поиск представителя в задаче о метрической близости*, Международная конференция «Современные проблемы математики, механики и их приложений», посвящённая 70-летию ректора МГУ академика В.А.Садовниченко (30 марта – 2 апреля 2009 г., Москва). Материалы конференции. С. 369.
8. Пивоваров А. П. *Информационные графы с автоматными функциями*, Сборник тезисов XVI Международной конференции студентов, аспирантов и молодых учёных «Ломоносов-2009», секция «Вычислительная математика и кибернетика» (Москва, 13-18 апр. 2009). С. 66.
9. Пивоваров А. П. *Математическая модель неперечислительных задач поиска*, Материалы Международного семинара «Дискретная математика» (Москва, 1-6 февраля 2010 г.). С. 407-410.
10. Пивоваров А. П. *О сложности вычисления ответа одномерного интервального поиска*, Тезисы докладов Секции «Математика и механика» Международной конференции студентов, аспирантов и молодых учёных «Ломоносов-2010».