

# СЕКЦИЯ «СИСТЕМЫ И СЕТИ ИНФОКОММУНИКАЦИЙ»

УДК 004.777

## МЕТОДОЛОГИЯ ВЫБОРА ЭФФЕКТИВНОГО АЛГОРИТМА ПОИСКА

*Довгулевич Е.В., Шука В.С.*

*гр.267041, гр. 263001*

*Белорусский государственный университет информатики и радиоэлектроники,  
г. Минск, Республика Беларусь*

*Научный руководитель: Шевчук О.Г. – канд. техн. наук, доцент кафедры ИКТ*

**Аннотация.** В данном тезисе предлагается анализ процесса выбора алгоритма поиска, включая рассмотрение основных критериев, влияющих на это решение, а также предоставление рекомендаций по оптимальному подходу, учитывая разнообразие задач и требований современных информационных систем.

**Ключевые слова:** алгоритмы поиска, критерии поиска, матрица сравнения алгоритмов

**Введение.** Одной из важнейших процедур обработки информации является поиск. Алгоритмы поиска успешно применяются в различных сферах современных информационных систем (ИС), включая анализ данных, биоинформатику, машинное обучение, криптографию, робототехнику, игровую индустрию и др.

Выбор эффективного алгоритма поиска играет ключевую роль в обеспечении производительности и функциональности ИС. Существует ряд актуальных проблем, с которыми сталкиваются специалисты, включающие в себя учет различных требований задач, высокая производительность с большими объемами данных, а также необходимость адаптации к изменяющимся условиям.

**Основная часть.** Существует множество различных алгоритмов поиска, имеющих свои преимущества и недостатки [1]. Рассмотрим основные классы алгоритмов, представляющие собой фундаментальные подходы к решению задач поиска.

**Линейный.** Один из наиболее простых методов, при котором элементы последовательно проверяются на соответствие критерию поиска, однако, эффективность снижается с увеличением объема данных. Применяется в случаях, когда структура данных не предоставляет дополнительной информации о распределении элементов.

**Бинарный.** Метод для упорядоченных данных, при котором диапазон поиска сокращается в два раза на каждом шаге, обеспечивая значительный прирост производительности, особенно в случае больших объемов данных.

**Хеширование.** Алгоритмы хеширования преобразуют входные данные в уникальные значения, обеспечивая быстрый доступ к ним. Однако эффективность зависит от качества хеш-функции и уровня коллизий. Идеально подходит для поиска в базах данных и словарях.

**Алгоритмы поиска в графах.** Методы для решения задач поиска в сетевых и графовых структурах данных, таких как алгоритм Дейкстры, алгоритмы поиска в ширину (BFS) или в глубину (DFS) [2].

**Алгоритмы машинного обучения.** Поиск информации в собранных данных на этапе моделирования производится математической функцией. Обучение зависит от типа используемой модели [3].

Выбор подходящего алгоритма для конкретных требований – это важная задача, и существует несколько методов для определения, какой алгоритм может быть наиболее оптимальным. Однако, универсального метода нет, а эффективность каждого зависит от конкретной задачи, данных и условий применения. Необходимо предварительно систематизировать задачи по их характеристикам. Разработка методологии систематизации задач позволит более точно определить, какие критерии следует учитывать при выборе алгоритма. Рекомендованы следующие критерии:

1. Эффективность по времени и ресурсам. Скорость выполнения является главным фактором при обработке запросов в реальном времени. Ресурсоемкость актуальна в условиях ограниченных вычислительных мощностей или при работе в облачных сервисах.

2. Работа с большими объемами данных. Алгоритмы должны быть спроектированы таким образом, чтобы обеспечивать высокую скорость поиска и обработки больших объемов данных. Эффективность алгоритма может зависеть от того, каким образом данные организованы в памяти. Существенно улучшить производительность позволит использование оптимальных структур данных, таких как деревья или хеш-таблицы.

3. Устойчивость к изменениям в данных. Способность динамически адаптироваться к изменениям входных данных, включая добавление или удаление элементов. Особенно актуально в системах, в которых данные поступают или изменяются часто.

Определение алгоритма в рамках каждой категории задач производится созданием матрицы сравнения, в которой алгоритмы оцениваются по различным критериям. Матрица приоритетов (матрица критериев) – это инструмент, с помощью которого можно ранжировать по степени

важности данные и информацию [4]. Критерии не ограничены перечисленными выше, к примеру, критериями могут являться простота реализации, специфические требования к задаче и другие. Оценки могут быть представлены в числовой форме или использоваться качественные оценки, в зависимости от конкретных требований. Матрица представляет собой таблицу, в которой в строках указываются алгоритмы, а в столбцах – критерии сравнения. Пример матрицы сравнения алгоритмов приведен в таблице 1.

Таблица 1 – Матрица сравнения алгоритмов поиска.

Класс алгоритмов	Основные характеристики	Эффективность по времени	Работа с большими объемами данных	Устойчивость к изменениям в данных
Линейный поиск	Простота реализации	Низкая Линейная	Затраты растут экспоненциально с увеличением данных	Неустойчив
Бинарный поиск	Требуется отсортированный массив	Логарифмическая	Не подходят для неупорядоченных данных	Требуется пересортировка данных при изменениях
Хеширование	Быстрый доступ к данным	Константная	Зависит от хеш-функции. Возможны коллизии	Устойчив
Алгоритмы поиска в графах	Требуется оптимизация и адаптация	Зависит от конкретной задачи и характеристик графа (количество вершин и ребер)	Некоторые алгоритмы становятся вычислительно сложными с увеличением размера графа	Зависит от конкретного алгоритма. DFS и BFS -устойчивы
Алгоритмы машинного обучения	Высокая точность, но требуется обучение	Зависит от модели и данных	Подходят для обработки больших объемов данных	Устойчив

Следующим этапом методологии анализа алгоритмов предполагается создание методов количественной оценки критериев для каждого алгоритма. Эти методы могут включать в себя разработку метрик. Позволяет выразить оценки по каждому критерию в числовой форме, что упрощает анализ и сравнение алгоритмов, обеспечивает систематизацию подхода к этому процессу.

Применим методологию выбора алгоритма на примере конкретной задачи: анализ данных в области медицинского исследования. Характеристики задачи: объем данных – большой (медицинские записи, изображения и результаты исследований); временные ограничения – разные, в зависимости от конкретной задачи; структура данных – разнообразная (текст, изображения, числовые данные).

Систематизация задачи по поиску в базе данных медицинских записей включает в себя разделение задачи на подзадачи в соответствии с характеристиками данных (текстовые, изображения, числовые данные) и определение ключевых критериев для каждой подзадачи. Таким образом, можно выделить три подзадачи. Подзадача 1: поиск в текстовых медицинских записях. Значимо точное нахождение информации, учитывая особенности медицинской терминологии и семантики. Подзадача 2: распознавание и поиск в медицинских изображениях, в снимках с медицинских устройств (например, в рентгеновских снимках). Подзадача 3: поиск в числовых данных, таких как результаты лабораторных исследований. Каждая подзадача подчиняется своим уникальным требованиям, и выбор алгоритмов для решения этих подзадач должен быть согласован с соответствующими ключевыми критериями для обеспечения оптимальной эффективности системы поиска в базе данных медицинских записей.

Создадим матрицу сравнения основных алгоритмов (представлена в таблице 2) в базе данных медицинских записей, учитывая следующие критерии: точность, возможность работы с различными типами данных и устойчивость к изменениям в данных.

Таблица 2 – Матрица сравнения алгоритмов в базе данных медицинских записей.

Алгоритм поиска	Точность	Возможность работы с различными типами данных	Устойчивость к изменениям данных
Линейный	++	++++	---
Бинарный	+++	+++	---
Хеширование	++++	++++	--
Алгоритмы поиска в графах	+++	++++	++++
Машинное обучение	++++	++++	++++

Объяснение обозначений:

++++ : очень высокий уровень соответствия критерию;

+++ : высокий уровень соответствия критерию;

++ : умеренный уровень соответствия критерию;

--- : низкий уровень соответствия критерию;

-- : очень низкий уровень соответствия критерию.

Оценив различные алгоритмы поиска с учетом требований задачи в базе данных медицинских записей, резюмируем:

- не рекомендованы линейный и бинарный поиски. Первый, т.к. менее эффективен для больших объемов данных, второй – по причине того, что требует упорядоченных данных;
- предпочтительны в использовании: для текстовых данных и изображений – машинное обучение, для числовых данных – алгоритмы в графах, при хорошей хеш-функции и ограниченном объеме данных – хеширование.

**Заключение.** Методология выбора алгоритма является критическим элементом в области информационных систем и приложений, где эффективность поиска напрямую влияет на общую производительность. Анализ различных критериев выбора, таких как эффективность по времени и ресурсам, работа с большими объемами данных и устойчивость к изменениям, позволил выделить важные аспекты, которые следует учитывать при выборе алгоритма. Методология, предложенная в тезисе, предоставляет систематизированный и комплексный подход к этому процессу. Применение методологии на практике, через рассмотрение конкретной задачи и сравнение результатов выбранных алгоритмов, демонстрирует ее эффективность в различных областях, и подтверждает необходимость индивидуального подхода к выбору алгоритмов в зависимости от контекста задачи.

### Список литературы

1. Алгоритмы поиска данных [Электронный ресурс]. – Режим доступа : <https://top-technologies.ru/ru/article/view?id=24620> - Дата доступа : 13.02.2024.
2. Базовые алгоритмы на графах [Электронный ресурс]. – Режим доступа : <https://habr.com/ru/companies/timeweb/articles/751762/> - Дата доступа : 13.02.2024.
3. Машинное обучение как инструмент анализа данных [Электронный ресурс]. – Режим доступа : <https://rep.bntu.by/bitstream/handle/data/110567/304-309.pdf?sequence=1&isAllowed=y> - Дата доступа : 16.02.2024.
4. Матрица приоритетов [Электронный ресурс]. – Режим доступа : [https://www.kpms.ru/Implement\\_Qms\\_Prioritization\\_Matrix.htm](https://www.kpms.ru/Implement_Qms_Prioritization_Matrix.htm). - Дата доступа : 19.02.2024.

UDC 004.777

## METHODOLOGY FOR SELECTING AN EFFECTIVE SEARCH ALGORITHM

*Dovgulevich E.V.*

*gr. 267041*

*Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus*

## 60-я научная конференция аспирантов, магистрантов и студентов

*Scientific supervisor: Shevchuk O.G. - Kand. techn. sciences, associate professor of the department of ICT*

**Annotation.** This report provides an analysis of the process of selecting a search algorithm, including consideration of the main criteria influencing this decision, as well as recommendations for an optimal approach, taking into account the diversity of tasks and requirements of modern information systems.

**Keywords:** search algorithm, search criteria, algorithm comparison matrix