

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ

**Федеральное государственное автономное образовательное учреждение
высшего образования**

**Национальный исследовательский университет
«Высшая школа экономики»**

**Факультет компьютерных наук
Образовательная программа
«Прикладная математика и информатика»**

КУРСОВАЯ РАБОТА

На тему: Различные способы построения графовых структур данных для поиска
ближайшего соседа

Тема на английском: Several Ways to Form Graph Based Nearest Neighbour Search
Structure

Студент / студентка 2-го курса
группы № БПМИ2110/21ПМИ-2:

Рябков Игорь Дмитриевич
(Ф.И.О.)

Научный руководитель:

Пономаренко Александр Александрович
(Ф.И.О.)

Доцент, НН Кафедра прикладной математики и информатики
(должность, звание)

Содержание

1	Введение	3
2	Основная часть	5
2.1	История развития структур для поиска ближайшего соседа	5
2.2	Феномен тесного мира	5
2.3	Подход Клайнберга	5
2.4	Подход Пономаренко	5
2.5	Предложения по модификации	5
3	Практическая часть	6
3.1	Описание абстракций	6
3.2	Эксперименты и наблюдения(параметры)	6
3.3	Проверка ранее выдвинутых гипотез	6
4	Заключение	7
5	Список литературы	8

1 Введение

В последнее десятилетие наша жизнь стала тесно связана с удобными приложениями и сервисами. Многие из них базируются на рекомендательных системах, для предоставления целевого товара на основе наших интересов. Некоторые используют компьютерное зрение для решения огромного кол-ва задач (от масок в социальных сетях, до автопилота в электроавтомобилях). Поиск синонимов и системы автоматического дополнения текста (T9) также используются каждый день миллионами пользователей. Это лишь малая часть задач, которые можно решить используя алгоритм поиска ближайшего соседа (или поиска К-ближайших соседей).

Вот ещё некоторые задачи, о которых хотелось бы упомянуть:

- Поиск дубликатов (определить являются ли 2 текстовых документа одинаковыми)
- Задача кластеризации (Определить, как какой группе относится выбранный объект)
- Поиск ближайших географических объектов (карты)
- Поиск схожих фрагментов в фильмах или музыке

Именно поэтому так важно искать новые подходы для улучшения скорости данного алгоритма. Чтобы достичь поставленную цель, необходимо разработать структуру данных, которая сможет наиболее эффективно осуществлять две операции добавления и поиска. Вариантов подходящих структур - огромное множество. Например, некоторые могут быть построены на базе вектора, списка, дерева, графа (в виде сети). Некоторые поддерживают точные поиск, а некоторые только приближённый. Некоторые формируются по средствам детерминированных алгоритмов, а некоторые используют рандомизированный подход.

Моё исследование будет в основном основываться на изучении графовых структур (в виде сетей), так как они более современные и эффективные (подробнее ниже). Перед собой я ставлю следующие задачи:

- Изучить какие структуры для поиска ближайшего соседа существуют
- сравнить их асимптотику построения этих структур и поиска внутри них, выявить преимущества и недостатки
- Исследовать феномен тесного мира.
- Обосновать выбор именно графовых структур, а также подробнее исследовать те из них, которые имеют свойств тесного мира.

- Выдвинуть несколько предположений о модификациях, которые смогли бы улучшить имеющиеся методы.
- Разработать необходимые абстракции для работы с подобными структурами
- На основе сравнительного анализа определить наилучшую конфигурацию параметров в разработанных классах для наибольшей эффективности алгоритма поиска ближайшего соседа
- А также проверить эффективность предложенные ранее модификации

2 Основная часть

2.1 История развития структур для поиска ближайшего соседа

2.2 Феномен тесного мира

2.3 Подход Клайнберга

2.4 Подход Пономаренко

2.5 Предложения по модификации

3 Практическая часть

3.1 Описание абстракций

3.2 Эксперименты и наблюдения(параметры)

3.3 Проверка ранее выдвинутых гипотез

4 Заключение

5 Список литературы