**Министерство науки и высшего образования Российской Федерации**

**федеральное государственное автономное образовательное учреждение высшего образования**

**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО»**

**(Университет ИТМО)**

**Факультет: Информационных технологий и программирования**

**Направление подготовки: 01.03.02 Прикладная математика и информатика**

Образовательная программа: Информатика и программирование

О Т Ч Е Т

об *(учебной)*практике

Тема задания: Разработка алгоритма нахождения релевантности пары.

Обучающийся: Осадчий Максим Дмитриевич

Руководитель практики от профильной организации: Кочерга Евгений, программист

Руководитель практики от университета: Аксенов Виталий Евгеньевич, научный сотрудник, факультет информационных технологий и программирования

Практика пройдена с оценкой:

Дата

Санкт-Петербург

2020 г.

1. **Цели и задачи практики**

Цели практики — изучение новых технологий, применение их на практике. Так же целью практики является получение опыта в области командной работы. Практика позволяет получить опыт в работе над реальными проектами, предложенными компаниями, вероятно, являющимися будущими работодателями. Так же одна из целей практики - это возможность пойти на стажировку в компанию.

1. **Сведения об организации**

Huawei - это мультинациональная корпорация, крупнейший в мире поставщик сетевого и телекоммуникационного оборудования. Компания основана в 1988 году отставным офицером Народно-освободительной армии Китая Жэнь Чжэнфэем (Ren Zhengfei). Штаб-квартира располагается в городе Шэньчжэнь, провинция Гуандун, Китай.

Ключевыми направлениями деятельности компании являются:

1. Телекоммуникационные сети, в частности проектирование и строительство телекоммуникационных сетей;
2. Глобальные сервисы, включая комплексное оснащение, оперативное обслуживание и консалтинговые услуги для корпоративных клиентов;
3. Мобильные устройства, а именно производство и продвижение мобильных устройств под собственным брендом.

 Научно-исследовательские центры Huawei, количество которых перевалило за 20, расположены в Китае, США, Германии, Турции, Индии, Швеции и России. Кроме этого, компания создает совместные инновационные центры с крупнейшими мировыми операторами связи, такими как: Vodafone Group, Telecom Italia, BT Group, Telefonica, Deutsche Telekom, France Telecom и рядом других. Продукция и технические решения компании используются более чем в 140 странах мира.

Huawei – ведущий мировой поставщик инфокоммуникационных решений, цель которого – обогащать жизнь людей через общение. Благодаря нашим инновационным разработкам, ориентированным на потребности клиентов, и развитой партнерской сети мы достигли высоких результатов в разработке телекоммуникационных сетей, терминальных решений и систем облачных вычислений. 180 тыс. сотрудников Huawei по всему миру стремятся создавать максимально выгодные условия для операторов, предприятий и конечных пользователей, предоставляя им конкурентоспособные решения и услуги. Одна треть населения планеты использует наши инновационные продукты, решения и сервисы более чем в 170 странах мира. Основанная в 1987 году Huawei является частной компанией, полностью находящейся во владении своих сотрудников.

1. **Занимаемая должность**

Программист**-**стажер.

1. **Цели проекта**

Проект содержит в себе задачу по реализации алгоритма, дающего оценку релевантности каждой пары, исходя только из оценок анонимных пользователей. Цель проекта — ознакомиться с простейшими формами оценки релевантности, разработать алгоритм с учетом возможности существования недобросовестных оценщиков, что повышает реалистичность задачи и получении навыков тестирования программ. А также составление тестов исходя из различных ситуаций: 1) нет недобросовестных оценщиков; 2) какое-то количество людей ошиблось; 3) есть и недобросовестные, и те, кто ошибся.

Получение ценного опыта по реализации проекта очень востребованного на рынке.

1. **Использованные технологии**

Для составления алгоритма были использованы знания математической статистики. Для написания программного кода был использован язык C++ и знания структур данных — таких как map.

1. **Описание выполненного проекта**

Формулировка задачи:

Даны пары (запрос — URL) и пары могут повторятся. Для каждой пары ставится оценка [0..4] анонимным пользователем. Нужно для каждой уникальной пары (запрос — URL) найти среднюю оценку соответствующую действительности.

Реализация : <https://github.com/maxosad/Relevance-score>.

Задача примечательна тем, что в ней нам даны данные, которые не совсем соответствуют действительности. Некоторые из оценок, поставленных парам могу быть не корректными как в одну, так и в другую сторону. Они могут отличаться незначительно, если мы полагаемся на субъективное мнение честного человека. Или же она может отличаться значительно, если злоумышленник захотел испортить нам статистику намеренно. Мы не знаем кто ставит эти оценки, поэтому не можем вычислять злоумышленников. Следуя из этого мы будем работать только с самими оценками. Не будем учитывать самые подозрительные. Мы надеемся, что большая часть оценок верна. Пусть у каждой пары есть какая — то реальная оценка. Тогда все оценки, которые не сильно отклоняются от нее нам не мешают и не будем их трогать. Давайте тогда для каждой пары выделим все оценки и посчитаем для них дисперсию.

Не будем учитывать самую большую оценку и самую маленькую. Если одна из этих оценок была близка к реальному ответу, то мы мало потеряли от того что перестали учитывать ее, другое же число было либо не далеко от нашего реального ответа, тогда мы тоже ничего не потеряли, либо далеко и тогда дисперсия уменьшиться. Экспериментальным путем выведено, что при дисперсии 0.5 оценки для каждой пары будут соответствовать реальности. Продолжим отбрасывать крайние оценки, пока дисперсия не станет меньше или равна 0.5 . Были разработаны тесты, рассматривающие разные состояния: все идеально, много специальных ошибок, случайные. На них решение показало правильный результат — результат соответствующий идеальным тестам. Также эти тесты были объединены. На объединенных тестах решение тоже показало результат схожий с идеальным.

1. **Выводы**

Оценка релевантности — это не такая уж и простая задача. В реальном мире мы встречаемся со злоумышленниками, поэтому мы не можем опираться на исходные данные. Чтобы получить какие-то реальные данные нам нужно провести обработку исходных. Было обращено внимание, что мы опираемся только на мнение оценщиков, поэтому результаты, которые были получены являются показателями полезности URL для людей искавших по конкретным запросам. Решение успешно протестировано и цели и задачи практики были выполнены, ценный опыт был получен, и таким образом, обучающийся теперь готов к работе в настоящей команде и реализации реального проекта.

1. **Список источников**

1.Система контроля версий. /Электронный ресурс/. Режим доступа <https://github.com/> (Дата обращения 07.02.2020 г.)

2.Справочник по языку С++. /Электронный ресурс/. 2019. Режим доступа <https://docs.microsoft.com/ru-ru/cpp/cpp/cpp-language-reference?view=vs-2019>(дата обращения 07.02.2020 г.)

3. Сайт Huawei. /Электронный ресурс/ Режим доступа <https://www.huawei.com/ru/> (Дата обращения 05.02.2020 г.)

4. Статья Ашманова "Что такое релевантность" /Электронный ресурс/ 21.12.2017. Режим доступа <https://www.ashmanov.com/education/articles/relevantnost/> (Дата обращения 06.02.2020 г.)

5.Статья "Дисперсия случайной величины" /Электронный ресурс/ Режим доступа [https://ru.wikipedia.org/wiki/%D0%94%D0%B8%D1%81%D0%BF%D0%B5%D1%80%D1%81%D0%B8%D1%8F\_%D1%81%D0%BB%D1%83%D1%87%D0%B0%D0%B9%D0%BD%D0%BE%D0%B9\_%D0%B2%D0%B5%D0%BB%D0%B8%D1%87%D0%B8%D0%BD%D1%8B](https://ru.wikipedia.org/wiki/Дисперсия_случайной_величины) (Дата обращения 06.02.2020 г.)

6.Статья "Среднеквадратичное отклонение" /Электронный ресурс/ Режим доступа [https://ru.wikipedia.org/wiki/%D0%A1%D1%80%D0%B5%D0%B4%D0%BD%D0%B5%D0%BA%D0%B2%D0%B0%D0%B4%D1%80%D0%B0%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%BE%D0%B5\_%D0%BE%D1%82%D0%BA%D0%BB%D0%BE%D0%BD%D0%B5%D0%BD%D0%B8%D0%B5](https://ru.wikipedia.org/wiki/Среднеквадратическое_отклонение) (Дата обращения 06.02.2020 г.)