|  |
| --- |
| Добрый день, уважаемые члены аттестационной комиссии. Вашему вниманию предлагается выпускная квалификационная работа Сабуровой Ви на тему «критерий на открытый текст на базе решающего леса». |
| Обозначим цели и задачи работы:  Целью работы является реализация различных вариантов критериев на открытый текст и их применения в различных задачах.  Средствами для достижения данной цели будут являться задачи, приведенные на слайде. |
| Все выше перечисленные задачи относятся к науке криптографии, изучающей построение и использование систем шифрования.  Для того чтобы оценить используемые критерии необходимо выбрать способ шифрования и деширфования. |
| В данной работе для дешифрования текста был выбран метод полного перебора ключей. Метод основан на поиске решения, которое исчерпывает всевозможные варианты. Для шифрования был использован регистр сдвига с линейно обратной связью, где значение ключа задается вручную. Чем выше изначально задавалось значение ключа, тем дольше по времени шло расшифровывание. |
| Рассмотри первые два критерия.  В критерии Вальда на каждой стадии эксперимента совокупность всевозможных выборок различных частей текста объема - символов разбивается на три непересекающихся множества. При попадании выборки:   * в множество принимают гипотезу , испытания заканчивают; * при попадании в множество принимают гипотезу и испытания заканчивают.   В случае попадания выборки в область не принимают ни одной гипотезы, а производят следующее испытание и анализируют аналогично предыдущему выборку .  В данной работе к качестве гипотезы выступает гипотеза о том, что текст является открытым, а в - закрытым. Границы множества выбираются путем подсчета значений порогов, основанных на коэффициенте правдоподобия.  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Использование критерия запрещенных биграмм основано на применение статистики для анализа текстовой информации.  Одной из основных характеристик текста является повторяемость и сочетаемость пар букв – биграмм.  Основная идея критерия состоит в следующем:   * подсчет числа вхождений каждой возможных биграмм текстах , составленных из букв алфавита для открытого текста; * Сравнение биграмм поданного на вход текста с биграммами открытого текста:   + При отсутствии запрещенных (несуществующих) биграмм текст объявляется открытым (не зашифрованным);   + При появлении запрещенной биграммы поданный текст объявляется закрытым. |
| При тестировании разработанного приложения было выявлено, что критерий Вальда лучше определяет открытый текст, а критерий запрещенных биграмм работает лучше с закрытыми текстами. Данный факт обусловлен быстрым появлением запрещенной биграммы в шифрованном тексте. В приведенном отрывке уже вторая биграмма является запрещенной, программа сразу объявит текст закрытым, в связи с чем нет необходимости высчитывать пороги. |
|  |
| Критерии на открытый текст находят применение в различных областях и решениях задач:   * Распознавание типов файлов; * Определение авторства текста; * Распознавание кодировок; * И еще нужно один пункт придумать.   Так, например, авторство текста можно определять из статистических данных, основанных на частоте появления каждой биграммы данного автора и подаваемого на вход текста. Если одна из биграмм, свойственных автору отсутствует, текст объявляется «закрытым», то есть принадлежность автору не установлена. |
| Результатов данной работы является разработанное на языке C++ приложение, которое содержит в себе функции все вышеперечисленные задачи.  Основными направления развития приложения будут служить:   * Оптимизация программы и ускорение ее работы; * Реализация критериев с помощью решающих лесов; * Применение реализованного критерия в задаче классификации открытых и закрытых текстов. |