DOI: 10.17117/na.2015.11.03.254

Поступила (Received): 28.11.2015

<http://ucom.ru/doc/na.2015.11.03.254.pdf>**Лапшова М.А.****Использование многопоточного программирования для создания программного продукта, контролирующего и модифицирующего текст, для безопасной работы детей за компьютером и в сети Internet****Lapshova M.A.****Multithreaded programming for creating software of controlling and modifying text for the safe operation computer and the Internet by children**

В работе рассматривается решение проблемы безопасной работы детей за компьютером и в сети Internet с помощью программного обеспечения. Сравниваются разные пути решения этой проблемы по эффективности и удобству использования. Предложен алгоритм написания программы, контролирующей и модифицирующей текст, содержащий нецензурные выражения, для безопасной работы детей

This paper considers the question of using software for the safe operation computer and the Internet by children. Different ways about solving this problem are compared. The algorithm of writing software(it controls and modifies text with obscene language) is offered in this article

Ключевые слова: многопоточное программирование, критическая секция, обработка текста

Key words: multithreaded programming, critical section, word processing

Лапшова Марина Александровна

Студент

Самарский государственный аэрокосмический университет им. С.П. Королева
г. Самара, Московское шоссе, 34

Lapshova Marina Aleksandrovna

Student

Samara state aerospace university named S.P.
Korolev
Samara, Moskovskoye Shosse, 34

В современном мире люди постоянно пользуются благами цивилизации. Разнообразные электронные устройства используются не только для работы и обучения, но и для развлечения и общения. Иногда пользователи не ограничивают себя литературным языком, нормами приличия и используют в своем общении профессиональный сленг, нецензурные слова и выражения. Разумеется, использование ненормативной лексики недопустимо, особенно учитывая то, что многие дети сейчас пользуются Интернетом. Для решения этих проблем

необходимо контролировать поток текстовой информации. Некоторые владельцы сайтов, чатов, форумов борются с нецензурной лексикой, сленгом и жаргонизмами организационными методами, устанавливая правила. Но этот способ полностью определяется честностью и воспитанием гостей сайта, поэтому полагаться на него не стоит. Чтобы наверняка предотвратить появление ненормативной лексики в тексте, необходима специальная программа автоматического исправления неподходящих слов при обработке текстовой информации.

Есть несколько способов поиска нежелательных слов или словосочетаний в текстовой информации:

1. Создание библиотеки корней слов. Поиск вхождений этих корней в словах текста. Выделение всего слова (от первой буквы до пробела или знака препинания). Этот подход имеет большой недостаток. При нем возможно исправление слов, которые соответствуют нормативной лексике, так как корни слов из библиотеки могут быть частью обычного слова.

2. Создание библиотеки корней искомых слов и всевозможные формы слов с подобными корнями. На первом этапе находится слово, в которое входит данный корень; на втором проверяется, является ли оно целиком словом, которое нужно заменить.

3. Создание библиотеки «запрещенных» слов, в которую войдут всевозможные слова. В ней необходимо учесть, что слово в русском языке может включать в себя следующие морфемы: приставка, корень, интерфикс, суффикс, окончание. Для этого используются регулярные выражения. Далее происходит исправление всех слов, совпадающих с библиотечными словами.

Способы исправления найденных нежелательных слов:

1. Удаление сообщения (предложения), содержащего слово из библиотеки.
2. Замена корня слова из библиотеки на знаки (например звездочки).
3. Замена слова из библиотеки целиком на знаки.
4. Замена слов из библиотеки на синонимичные слова в нормативной лексике, учитывая все особенности русского языка (формы, времена, рода и т.д.).
5. Замена слов из библиотеки на случайное слово из библиотеки «хороших» слов, учитывая части речи и особенности русского языка.

Чтобы программа работала наиболее надежно, выбран второй способ обнаружения нежелательных слов. Для сохранения смысла и содержания текстового сообщения замена слов будет происходить на их синонимы с учетом регистра и формы слова.

Возможны два варианта работы программы: последовательный или параллельный. Для более быстрой и эффективной работы программы выбран способ с параллельными вычислениями. Для этого применим многопоточное программирование, которое позволит сократить время работы программы. Поиск слов, требующих замены, выполняется с помощью массива потоков. Каждый поток выполняет функцию поиска вхождения форм одного корня. В этой функции по мере обнаружения слов заполняется массив замен. Заполнение каждого элемента массива происходит в критической секции, чтобы обеспечить синхронизацию потоков. Программа продолжит работу только тогда, когда завершат

свою работу все потоки из массива. Это возможно с применением функции ожидания завершения работы всех потоков массива.

Для реализации этой программы выбран язык программирования C++, так как он один из самых распространенных языков и позволяет вызывать системные функции и работать с несколькими потоками. Для этого в C++ существуют следующие функции:

- `CreateThread()` – создает новый поток внутри вызывающего потока;
- `ExitThread()` – вызывается извне потока и завершает его;
- `WaitForMultipleObjects()` – приостанавливает поток до момента проявления одного из объектов, указанных в списке;

В ходе моей работы был разработан алгоритм работы программы.

Программа считывает информацию о библиотеке слов и их замен, размещенную в текстовом документе, который можно редактировать и создает в памяти массив структур, каждый элемент которого состоит из корня слов и пар слов: «заменяемое слово – синоним». Этот файл записывается по правилам: все слова записываются в нижнем регистре; каждая строка – корень со всеми его формами и их замены; сначала записан корень слова, далее знак «;», далее следуют пары слов в формате: «слово>синоним;»; в конце строки знак «;» не ставится.

Следующим шагом программа считывает исходный текст и ищет вхождения корней слов из библиотеки в тексте. Для верного поиска все символы переводятся в нижний регистр. Номер позиции первого вхождения корня передается в функцию выделения слова. Выделенное слово проверяется на совпадения в библиотеке слов с текущим корнем. Если слово целиком присутствует в библиотеке, то в массив замен записываются следующие данные: номер позиции начала заменяемого слова, номер позиции конца этого слова, новое слово синоним. Массив замен – массив структур из трех полей (номер начала, конца слова и синоним).

Поиск заменяемых слов и заполнение массива происходит несколькими потоками, которые ищут вхождения форм определенного корня. Заполнение массива происходит в критической секции.

Массив замен сортируется по номеру позиции начала заменяемого слова.

Новый текстовый файл исправленного текста заполняется по правилам: фрагмент текста, не содержащего искомого слова, копируется до позиции начала ближайшего заменяемого слова. Далее в конец строки добавляется слово-синоним из массива замен. До следующей позиции начала ближайшего «плохого» слова строка копируется без изменений и т.д.

Таким образом, после работы программы, создается новый, содержащий исправления, текст и не изменяется исходный файл. В библиотеку слов можно добавлять новые корни и его формы по мере необходимости.

При использовании данной программы, пребывание в сети Internet и работа за компьютером станет безопаснее для детей.

Список используемых источников:

1. Бьерн Страуструп. Язык программирования C++. Специальное издание 2011. Бином, 2011. 1136 с.
2. Энтони Уильямс. Параллельное программирование на C++ в действии. Москва, 2012. 672с.

© 2015, Лапшова М.А.

Использование многопоточного программирования для создания программного продукта, контролирующего и модифицирующего текст, для безопасной работы детей за компьютером и в сети Internet

© 2015, Lapshova M.A.

Multithreaded programming for creating software of controlling and modifying text for the safe operation computer and the Internet by children