УДК 378.02

Реализация эффективных алгоритмов поиска подстроки в строке

Тихомирова Инга Николаевна
Ярославский государственный педагогический университет им. К.Д.Ушинского
студент

Воронцова Ксения Максимовна
Ярославский государственный педагогический университет им. К.Д.Ушинского
студент

Корнилов Петр Анатольевич
Ярославский государственный педагогический университет им. К.Д.Ушинского
к.ф.-м.н., доцент, зав. кафедрой теории и методики обучения информатике

Аннотация
В статье анализируется актуальность простой, но тем не менее очень важной прикладной задачи поиска подстроки в строке. Большинство источников предоставляют готовые алгоритмы и не стремятся дать пользователю понимание того, как этот алгоритм работает. В этой статье рассматриваются также и способы решения этой проблемы.
Ключевые слова: обработка данных, алгоритмы, программные среды, поиск подстроки, строки

Implementation of efficient string-searching algorithms

Tikhomirova Inga Nikolaevna
Yaroslavl State Pedagogical University named after K. D. Ushinsky
student

Vorontsova Ksenia Maksimovna
Yaroslavl State Pedagogical University named after K. D. Ushinsky
student

Kornilov Peter Anatolyevich
Yaroslavl State Pedagogical University named after K. D. Ushinsky
Candidate of Physical and Mathematical Sciences, Associate Professor, Head of the Department of Theory and Methods of Teaching Computer Science

Abstract
The article analyzes the relevance of a simple, but nevertheless very important applied problem of finding a substring in a string. Most sources provide ready-made algorithms and do not seek to give the user an understanding of how this algorithm works. This article also discusses ways to solve this problem.
Keywords: data processing, algorithms, software environments, substring search, strings

В учебной и профессиональной деятельности нам часто приходится работать с текстовой информацией в различных редакторах, что значительно упрощает работу с данным типом информации. Многие из встроенных в них алгоритмов используются нами почти ежедневно, и нам, вероятно, не нужно знать, как они работают. Но все кардинально меняется, когда речь заходит о разработке программ, предполагающих работу с текстовой информацией. Тогда нам становится просто необходимо знать эффективный алгоритм и понимать его работу [1].

К сожалению, довольно небольшое количество ресурсов предлагают информацию о том, как работают строковые алгоритмы, однако работа со значительными объемами данных требует эффективных и быстрых решений данных нам целей и задач. Одной из таких важных задач, требующих особого подхода, является поиск подстроки в строке [3]. Мы используем его везде, не только в текстовых редакторах, но и в поисковых системах, инструментах анализа и распознавания речи, извлечении знаний, архиваторах данных и многих других целях [2]. Все это свидетельствует об актуальности проблемы, которую мы рассматриваем в данной работе.

В результате анализа информации, представленной в различных источниках, было принято решение о разработке учебной программной среды (далее – УПС), основной целью которой является детальное описание и доступное объяснение работы алгоритмов поиска подстрок в строке.

В разработанной УПС рассматриваются два основных алгоритма: эффективный алгоритм вычисления Z-функции и алгоритм Кнута-Морриса-Пратта. Данные алгоритмы имеют важное практическое значение и являются начальным этапом усвоения и понимания того, что стандартная процедура поиска информации в тексте может быть оптимизирована и выходит далеко за рамки простого переборного алгоритма. Изучение алгоритма вычисления Z-функции и алгоритма Кнута-Морриса-Пратта является ключом к пониманию более сложных алгоритмов поиска информации, на которых основываются современные информационные системы. Именно данный факт является фундаментально значимым при изучении данных алгоритмов студентами IT-специальностей и школьниками, заинтересованными в изучении информатики.

Каждый алгоритм имеет подробное пошаговое объяснение, которое включает в себя визуальное представление и разбор каждого этапа алгоритма. Программа содержит множество интерактивных элементов, которые способствуют более прочному усвоению материала.

УПС имеет удобный интерфейс и включает в себя теоретическую часть, поэтапную реализацию каждого алгоритма, а также закрепление материала путем тестирования.

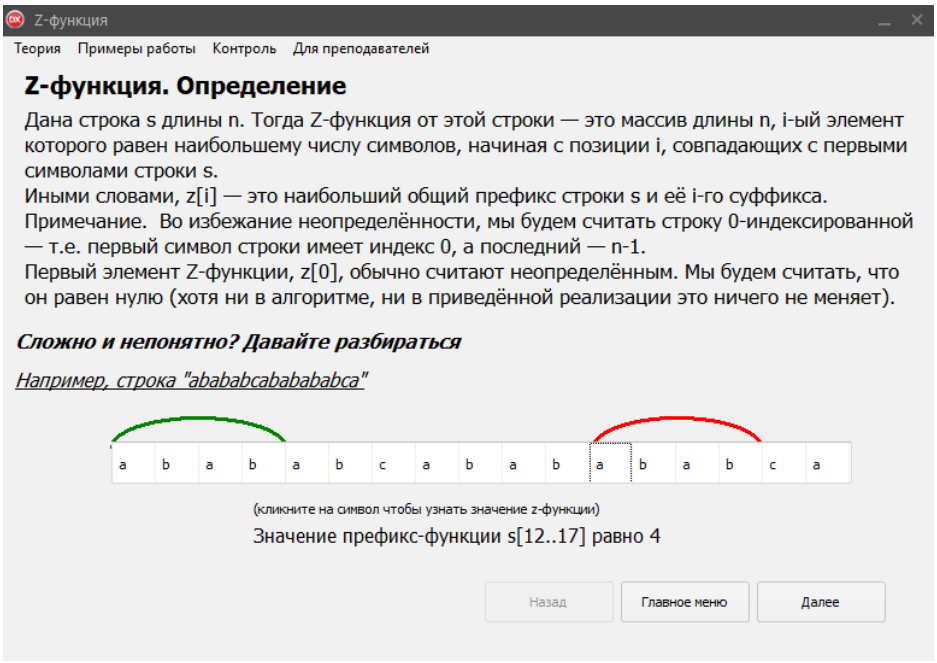

Рисунок 1 – Раздел «Теория». Определение Z-функции

Раздел «Теория» (рис. 1) включает в себя определение, алгоритмы реализации и их подробное описание. Также в данном разделе присутствуют интерактивные элементы демонстрации того, что такое префикс-функция, так как данный аспект не всегда понятен тем, кто занимается изучением данной темы. Особое внимание уделено тому, чтобы обучающийся понял, насколько более эффективны данные алгоритмы по сравнению с простым переборным алгоритмом.

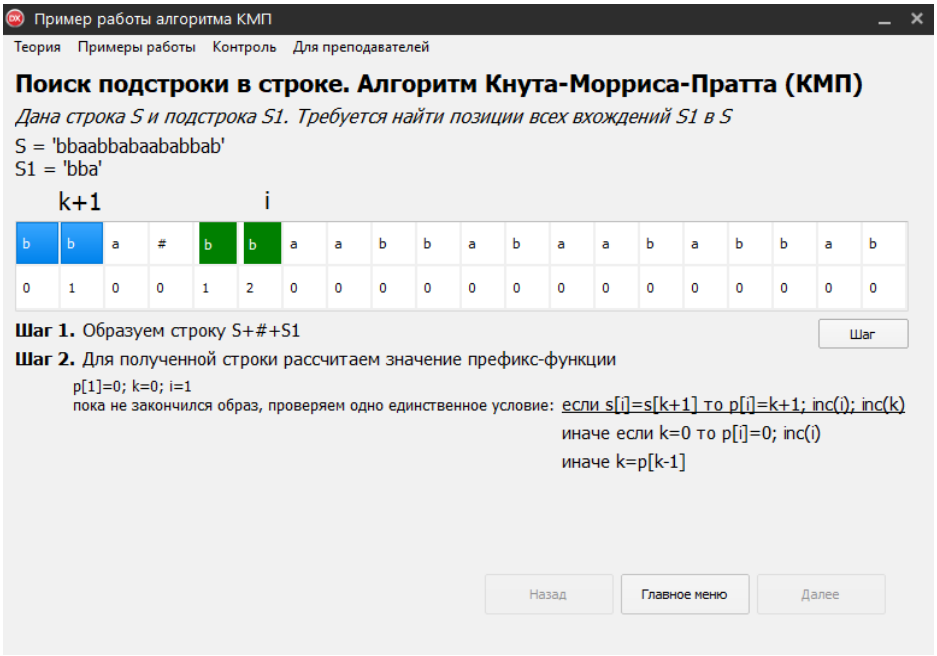

Рисунок 2 – Раздел «Примеры работ». Пример работы алгоритма Кнута-Морриса-Пратта

Раздел «Примеры работ» (рис. 2) обеспечивает пошаговый разбор каждого этапа алгоритма с использованием примеров. Каждый шаг алгоритма имеет текстовое объяснение и графическую демонстрацию. Данный подход даёт возможность пользователю УПС проследить за каждым шагом алгоритма и таким образом сформировать более прочное понимание его работы.

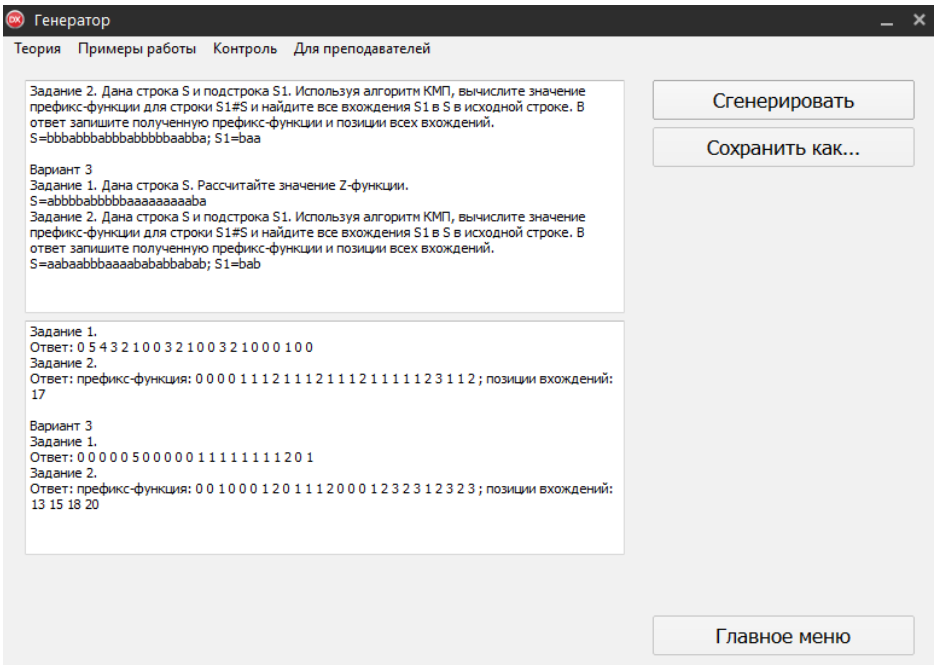

Рисунок 3 – Раздел «Для преподавателей». Генератор заданий

В разделе «Контроль» предлагается проверить прочность полученных знаний и пройти тест по данной теме. Этот раздел может быть использован не только обучающимися в целях самоконтроля, но и преподавателями во время занятия для проверки знаний изученного учащимися материала.

Раздел «Для преподавателей» (рис. 3) включает в себя генератор вариантов проверочных работ. Вариант проверочной работы включает в себя два задания на вычисления Z-функции и префикс-функции. Каждый вариант строго индивидуален, так как все задания генерируются программой.

Чтобы сэкономить время преподавателей, УПС автоматически рассчитывает ответы на сгенерированные задания и формирует текстовые файлы, которые полностью готовы к печати.

Таким образом, разработанная программа полностью готова к использованию в школах и других образовательных учреждениях. В ней отражены все ключевые моменты, важные для понимания решения такой, казалось бы, простой, но очень важной прикладной задачи поиска подстроки в строке.

Библиографический список

1. Окулов С.М. Алгоритмы обработки строк. М.: БИНОМ. Лаборатория знаний, 2015. 256 с.
2. Солдатова Г.П., Татаринов А.А., Болдырихин Н.В. Основные алгоритмы поиска подстроки в строке // Журнал «Academy». 2018. С. 8-10. URL: https://cyberleninka.ru/article/n/osnovnye-algoritmy-poiska-podstroki-v-stroke (дата обращения: 08.03.2020).
3. Царев Р.Ю., Царева Е.А., Черниговский А.С. Комбинированный алгоритм поиска образа в строке // Журнал Сибирского федерального университета. Техника и технологии. 2017. С. 126-135. URL: https://cyberleninka.ru/article/n/kombinirovannyy-algoritm-poiska-obraza-v-stroke/viewer (дата обращения: 07.03.2020).