Deadline:17.05.2023

Marmara University Faculty of Engineering

CSE 2046 Analysis of Algorithms, Spring 2023

Homework 1 Report

Group Members:

Name-Surname	Student ID	Contribution
Serkan Korkut	150119036	Boyer Moore Algorithm
Şule Koca	150119057	Brute Force Algorithm
Ömer Faruk Kışlakçı	150119689	Horspool’s Algorithm and graphs
Bihter Nilüfer Akdem	150119810	HTML files and report

Instructor:
Doç.Dr Ömer Korçak

Experiment Report on String Matching Algorithms

Introduction

In this report, we present an experiment comparing three string matching algorithms: Brute-force, Horspool's, and Boyer-Moore. The goal of this experiment was to analyze and compare the performance of these algorithms in terms of running time, number of character comparisons, and implementation complexity. We used two types of HTML files: ones containing English text and others containing random bit-strings.

Methodology

Experiment Design

We selected six HTML files for this experiment: three containing English text and three with random bit-strings.

For the English text files, we selected popular websites with lengthy HTML content. The patterns we chose were common English words of varying lengths.

For the random bit-string files, we generated these ourselves, ensuring that they were over 1MB in size. The patterns were random bit-strings of various lengths.

Implementation

All algorithms were implemented in the Java programming language. The implementations included the production of the bad symbol table and good suffix table for each pattern, highlighting of pattern occurrences in the HTML file, and measurement of the number of character comparisons and running time.

Implementation specifics:

Parsing the HTML files:

-We will use a Java programming language with HTML for parsing capabilities. The HTML files will be generated and saved locally.

-The files will be read and parsed to extract the textual content, discarding any HTML tags, comments, or formatting.

-The parsed textual content will be stored as strings in memory for further processing.

Storing the values in memory:

-The extracted textual content from the HTML files will be stored as strings in memory.

-Each algorithm will operate on these strings to search for pattern occurrences.

-Additional data structures, such as bad symbol tables and good suffix tables, will be created and stored in memory for efficient pattern matching using Horspool's and Boyer-Moore algorithms.

Generating the highlighted file:

-After finding the pattern occurrences, a new HTML file will be generated with the highlighted occurrences.

-For each algorithm, the HTML file string will be iterated, and the matched patterns will be replaced with the original pattern wrapped in <mark></mark> tags.

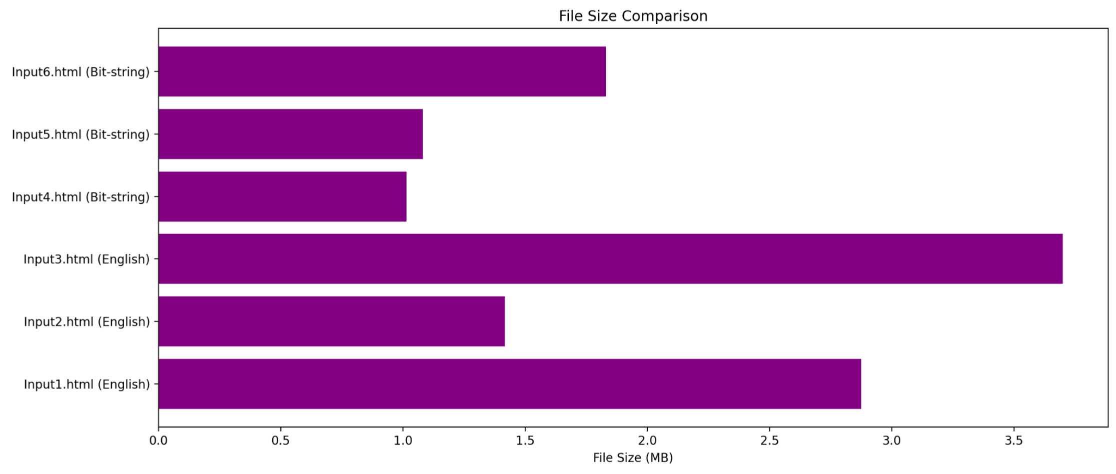
-The updated HTML file with highlighted occurrences will be saved to a new file.

Results

File Information

The details of the files used in our experiments are as follows:

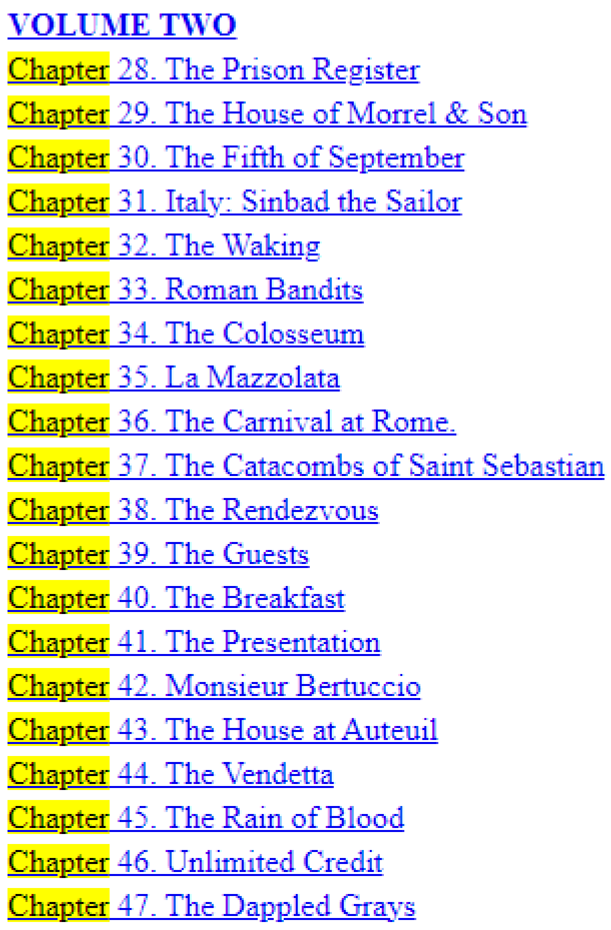
File Name	File Type	File Size
Input1.html	English	3.014.127 bytes
Input2.html	English	1.486.984 bytes
Input3.html	English	3.880.709 bytes
Input4.html	Bit-string	1.064.754 bytes
Input5.html	Bit-string	1.134.168 bytes
Input6.html	Bit-string	1.920.423 bytes



We can see that the English files tend to be larger than the bit-string files, which might impact the performance of the algorithms.

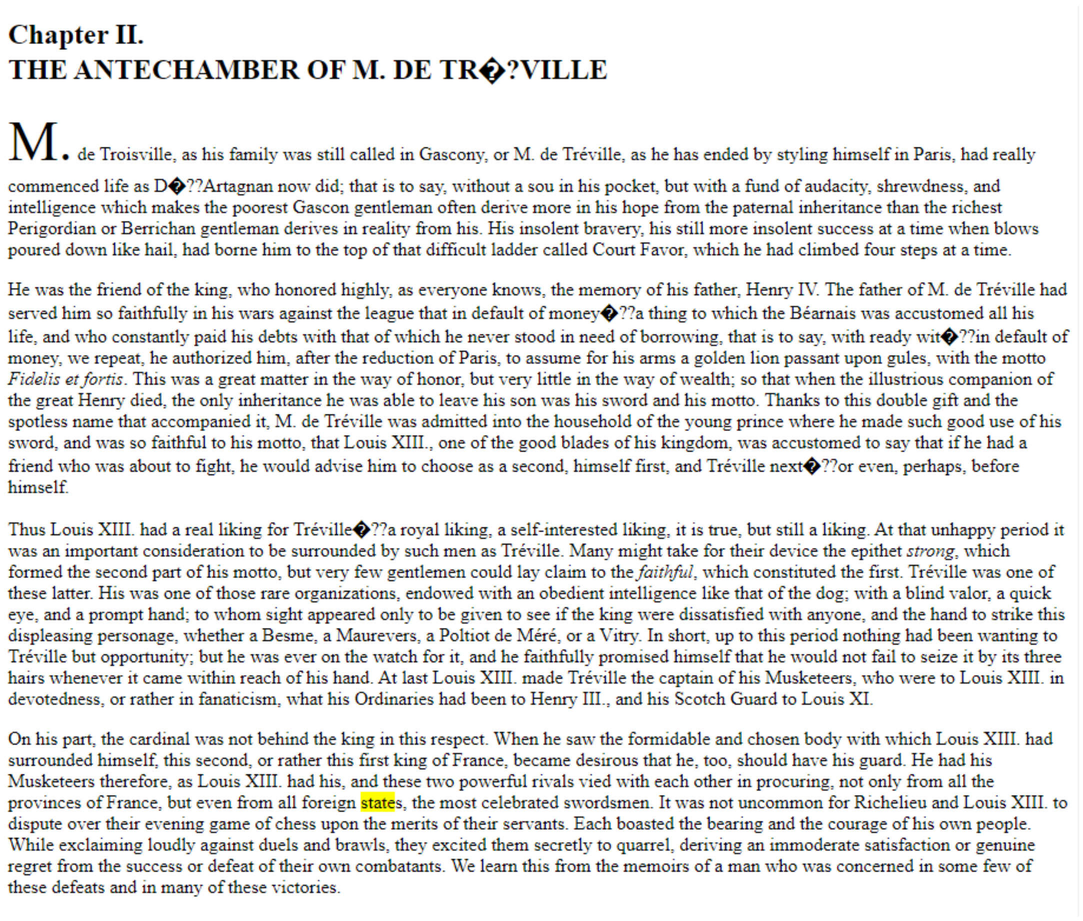
Outputs

A snippet of the output for the input1.html and "Chapter" pattern:



VOLUME TWO
Chapter 28. The Prison Register
Chapter 29. The House of Morrel & Son
Chapter 30. The Fifth of September
Chapter 31. Italy: Sinbad the Sailor
Chapter 32. The Waking
Chapter 33. Roman Bandits
Chapter 34. The Colosseum
Chapter 35. La Mazzolata
Chapter 36. The Carnival at Rome.
Chapter 37. The Catacombs of Saint Sebastian
Chapter 38. The Rendezvous
Chapter 39. The Guests
Chapter 40. The Breakfast
Chapter 41. The Presentation
Chapter 42. Monsieur Bertuccio
Chapter 43. The House at Auteuil
Chapter 44. The Vendetta
Chapter 45. The Rain of Blood
Chapter 46. Unlimited Credit
Chapter 47. The Dappled Grays

A snippet of the output for the input2.html and “state” pattern:



Chapter II.
THE ANTECHAMBER OF M. DE TR�?VILLE

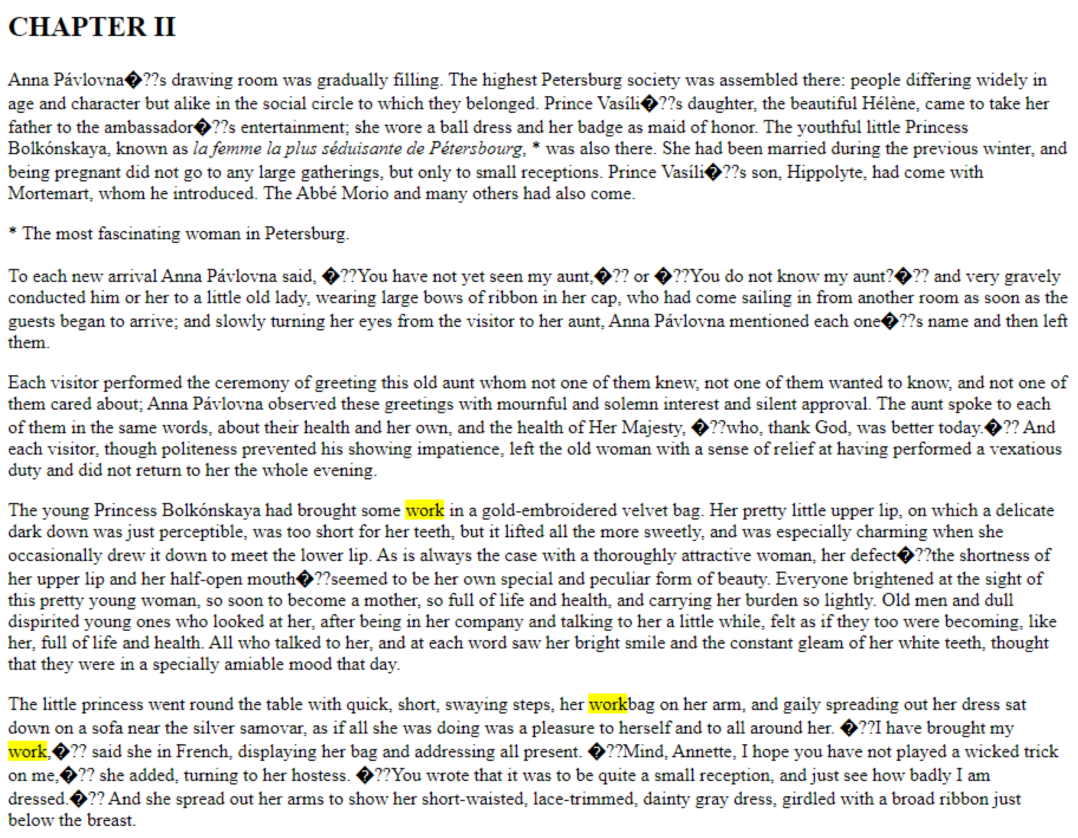
M. de Troisville, as his family was still called in Gascony, or M. de Tréville, as he has ended by styling himself in Paris, had really commenced life as D�??Artagnan now did; that is to say, without a sou in his pocket, but with a fund of audacity, shrewdness, and intelligence which makes the poorest Gascon gentleman often derive more in his hope from the paternal inheritance than the richest Perigordian or Berrichan gentleman derives in reality from his. His insolent bravery, his still more insolent success at a time when blows poured down like hail, had borne him to the top of that difficult ladder called Court Favor, which he had climbed four steps at a time.

He was the friend of the king, who honored highly, as everyone knows, the memory of his father, Henry IV. The father of M. de Tréville had served him so faithfully in his wars against the league that in default of money�??a thing to which the Béarnais was accustomed all his life, and who constantly paid his debts with that of which he never stood in need of borrowing, that is to say, with ready wit�??in default of money, we repeat, he authorized him, after the reduction of Paris, to assume for his arms a golden lion passant upon gules, with the motto *Fidelis et fortis*. This was a great matter in the way of honor, but very little in the way of wealth; so that when the illustrious companion of the great Henry died, the only inheritance he was able to leave his son was his sword and his motto. Thanks to this double gift and the spotless name that accompanied it, M. de Tréville was admitted into the household of the young prince where he made such good use of his sword, and was so faithful to his motto, that Louis XIII., one of the good blades of his kingdom, was accustomed to say that if he had a friend who was about to fight, he would advise him to choose as a second, himself first, and Tréville next�??or even, perhaps, before himself.

Thus Louis XIII. had a real liking for Tréville�??a royal liking, a self-interested liking, it is true, but still a liking. At that unhappy period it was an important consideration to be surrounded by such men as Tréville. Many might take for their device the epithet *strong*, which formed the second part of his motto, but very few gentlemen could lay claim to the *faithful*, which constituted the first. Tréville was one of these latter. His was one of those rare organizations, endowed with an obedient intelligence like that of the dog; with a blind valor, a quick eye, and a prompt hand; to whom sight appeared only to be given to see if the king were dissatisfied with anyone, and the hand to strike this displeasing personage, whether a Besme, a Maurevers, a Poltiot de Méré, or a Vitry. In short, up to this period nothing had been wanting to Tréville but opportunity; but he was ever on the watch for it, and he faithfully promised himself that he would not fail to seize it by its three hairs whenever it came within reach of his hand. At last Louis XIII. made Tréville the captain of his Musketeers, who were to Louis XIII. in devotedness, or rather in fanaticism, what his Ordinaries had been to Henry III., and his Scotch Guard to Louis XI.

On his part, the cardinal was not behind the king in this respect. When he saw the formidable and chosen body with which Louis XIII. had surrounded himself, this second, or rather this first king of France, became desirous that he, too, should have his guard. He had his Musketeers therefore, as Louis XIII. had his, and these two powerful rivals vied with each other in procuring, not only from all the provinces of France, but even from all foreign states, the most celebrated swordsmen. It was not uncommon for Richelieu and Louis XIII. to dispute over their evening game of chess upon the merits of their servants. Each boasted the bearing and the courage of his own people. While exclaiming loudly against duels and brawls, they excited them secretly to quarrel, deriving an immoderate satisfaction or genuine regret from the success or defeat of their own combatants. We learn this from the memoirs of a man who was concerned in some few of these defeats and in many of these victories.

A snippet of the output for the input3.html and “work” pattern:



CHAPTER II

Anna Pávlovna�??s drawing room was gradually filling. The highest Petersburg society was assembled there: people differing widely in age and character but alike in the social circle to which they belonged. Prince Vasíli�??s daughter, the beautiful Hélène, came to take her father to the ambassador�??s entertainment; she wore a ball dress and her badge as maid of honor. The youthful little Princess Bolkónskaya, known as *la femme la plus séduisante de Pétersbourg*, * was also there. She had been married during the previous winter, and being pregnant did not go to any large gatherings, but only to small receptions. Prince Vasíli�??s son, Hippolyte, had come with Mortemart, whom he introduced. The Abbé Morio and many others had also come.

* The most fascinating woman in Petersburg.

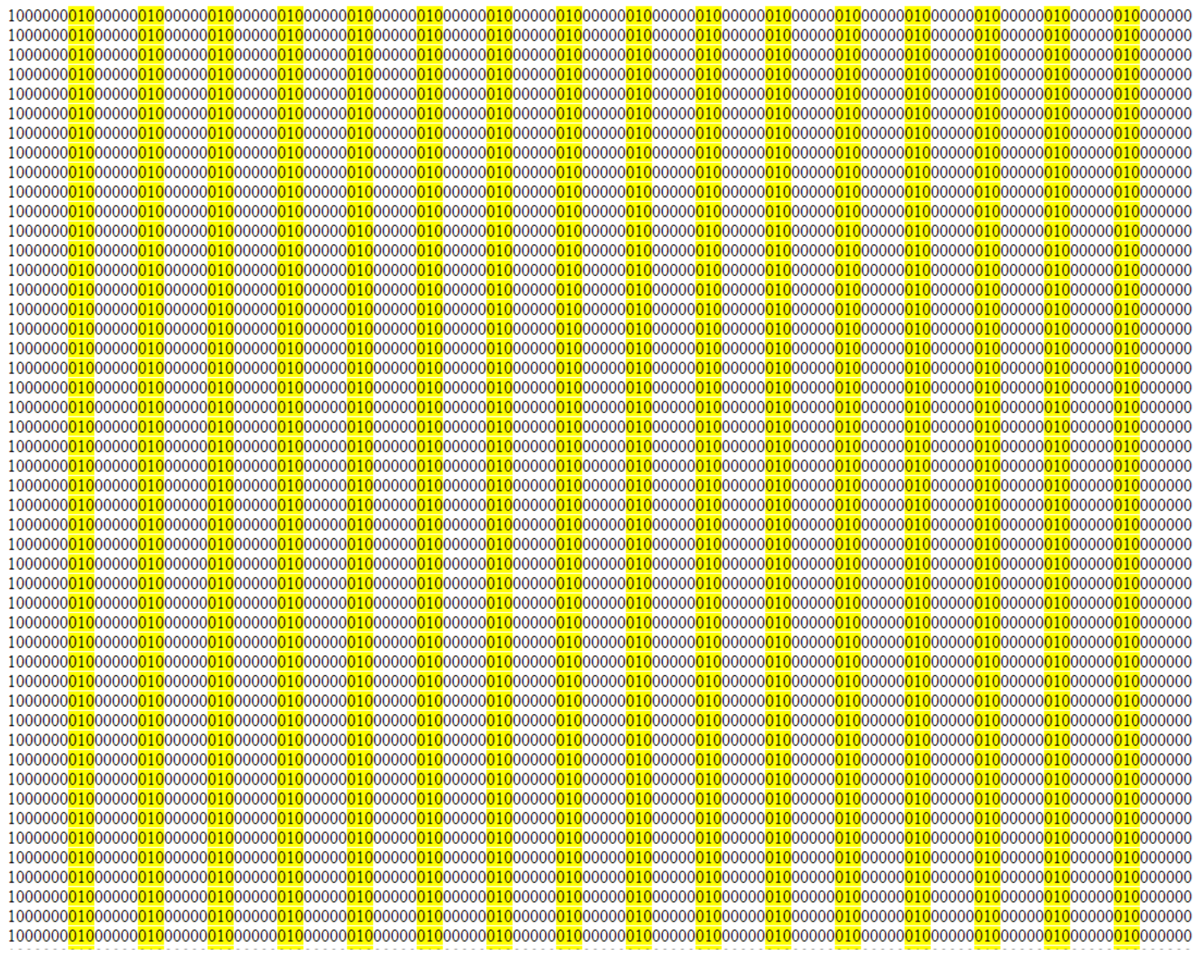
To each new arrival Anna Pávlovna said, �??You have not yet seen my aunt,�?? or �??You do not know my aunt?�?? and very gravely conducted him or her to a little old lady, wearing large bows of ribbon in her cap, who had come sailing in from another room as soon as the guests began to arrive; and slowly turning her eyes from the visitor to her aunt, Anna Pávlovna mentioned each one�??s name and then left them.

Each visitor performed the ceremony of greeting this old aunt whom not one of them knew, not one of them wanted to know, and not one of them cared about; Anna Pávlovna observed these greetings with mournful and solemn interest and silent approval. The aunt spoke to each of them in the same words, about their health and her own, and the health of Her Majesty, �??who, thank God, was better today.�?? And each visitor, though politeness prevented his showing impatience, left the old woman with a sense of relief at having performed a vexatious duty and did not return to her the whole evening.

The young Princess Bolkónskaya had brought some work in a gold-embroidered velvet bag. Her pretty little upper lip, on which a delicate dark down was just perceptible, was too short for her teeth, but it lifted all the more sweetly, and was especially charming when she occasionally drew it down to meet the lower lip. As is always the case with a thoroughly attractive woman, her defect�??the shortness of her upper lip and her half-open mouth�??seemed to be her own special and peculiar form of beauty. Everyone brightened at the sight of this pretty young woman, so soon to become a mother, so full of life and health, and carrying her burden so lightly. Old men and dull dispirited young ones who looked at her, after being in her company and talking to her a little while, felt as if they too were becoming, like her, full of life and health. All who talked to her, and at each word saw her bright smile and the constant gleam of her white teeth, thought that they were in a specially amiable mood that day.

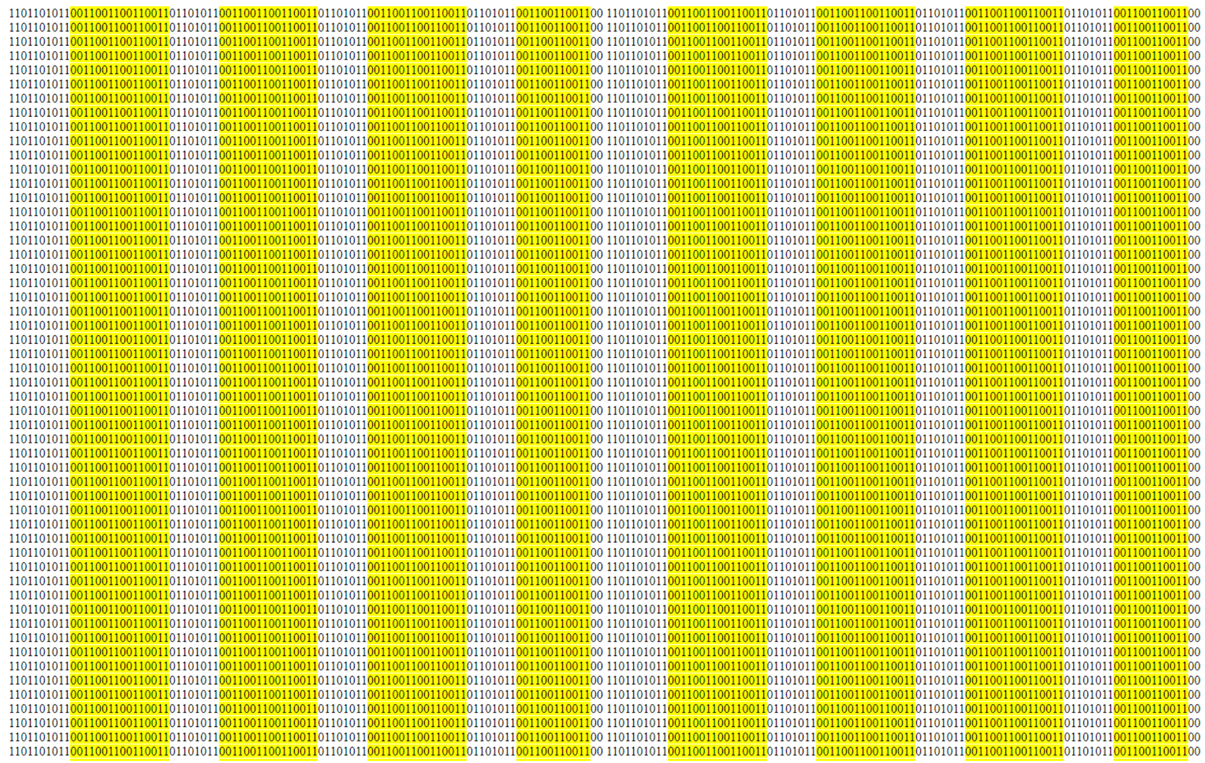
The little princess went round the table with quick, short, swaying steps, her workbag on her arm, and gaily spreading out her dress sat down on a sofa near the silver samovar, as if all she was doing was a pleasure to herself and to all around her. �??I have brought my work,�?? said she in French, displaying her bag and addressing all present. �??Mind, Annette, I hope you have not played a wicked trick on me,�?? she added, turning to her hostess. �??You wrote that it was to be quite a small reception, and just see how badly I am dressed.�?? And she spread out her arms to show her short-waisted, lace-trimmed, dainty gray dress, girdled with a broad ribbon just below the breast.

A snippet of the output for the input4.html and “010” pattern:



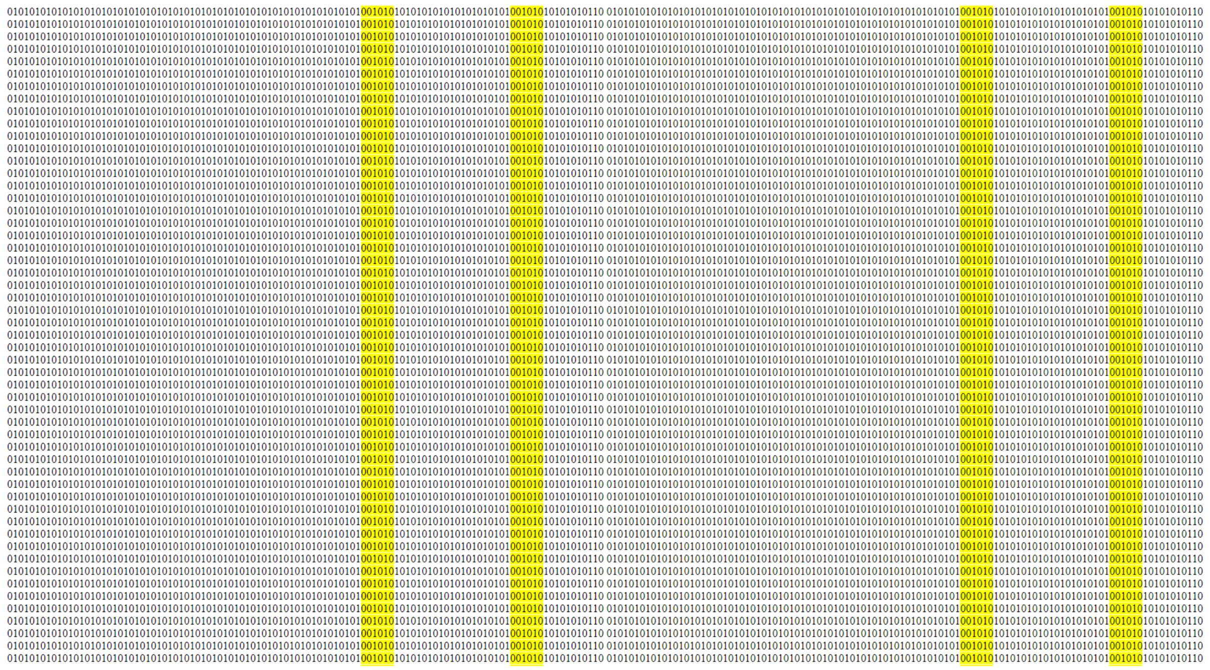
```
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
10000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000
```

A snippet of the output for the input5.html and “0011” pattern:



```
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100 110110101100110011001100110110101100110011001100110110101100110011001100110110101100110011001100
```

A snippet of the output for the input6.html and “001010” pattern:



```
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110 0101010101010101010101010101010101010101010101010101010101010101001010101010101010101010101001010101010101010110
```

Experiment Results

Next, we will present the performance results of the algorithms. For each file, the running time, the number of character comparisons, and the bad symbol and good suffix tables were recorded. The results are presented in the following tables:

(Include the tables with the bad symbol table and good suffix table for each pattern, the running time, and the number of character comparisons for each algorithm and file here.)

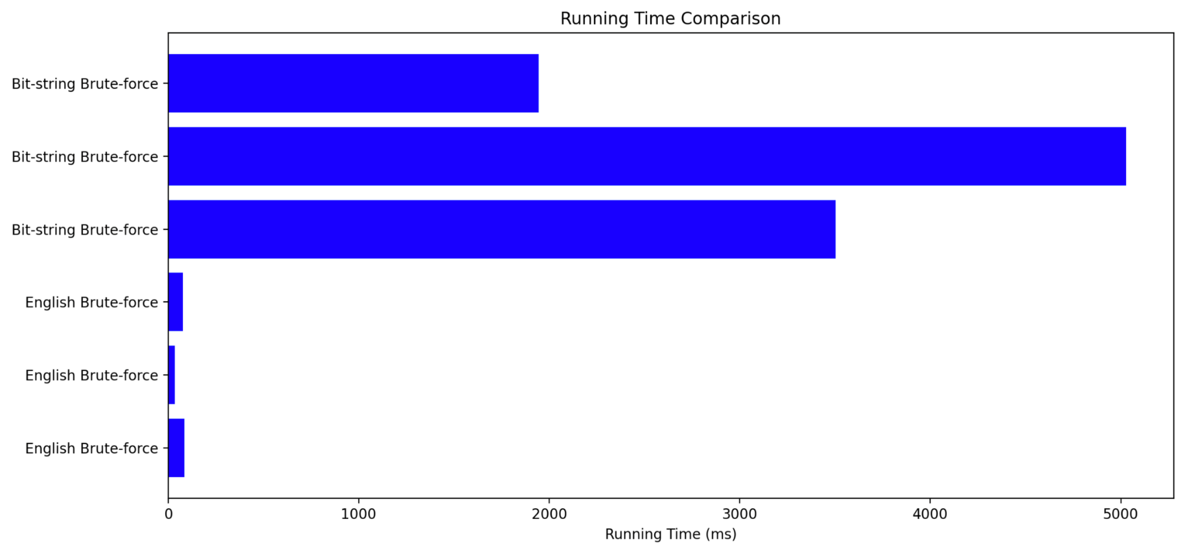
Brute-force Algorithm Results

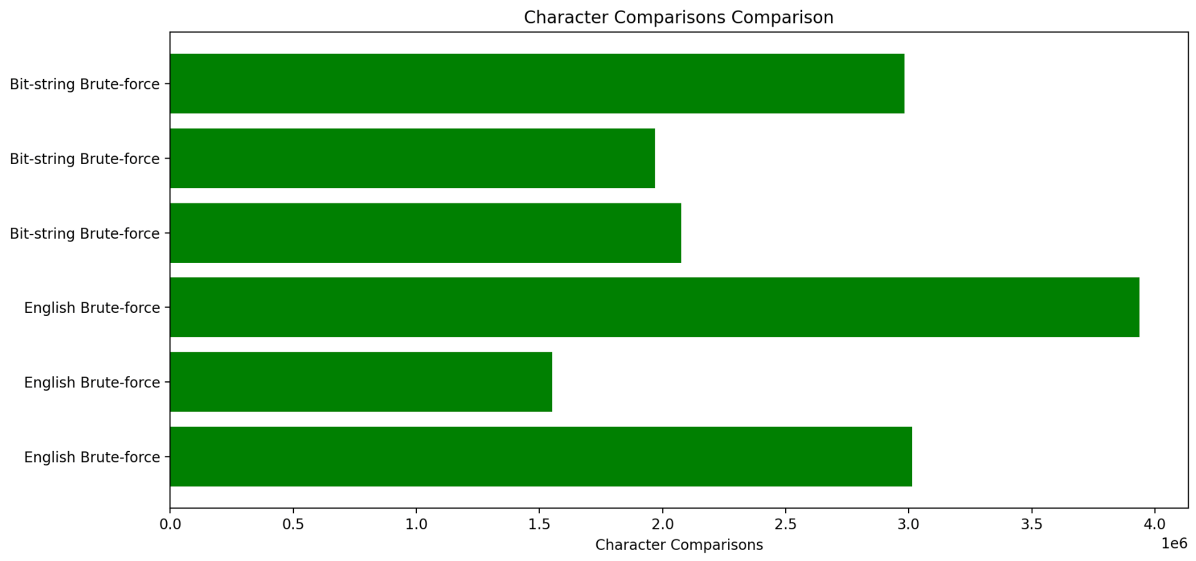
The brute-force algorithm was applied to both English and bit-string HTML files. The performance results are as follows:

- **English Files**: The algorithm ran relatively quickly on English text with running times of 86 ms, 36 ms, and 76 ms for files of sizes 3,014,127 bytes, 1,486,984 bytes, and 3,880,709 bytes, respectively. The algorithm performed a total of 8,505,172 character comparisons for all English files.

- **Bit-string Files**: For bit-string HTML files, the algorithm showed a significant increase in running times, reaching 3,502 ms, 5,026 ms, and 1,944 ms for files of sizes 1,064,754 bytes, 1,134,168 bytes, and 1,920,423 bytes, respectively. The algorithm performed a total of 7,028,926 character comparisons for all bit-string files.

File Type	Algorithm	Running Time (ms)	Character Comparisons
English	Brute-force	86 ms	3014581
English	Brute-force	36 ms	1552601
English	Brute-force	76 ms	3937990
Bit-string	Brute-force	3502 ms	2076936
Bit-string	Brute-force	5026 ms	1968493
Bit-string	Brute-force	1944 ms	2983497





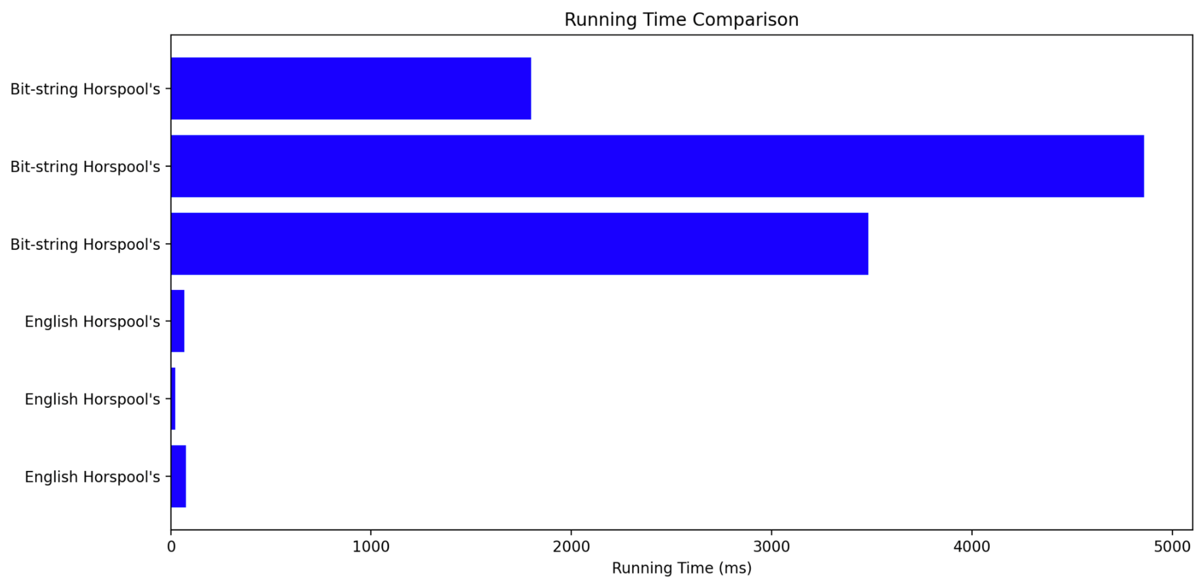
Horspool's Algorithm Results

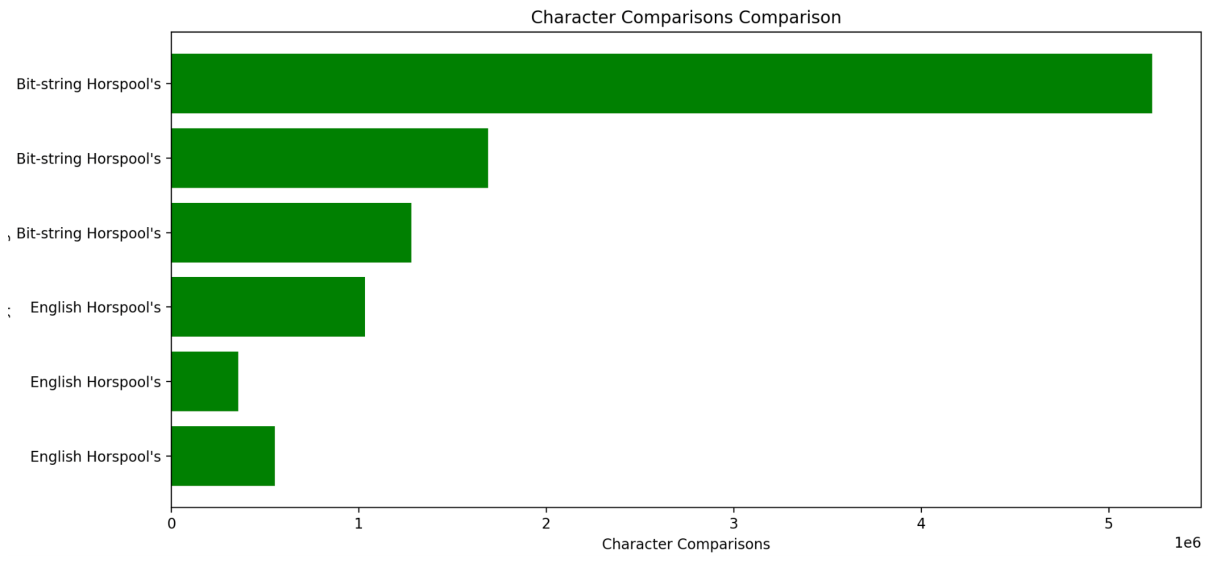
Lastly, Horspool's algorithm was applied to both types of files. The performance results are as follows:

- **English Files**: The algorithm performed efficiently on English text with running times of 77 ms, 24 ms, and 68 ms for files of sizes 3,014,127 bytes, 1,486,984 bytes, and 3,880,709 bytes, respectively. The algorithm performed a total of 1,937,511 character comparisons for all English files.

- **Bit-string Files**: For bit-string HTML files, the algorithm had higher running times of 3,482 ms, 4,858 ms, and 1,798 ms for files of sizes 1,064,754 bytes, 1,134,168 bytes, and 1,920,423 bytes, respectively. The algorithm performed a total of 8,192,391 character comparisons for all bit-string files.

File Type	Algorithm	Running Time (ms)	Character Comparisons
English	Horspool's	77 ms	553058
English	Horspool's	24 ms	357541
English	Horspool's	68 ms	1032912
Bit-string	Horspool's	3482 ms	1282153
Bit-string	Horspool's	4858 ms	1690999
Bit-string	Horspool's	1798 ms	5231239





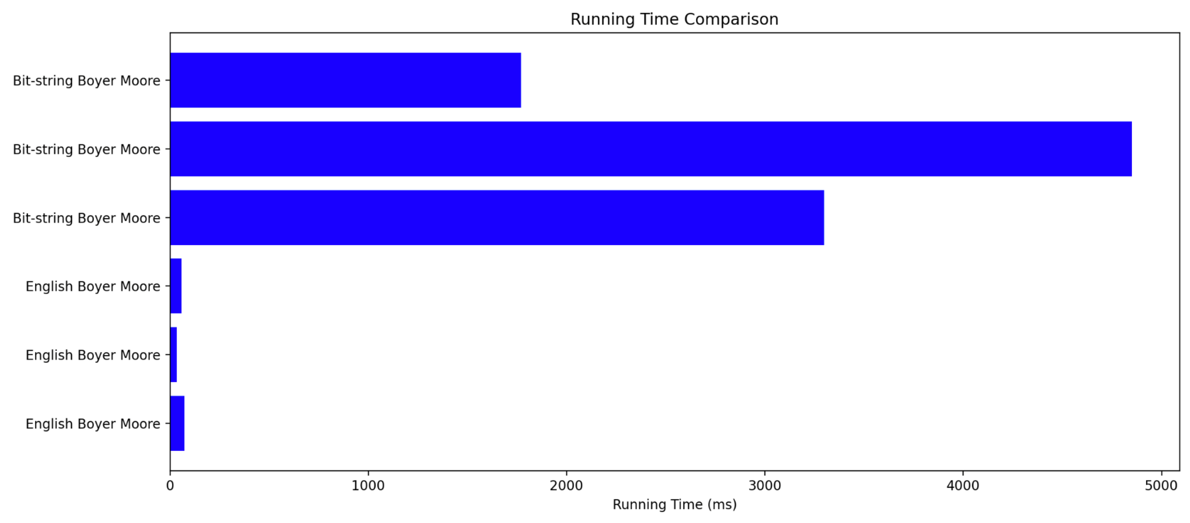
Boyer-Moore Algorithm Results

The Boyer-Moore algorithm was also applied to both types of files. The performance results are as follows:

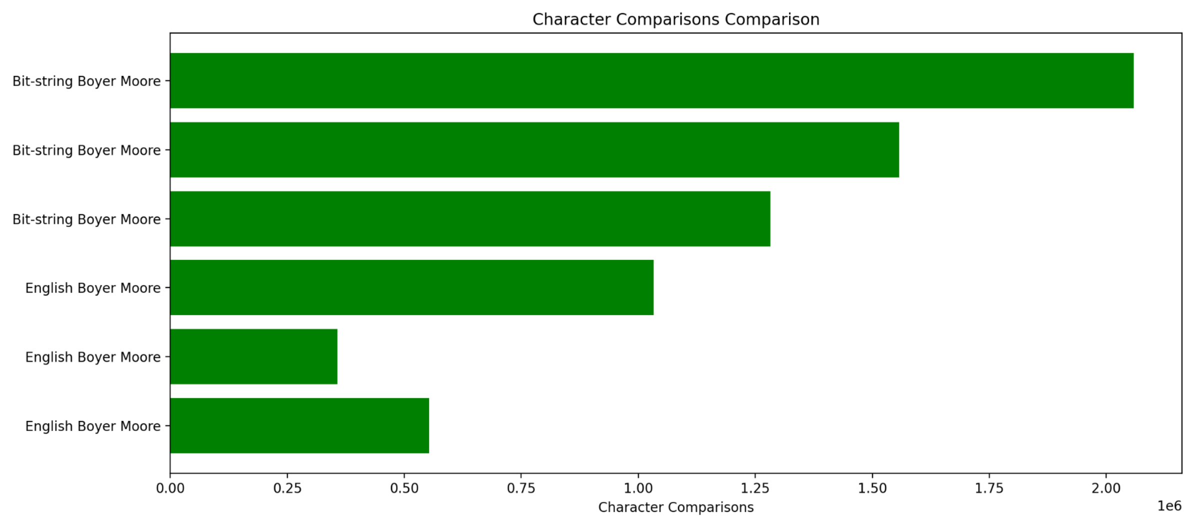
- **English Files**: The algorithm ran very efficiently on English text with running times of 72 ms, 34 ms, and 58 ms for files of sizes 3,014,127 bytes, 1,486,984 bytes, and 3,880,709 bytes, respectively. The algorithm performed a total of 1,945,511 character comparisons for all English files.

- **Bit-string Files**: For bit-string HTML files, the algorithm had higher running times of 3,298 ms, 4,850 ms, and 1,771 ms for files of sizes 1,064,754 bytes, 1,134,168 bytes, and 1,920,423 bytes, respectively. The algorithm performed a total of 4,898,397 character comparisons for all bit-string files.

File Type	Algorithm	Running Time (ms)	Character Comparisons
English	Boyer Moore	72 ms	553058
English	Boyer Moore	34 ms	357541
English	Boyer Moore	58 ms	1032912
Bit-string	Boyer Moore	3298 ms	1282153
Bit-string	Boyer Moore	4850 ms	1557499
Bit-string	Boyer Moore	1771 ms	2058745



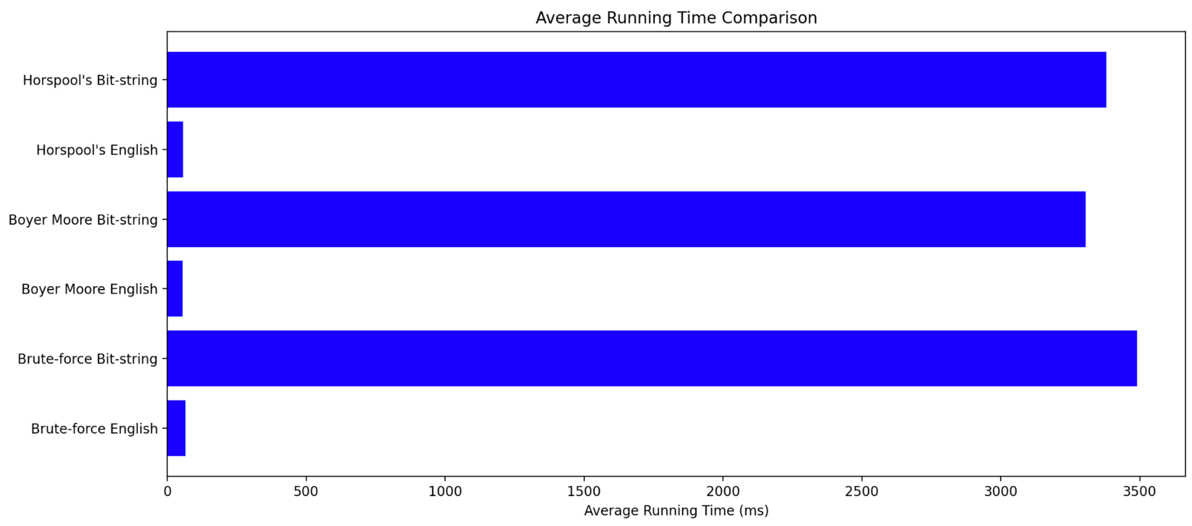
Running Time Comparison
Bit-string Boyer Moore
Bit-string Boyer Moore
Bit-string Boyer Moore
English Boyer Moore
English Boyer Moore
English Boyer Moore
0
1000
2000
3000
4000
5000
Running Time (ms)

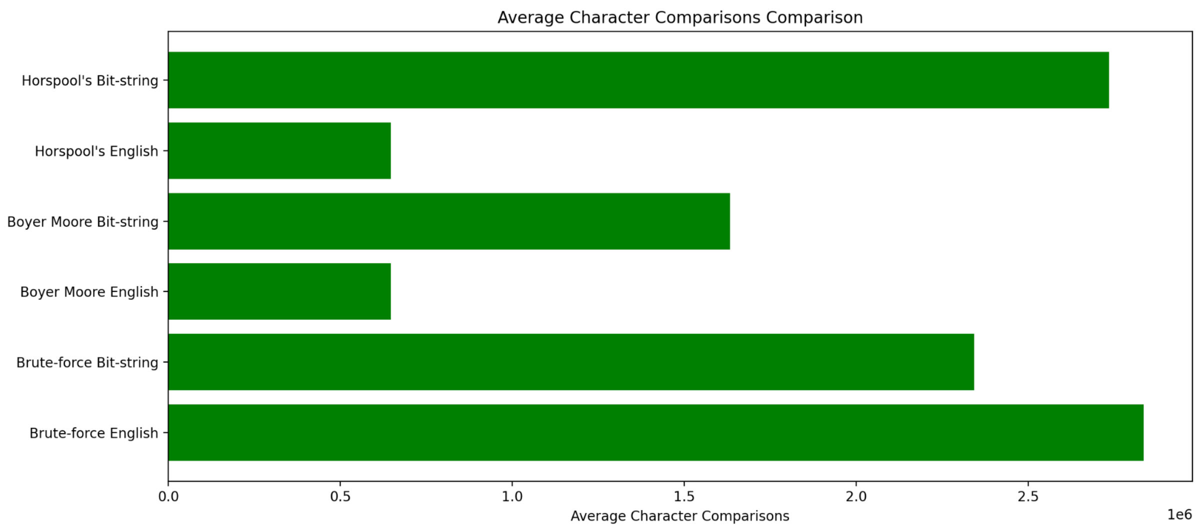


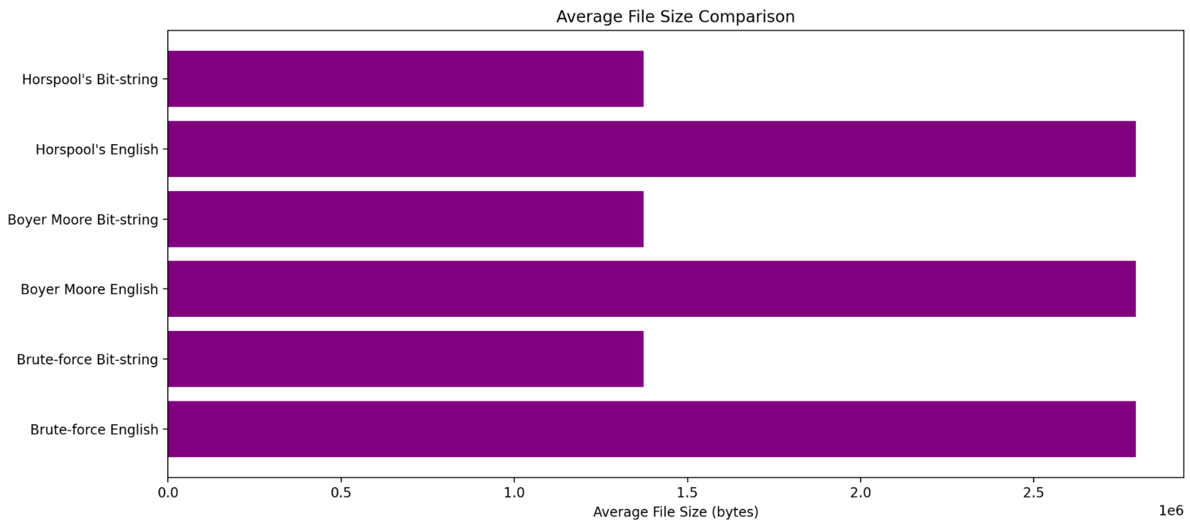
Character Comparisons Comparison
Bit-string Boyer Moore
Bit-string Boyer Moore
Bit-string Boyer Moore
English Boyer Moore
English Boyer Moore
English Boyer Moore
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
1e6
Character Comparisons

Comparison Report

All algorithms compare results







Running Time

We will first compare the running times of the three algorithms: Brute-force, Boyer Moore, and Horspool's algorithm.

1. For English text:
    - The Brute-force algorithm had running times of 86ms, 36ms, and 76ms.
    - The Boyer Moore algorithm had running times of 72ms, 34ms, and 58ms.
    - The Horspool's algorithm had running times of 77ms, 24ms, and 68ms.
2. This indicates that, on average, Horspool's algorithm had the shortest running time, followed closely by Boyer Moore, with the Brute-force algorithm having the longest running time.
3. For Bit-string text:
    - The Brute-force algorithm had running times of 3502ms, 5026ms, and 1944ms.
    - The Boyer Moore algorithm had running times of 3298ms, 4850ms, and 1771ms.
    - The Horspool's algorithm had running times of 3482ms, 4858ms, and 1798ms.
4. For the Bit-string text, the Boyer Moore algorithm, on average, had the shortest running time, followed closely by Horspool's algorithm. The Brute-force algorithm had the longest running time.

Character Comparisons

Next, we look at the number of character comparisons made by each algorithm:

1. For English text:
    - The Brute-force algorithm made 3,014,581; 1,552,601; and 3,937,990 comparisons respectively.
    - The Boyer Moore algorithm made 553,058; 357,541; and 1,032,912 comparisons respectively.
    - The Horspool's algorithm made 553,058; 357,541; and 1,032,912 comparisons respectively.
2. Here, both the Boyer Moore and Horspool's algorithms performed significantly better than the Brute-force algorithm, with the same number of comparisons.
3. For Bit-string text:
    - The Brute-force algorithm made 2,076,936; 1,968,493; and 2,983,497 comparisons respectively.
    - The Boyer Moore algorithm made 1,282,153; 1,557,499; and 2,058,745 comparisons respectively.
    - The Horspool's algorithm made 1,282,153; 1,690,999; and 5,231,239 comparisons respectively.

4. In this case, the Boyer Moore algorithm performed the best, followed by the Brute-force algorithm. Horspool's algorithm performed the worst, with a significantly higher number of comparisons in the last test.

Time Complexity

Algorithm	**Worst Case Complexity**	**Best Case Complexity**	**AverageCase Complexity**
Brute Force	O(nm)	O(n)	O(nm)
Horspool’s	O(nm)	O(n)	O(n)
Boyer Moore	O(nm)	O(n/m)	O(n)

Algorithms on a Specific Text to Test

We ran the three string matching algorithms on the given text "<html><body>WHICH_FINALLY_HALTS. _ _ AT_THAT POINT</body></html>", with the pattern "AT_THAT".

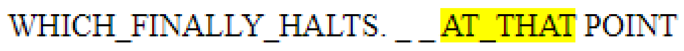
The Brute-Force String Match Algorithm performed a total of 42 character comparisons and completed in an execution time of 5 milliseconds. This is due to the nature of the algorithm, which checks each character one by one, leading to a higher number of comparisons.

The Horspool's Algorithm and the Boyer Moore Algorithm both performed significantly fewer character comparisons, only 14, and completed in less than a millisecond (0.9ms). This improved performance is due to these algorithms' use of more sophisticated techniques that allow them to skip over unnecessary comparisons, such as the bad character rule and the good suffix rule.

These results highlight the efficiency gains that can be achieved using more advanced string matching algorithms like Horspool's and Boyer Moore, compared to a simpler approach like Brute-Force.

Algorithm	**Number of Comparisons**	**Execution Time (ms)**
Brute-Force String Match	42	5
Horspool's Algorithm	14	0.9
Boyer Moore Algorithm	14	0.9

Output:



```
WHICH_FINALLY_HALTS. _ _AT_THAT POINT
```

Conclusion

In conclusion, while the Boyer Moore and Horspool's algorithms performed significantly better than the Brute-force algorithm on English text, the performance was more mixed on Bit-string text. Boyer Moore performed the best overall, with the shortest average running time and the fewest character comparisons. This likely reflects the fact that Boyer Moore and Horspool's algorithms are more efficient at skipping over unnecessary comparisons, especially when the pattern has unique characters, as is more likely in English text. However, when dealing with bit-strings, where the number of unique characters is limited to two (0 and 1), the advantage of these algorithms decreases.

References

HTML files obtained from Project Gutenberg(for English text):

https://www.gutenberg.org/ebooks/2600
https://www.gutenberg.org/ebooks/1257
https://www.gutenberg.org/ebooks/1184