

# Tandem Repeat Detection

This project implements [Main & Lorentz's \(1985\)](#) algorithm for tandem repeat/square detection.

`make` generates the following executables:

Executable	Time Complexity	Summary
<code>./tr.o</code>	$O(n \log n)$	Their main algorithm
<code>./tr_fast.o</code>	$O(n)$	Speedup to above through preprocessing
<code>./tr_bf.o</code>	$O(n^2)$	Naive approach for comparison

## Quickstart

In addition to running each executable, you can run testing scripts:

```
bash test_scripts/test.sh
bash test_scripts/test_manual.sh
bash test_scripts/test_time_complexity.sh
```

Script	Description	Optional Argument
<code>test.sh</code>	Compares $O(n \log n)$ algorithm to $O(n^2)$ on randomly generated strings	<code>./tr_fast.o</code> to test $O(n)$ alg
<code>test_manual.sh</code>	Runs $O(n \log n)$ algorithm on manually created strings with known answers	<code>./tr_fast.o</code> to test $O(n)$ alg
<code>test_time_complexity.sh</code>	Times $O(n \log n)$ and $O(n)$ algorithms on long generated squarefree strings	some integer $n$ to recreate input strings of length $n$

# Executable Input Format

---

Each executable reads data from stdin following this format:

```
num_tests start_char alphabet_size
query_string_for_test_1
query_string_for_test_2
...
```

and outputs 'YES' for each test case that contains a tandem repeat, and 'NO' otherwise.  
See `tests/in/100w_5n_4k.in` for a simple example.

The alphabet is defined as the `alphabet_size` ASCII characters beginning from the `start_char`.  
For example, `start_char = 'a'` and `alphabet_size = 4` gives an alphabet of {'a', 'b', 'c', 'd'}.

Note the following input requirements, which are not validated but must be followed:

- All query strings must only use characters from the defined alphabet
- The smallest allowed `start_char` is "!"
- The alphabet size should be defined so that the largest character doesn't go beyond "~"
- The alphabet size should be greater than 1. I did not bother to implement the trivial handling of a one letter alphabet.