Boyer Moore

Description

The main idea of this algorithm is to compare from right to left untill you find a mismatch. On a mismatch you shift the needle so that the mismatched character in the haystack lines up with the last occurence of that character in the needle, or shift the needle all the way past the character if it is not in the needle.

This algorithm performs well for large alphabets. This is because in a large alphabet it is more likely for a letter to appear in the haystack, but not the needle. In those cases you can shift by the entire length of the needle.

This is quicker in practice than brute force, but worst case it is still $O(nm)$. There are some optimizations that can make Booyer-Moore $O(n+m)$ in every case, but we will not cover them in this course.

Last Function

The last function determines how far you need to shift when you see each character. It is generated by creating a map for each letter. Then, for every character `c` in the needle, `map[c] = max(needle.length() - needle.lastIndexOf(c) - 1, 1)`. If a character `c` is not in the needle, then `map[c] = needle.length()`.

Algorithm

To run the algorithm line up the first character in the needle to the first character in the haystack. Then start comparing from the end of the needle. When there is a mismatch on character `c` in the haystack, shift the needle to the right by `map[c]` places. Then restart comparing from the end of the needle.

Example

Needle 1: "aardvark"

```
map[a] = max(8 - 1 - 5, 1) = 2
map[r] = max(8 - 1 - 6, 1) = 1
map[d] = max(8 - 1 - 3, 1) = 4
map[v] = max(8 - 1 - 4, 1) = 3
map[k] = max(8 - 1 - 7, 1) = 1
map[*] = 8 (* represents all other characters)
```

Haystack 1: "anteaters are aardvarks"

```
anteaters are aardvarks
aardvark.............          shift by map[r] = 1
.aardvark............          shift by map[s] = 8
.........aardvark......        shift by map[r] = 1
..........aardvark.....        shift by map[d] = 4
..............aardvark.        match at 14
```

Haystack 2: "crush kkyle with aardvarks"

```
crush kkyle with aardvarks
aardvark.................          shift by map[k] = 1
.aardvark................          shift by map[y] = 8
.........aardvark........          shift by map[ ] = 8
.................aardvark.         match at 17
```