BOI 2025, Day 1: Analysis.

If in the input we had numbers of occurrences of the character "B" in each substring, it would be easy to recover the positions of all "B"; the difficulty is that we do not know which characters are described by numbers given in the input. But maybe for some numbers we do know this? Indeed, the largest number on the input surely states the number of occurrences of "B"; the other characters are strictly less frequent. This allows us already to find the position f of the first "B" and the position ℓ of the last "B". Namely, f is the smallest index such that $M_{f+1,n} < M_{f,n}$; likewise ℓ is the greatest index such that $M_{1,\ell-1} < M_{1,\ell}$ (here and below we assume that each $M_{i,i-1}$, describing the empty substring, is 0).

Suppose now that we want to check whether "B" occurs at some position i strictly between f and ℓ. We have three cases.

$M_{f,i-1} > M_{f+1,i-1}$: we are sure that the only most common character at positions $f, f+1, \ldots, i-1$ is "B"; in this case "B" occurs at position i precisely when $M_{f,i} > M_{f,i-1}$.

$M_{i+1,\ell} > M_{i+1,\ell-1}$: then we are sure that the only most common character at positions $i+1, i+2, \ldots, \ell$ is "B"; in this case "B" occurs at position i precisely when $M_{i,\ell} > M_{i+1,\ell}$.

$M_{f,i-1} = M_{f+1,i-1}$ and $M_{i+1,\ell} = M_{i+1,\ell-1}$: this implies that some character a is strictly more frequent than "B" at positions $f+1, f+2, \ldots, i-1$, and some character c is strictly more frequent than "B" at positions $i+1, i+2, \ldots, \ell-1$. Then, if $M_{f+1,i} > M_{f+1,i-1}$, then the character at position i is $a \neq$ "B". Likewise, if $M_{i,\ell-1} > M_{i+1,\ell-1}$, then the character at position i is $c \neq$ "B". If none of these two inequalities hold, we know that at position i there is a character b other than a and other than c. Is it possible that $b \neq$ "B"? If so, since there are only three letters, we must have $a = c$. Therefore "B" does not occur more frequently than a at positions $f, f+1, \ldots, i-1$, and at positions $i+1, i+2, \ldots, \ell$. Additionally there is no "B" at positions before f, after ℓ, and at position i. This contradicts the assumption that "B" is strictly more frequent than a in the whole acronym. It follows that the character b at position i is "B". By this simple case-analysis, we can recover all positions at which "B" occurs.