ONE-ERROR CORRECTION

THE USER MIGHT HAVE INSERTED, REMOVED OR SUBSTITUTED ONE SINGLE CHARACTER.

WE WANT TO REMOVE, INSERT OR SUBSTITUTE THAT CHAR WITH THE CORRECT ONE

FERRRAGINA → FERRAGINA DELETION

FERAGINA → FERRAGINA INSERTION

FARRAGINA → FERRAGINA SUBSTITUTION

THERE MAY BE A LOT OF CANDIDATES !!

Indeed a string of length L over A possible chars may have

$$\#\text{variants} = L(A-1) + (L+1)A + L =$$
$$= LA - L + LA + L - L =$$
$$= 2LA + A = A(2L+1)$$

How to approach the problem and its cost

→ Create two dicts: $D1 = \{$ strings $\}$, $D2 = \{$ strings with 1 deletion $\}$

→ Given a query, made of a single string ?

↳ check if $P \in D1$ — 1 query

↳ check if $P \in D2$ — 1 query

↳ check if $P \setminus \{1 \text{ char}\} \in D1$ for each possible deletion — p queries, where $p = \#P$

↳ check if $P \setminus \{1 \text{ char}\} \in D2$ for each possible deletion — p queries, where $p = \#P$

$$\text{COST} = 2p + 2 \text{ queries for } P$$

($\equiv$ hash computations)

PRO = CPU efficient, no cache misses to compute p's hashes (but $O(p)$ cache misses to search in D1 or D2)

CONS = Large space because of the many strings in D2 which must be stored to search in the hashtable of D2 (unless we use perfect hashing to avoid collision)

FALSE MATCHES = $D1 = \{cat, cst\}$ $D2 = \{\dots\}$

$P = ctw$, ctw matches with cat and cst in D2 (ct matches with ct)

EXAMPLE

GIVEN $D = \{BINGO, BOSS, BOX, BULL, CAT\}$,

BUILD THE DATA STRUCTURES TO SOLVE THE 1-EDIT SEARCH AND RESOLVE THE QUERY $P_1 = bang$ and $P_2 = ball$

$D_1 = \{BINGO, BOSS, BOX, BULL, CAT\}$

$D_2 = \{INGO, BNGO, BIGO, BINO, BING,$
$OSS, BSS, BOS, \cancel{BOS}$
OX, BX, BO
$ULL, BLL, BUL, \cancel{BUL}$
$AT, CT, CA\}$

BANG

- BANG $\in D_1$ $\emptyset$
- BANG $\in D_2$ $\emptyset$
- ANG $\in D_1$ $\emptyset$
- ANG $\in D_2$ $\emptyset$
- BNG $\in D_1$ $\emptyset$
- BNG $\in D_2$ $\emptyset$
- BAG $\in D_1$ $\emptyset$
- BAG $\in D_2$ $\emptyset$
- BAN $\in D_1$ $\emptyset$
- BAN $\in D_2$ $\emptyset$

NO RESULTS

BALL

- BALL $\in D_1$ $\emptyset$
- BALL $\in D_2$ $\emptyset$
- ALL $\in D_1$ $\emptyset$
- ALL $\in D_2$ $\emptyset$
- BLL $\in D_1$ $\emptyset$
- BLL $\in D_2$ ✓
- BAL $\in D_1$ $\emptyset$
- BAL $\in D_2$ $\emptyset$

MAYBE YOU WERE SEARCHING FOR "BULL"?