

IN[34]120 Søketeknologi - Strengelgoritmer og oblighjelp

2023-09-15

TA: Oliver (oliverrij@ifi.uio.no)

- Aho–Corasick
- Levenstein distance
- Edit tables
- Oblighjelp



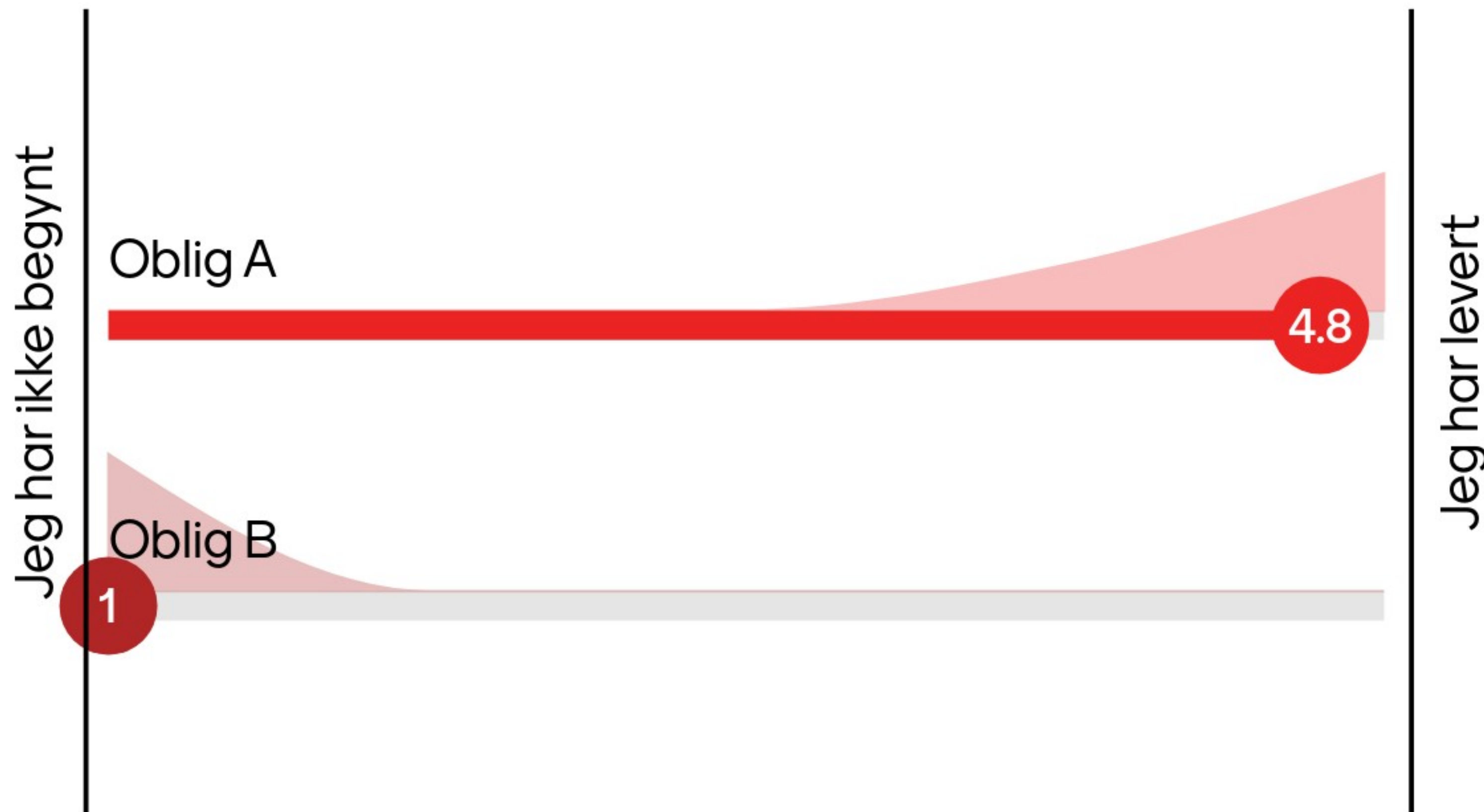
Agenda du jour

1. Strengalgoritmer: Aho–Corasick (ingen slide)
2. Introdusere oblig B
3. Et knippe oppgaver (på menti)
4. Oblig-workshop

Oblig A

- Inverted index
- OR mellom 2 posting lists
- AND mellom 2 posting lists
- Frist i dag
- Omfang: under 100 linjer v/optimal løsning

Status for oblig A



Neste "forelesning"

- Aleksander er borte
- Sergey og jeg holder
- Gi ønske til tema: <https://nettskjema.no/a/359385>
- Vanlig tid og sted

Oblig B

- Part 1: Suffix arrays
- Part 2: Simplified Aho–Corasick
- Harder than A (the hardest?)
- Prøv å begynne på obligen før neste seminar!!



Aho-Corasick

- Trie
- Trie walk
- Avgjør om tiden

Exercises

Stjålet fra ymse steder (*kopi fra ifjor*)



Select Answer

trie



The correct answer is: Binary search on a sorted array of the dictionary

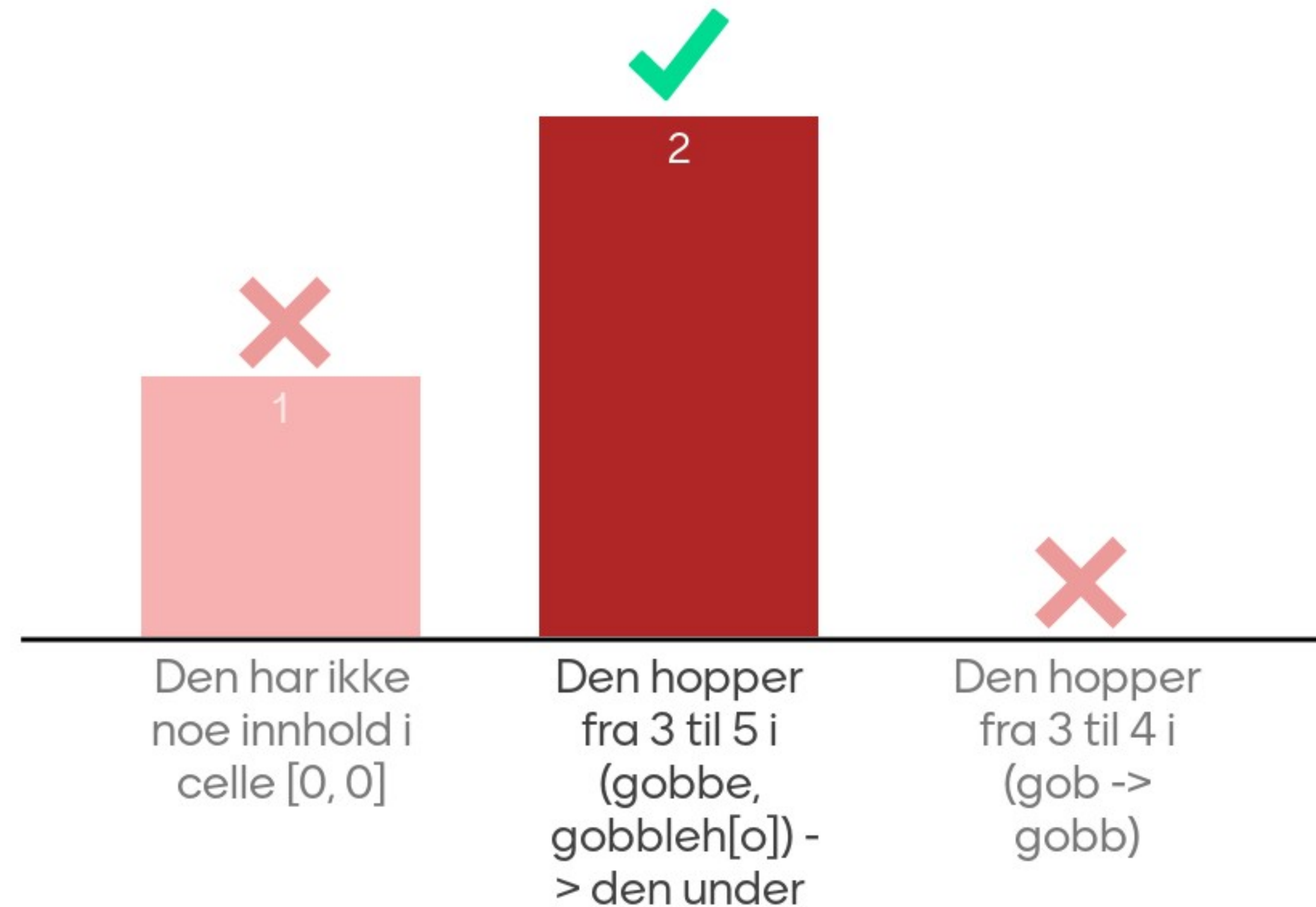
Leaderboard

0 p		ET
0 p		Magical Rhino
0 p		Bill
0 p		Bell

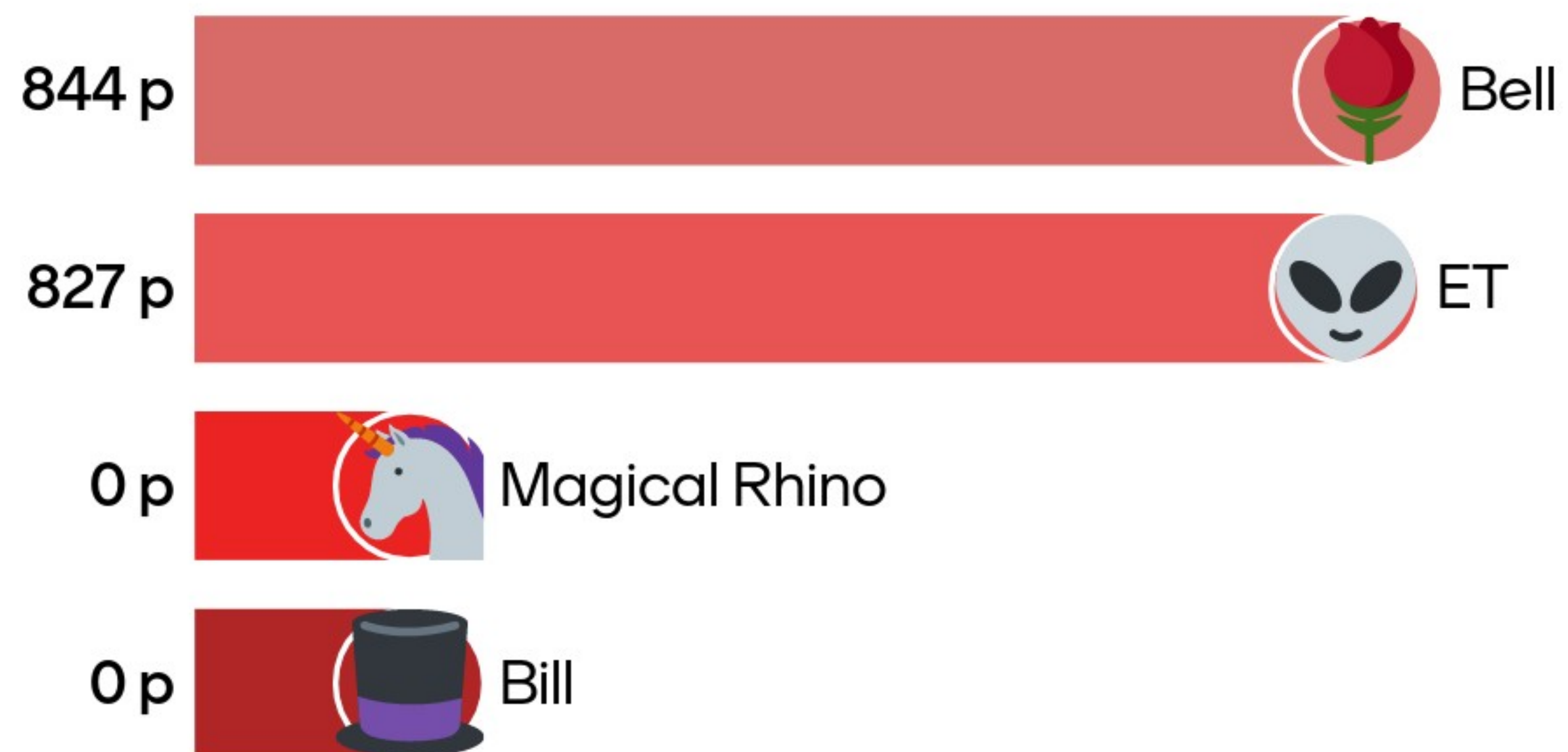
(e) [6%] If the requirement changes from exact prefix matching to allowing up to k edit errors when matching the prefixes, suggest a suitable data structure and describe the associated lookup algorithm. You can assume that the value of k is 0, 1 or 2 and determined as a function of the length of the prefix typed by the user.

~Les papers!~

Hva er feil med denne edit-tabellen hvis man har unit edit costs?



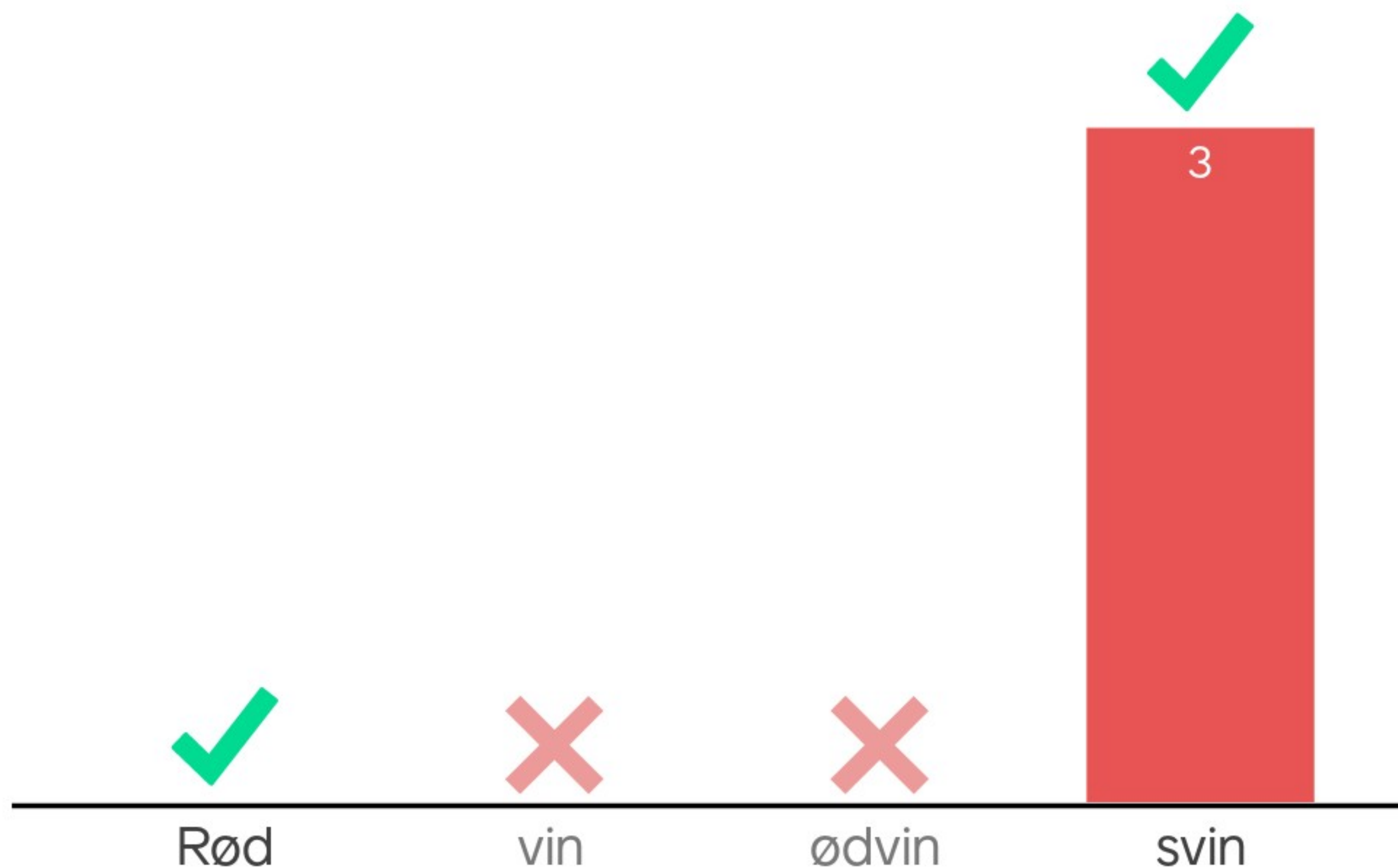
Leaderboard



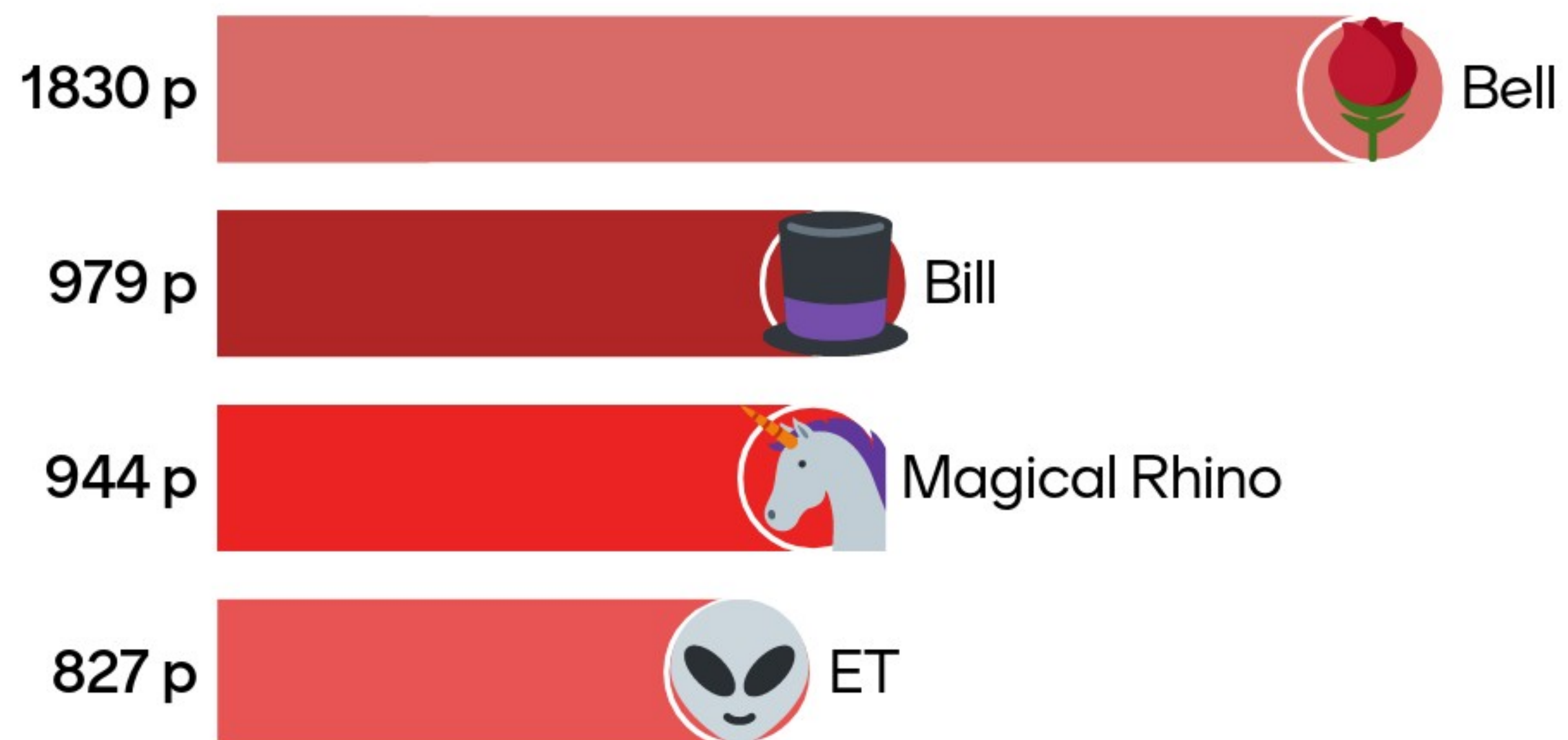
-	O	E	X	E	C	U	T	I	O	N
O	0	1	2	3	4	5	6	7	8	9
I	1	1	2	3	4	5	6	6	7	8
N	2	2	2	3	4	5	6	7	7	7
T	3	3	3	3	4	5	5	6	7	8
E	4	3	4	3	4	5	6	6	7	8
N	5	4	4	4	4	5	6	7	7	7
T	6	5	5	5	5	5	5	6	7	8
I	7	6	6	6	6	6	6	5	6	7
O	8	7	7	7	7	7	7	6	5	6
N	9	8	8	8	8	8	8	7	6	5

Edit table / edit distance

Hva er ikke et suffix av "rødvin"?



Leaderboard



Oblig A common fails + hints

- Bruk Mattermost! Mange gode tips i kanalene + spørre om hjelp
- Man trenger ikke mer enn ~10 linjer for å bygge indexen
- ALDRI ta list(p1) -> obligen stryker selv om testene funker
- Postingsmerger må skalere lineært med lengden på postinglistene

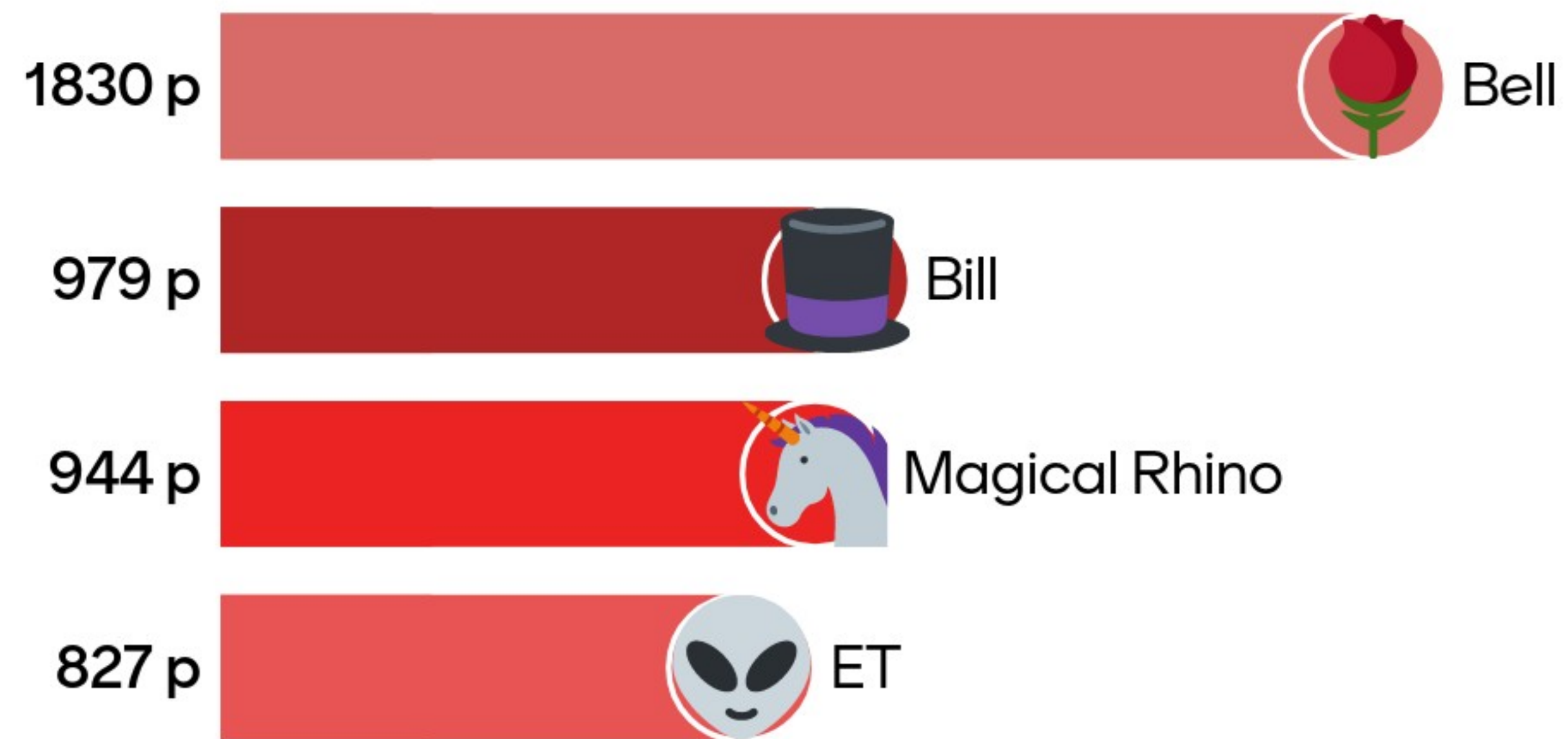
Jobbe med oblig. Enten A eller B

Oblig A har deadline i kveld! (Og det er Dagen@ifi)
Still spørsmål så du kommer i mål B)

Jeg har levert (ikke skriv navn, skriv 'X')

There's no correct answer!

Leaderboard



Si hva som helst:

Waiting for responses ...

