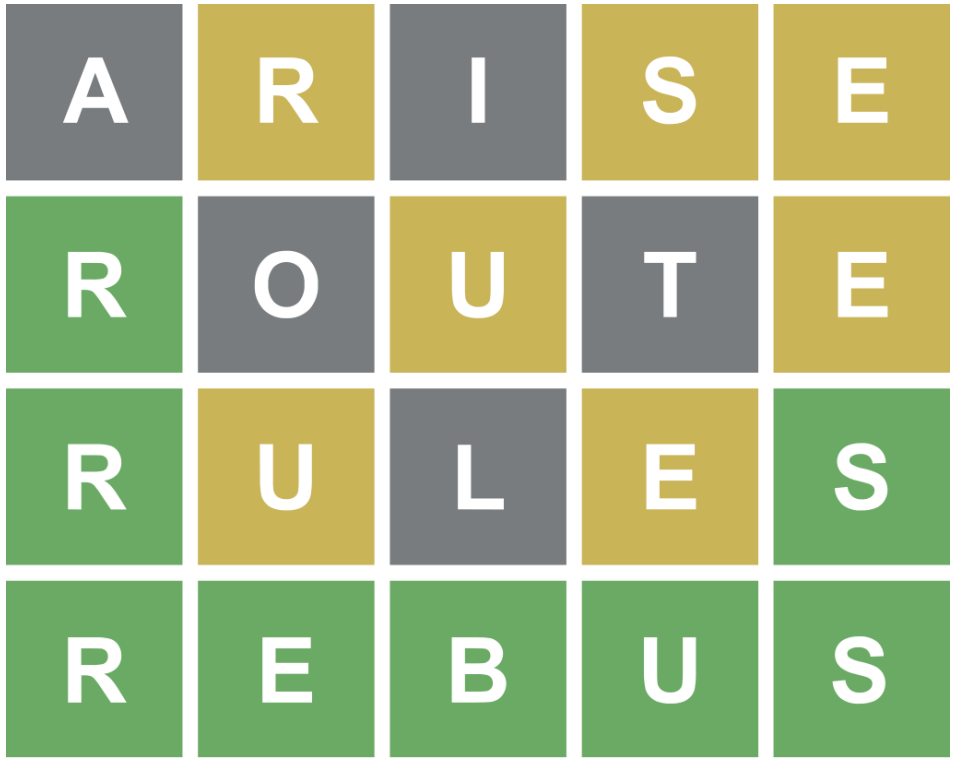
Løse wordle som en søketek student

Simon Halvorsen og Victor Moene

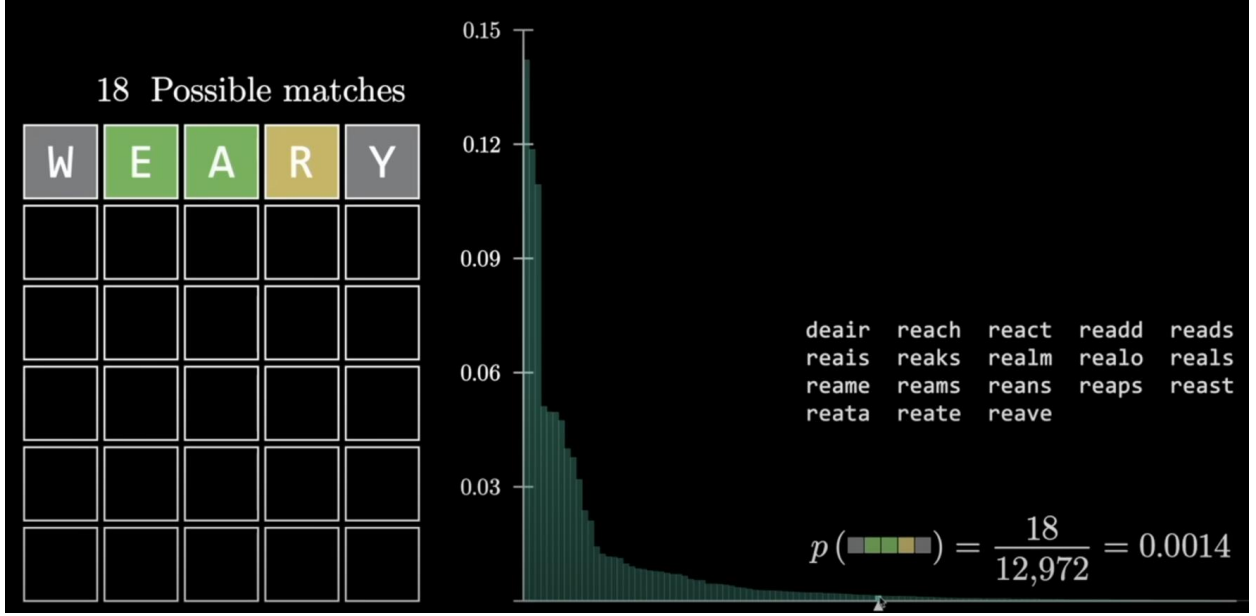
Hva er Wordle?

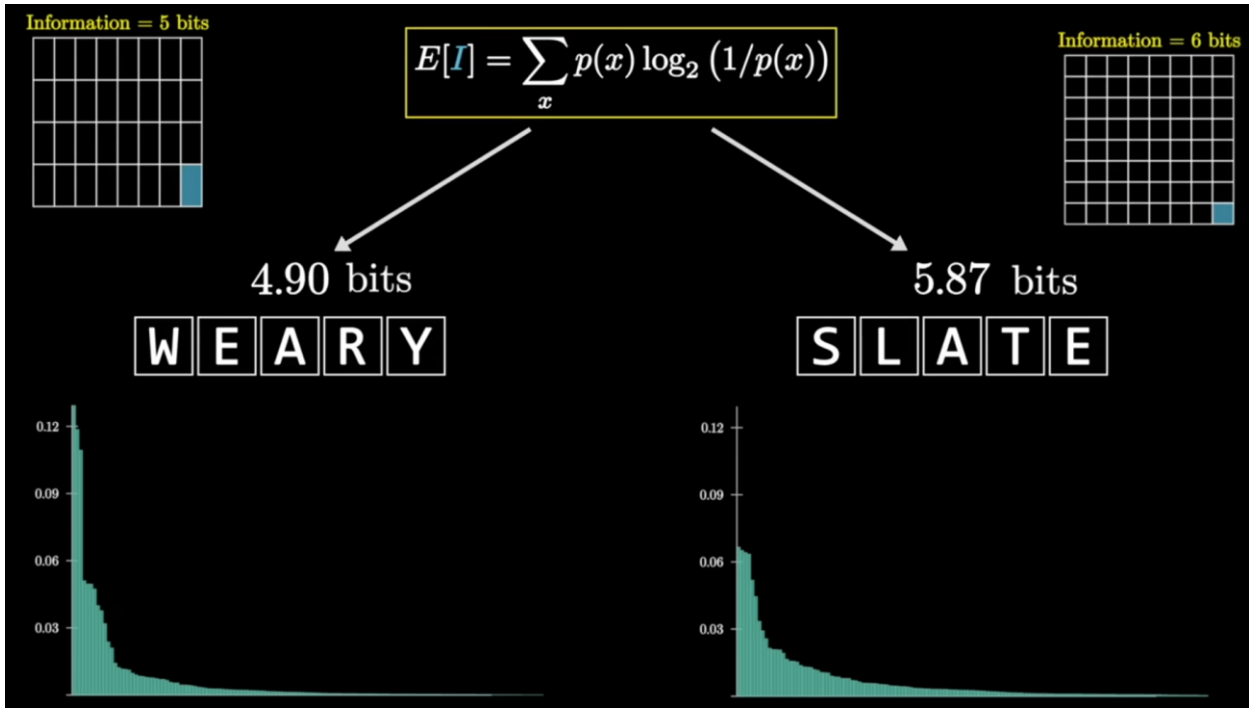
- Hva er det?
  - Svært populært nettbasert ordspill, laget av Josh *Wardle*
  - Ble kjøpt av NYT i 2022
- Hvordan fungerer det?
  - Finne ut hva det hemmelige, engelskspråklige ordet på fem bokstaver er
  - Bare 6 forsøk
- Hvordan løser man det?
  - Grå bokstaver: finnes ikke i ordet
    - Litt mer komplisert enn det
  - Gule: er på feil plass
  - Grønne: er på riktig plass



3blue1brown

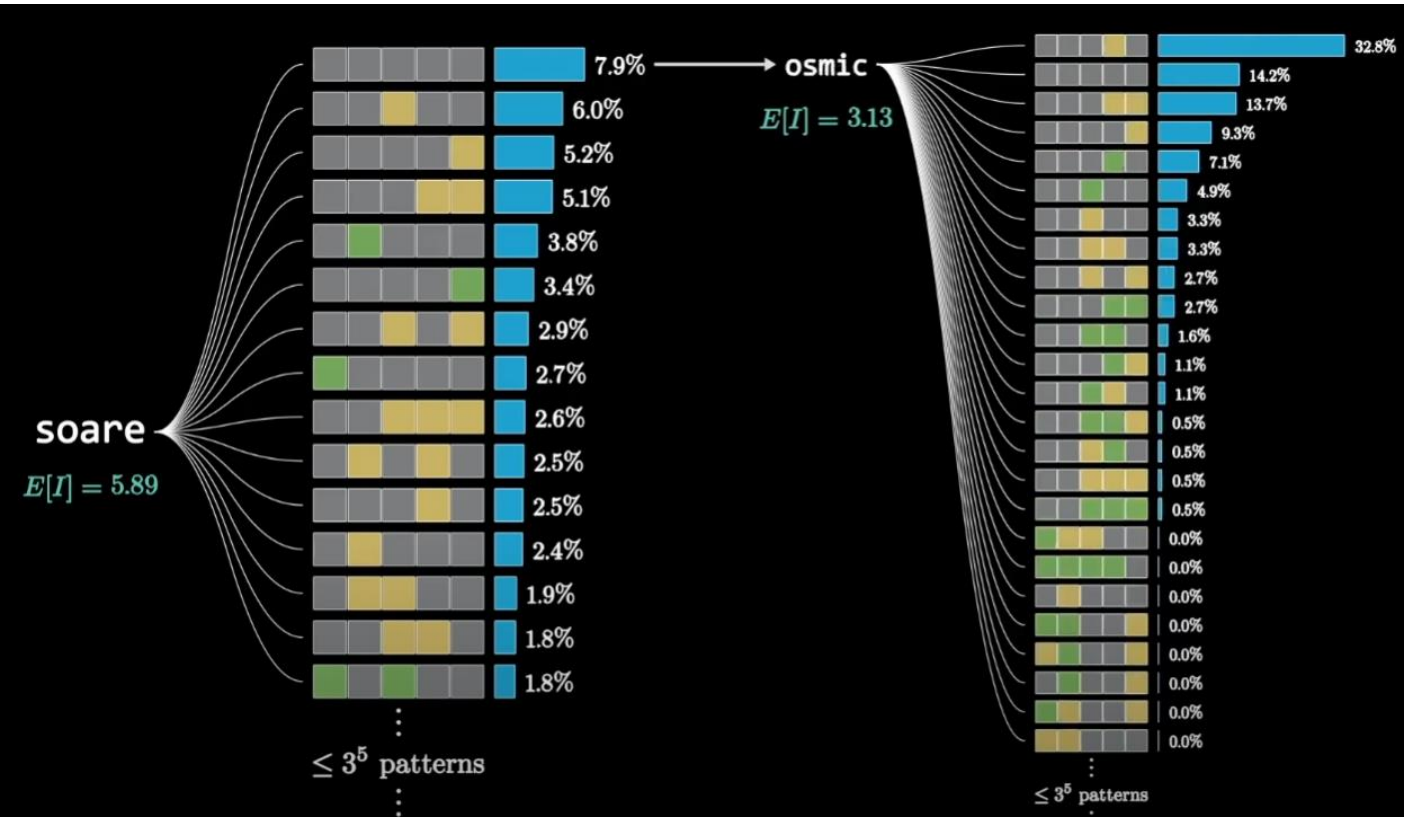
- Generell løsning
  - ~13000 ord
- Informasjonsgevinst
  - Mest sannsynlig utfall <-> minst informativt
  - I = log(1/p)
- Entropi
  - Snittet av informasjongevinster for alle matches
- Løse wordle
  - For hvert steg, velge ordet som maksimiserer entropi





3Blue1Brown - Beste åpningsord

- https://github.com/3b1b/videos/tree/master/_2022/wordle



Highest $E[I]$ (one step)	Highest $E[I]$ (two steps)
1. soare → 5.89	1. slane → 5.77 + 4.27 = 10.04
2. roate → 5.88	2. slate → 5.86 + 4.18 = 10.03
3. raise → 5.88	3. salet → 5.83 + 4.18 = 10.02
4. raile → 5.87	4. trace → 5.83 + 4.18 = 10.01
5. reast → 5.87	5. crate → 5.83 + 4.18 = 10.01
6. slate → 5.86	6. reast → 5.87 + 4.14 = 10.01
7. crate → 5.83	7. carle → 5.77 + 4.24 = 10.01
8. salet → 5.83	8. roast → 5.65 + 4.35 = 10.00
9. irate → 5.83	9. torse → 5.72 + 4.27 = 10.00
10. trace → 5.83	10. carse → 5.77 + 4.23 = 10.00
11. arise → 5.82	11. carte → 5.79 + 4.20 = 10.00
12. orate → 5.82	12. toile → 5.69 + 4.31 = 10.00
13. stare → 5.81	13. trone → 5.68 + 4.31 = 10.00
14. carte → 5.79	14. soare → 5.89 + 4.11 = 9.99
15. raine → 5.79	15. raile → 5.87 + 4.13 = 9.99

```
a df: 909 cf: 979
b df: 267 cf: 281
c df: 448 cf: 477
k df: 202 cf: 210
s df: 618 cf: 669
e df: 1056 cf: 1233
t df: 667 cf: 729
y df: 417 cf: 425
o df: 673 cf: 754
h df: 379 cf: 389
r df: 837 cf: 899
i df: 647 cf: 671
d df: 370 cf: 393
l df: 648 cf: 719
u df: 457 cf: 467
v df: 149 cf: 153
n df: 550 cf: 575
g df: 300 cf: 311
p df: 346 cf: 367
m df: 298 cf: 316
f df: 207 cf: 230
x df: 37 cf: 37
w df: 194 cf: 195
z df: 35 cf: 40
j df: 27 cf: 27
q df: 29 cf: 29
```

```
Attempt 4: Guessing 'ovary'
Solution found in 4 attempts: ovary
Attempt 1: Guessing 'slate'
Attempt 2: Guessing 'joint'
Attempt 3: Guessing 'toxin'
Attempt 4: Guessing 'tonic'
Solution found in 4 attempts: tonic
Attempt 1: Guessing 'slate'
Attempt 2: Guessing 'squib'
Attempt 3: Guessing 'swung'
Attempt 4: Guessing 'shuck'
Solution found in 4 attempts: shuck

=== Summary ===
Total words attempted: 500
Total successful solves: 428
Average attempts for successful solves: 3.808411214953271
Words not solved within 6 attempts: harry
, pagan
, drift
, piety
, manic
```

```
{}
[193, 196, 198, 202, 218, 219, 220, 221, 226, 227, 228, 229,
Attempt 2: Guessing 'biddy'
{'i': (1, 0), 'd': (2, 0), 'y': (1, 0)}
[854]
Attempt 3: Guessing 'giddy'
Solution found in 3 attempts: giddy

=== Summary ===
Total words attempted: 100
Total successful solves: 98
Average attempts for successful solves: 3.7244897959183674
Words not solved within 6 attempts:
older
sight

----19.57----
```

Improved pruning

```
' in 4 attempts.
Solved 'youth
' in 4 attempts.
Solved 'zebra
' in 2 attempts.
Solved 'zesty
' in 3 attempts.
Failed to solve 'zonal'.

=== Summary ===
Total words attempted: 2315
Total successful solves: 1991
Average attempts for successful solves: 3.848317
Words not solved within 6 attempts:
aback
abled
afoul
aloof
amiss
arson
awash
```

```
' in 4 attempts.
Solved 'youth
' in 4 attempts.
Solved 'zebra
' in 2 attempts.
Solved 'zesty
' in 3 attempts.
Solved 'zonal' in 3 attempts.

=== Summary ===
Total words attempted: 2315
Total successful solves: 2235
Average attempts for successful solves: 3.980760
Words not solved within 6 attempts:
aback
basis
bushy
catch
clack
crass
daddy
```

```
Target word: zesty
Attempt 1: Guessing 'slate'
{'s': (1, 0), 't': (1, 0), 'e': (1, 0)}
[1400, 2035, 2313]
Attempt 2: Guessing 'zesty'
Solution found in 2 attempts: zesty
```

```
Target word: sigma
Attempt 1: Guessing 'slate'
{'s': (1, 0), 'a': (1, 0)}
[1642, 1644, 1645, 1649, 1650, 1652, 1653, 1654, 1677, 1678, 1683, 1753, 1839, 1886, 1892, 1961, 1966]
Attempt 2: Guessing 'squad'
{'s': (1, 0), 'a': (1, 0)}
[1644, 1645, 1652, 1653, 1654, 1753]
Attempt 3: Guessing 'savvy'
{'s': (1, 0), 'a': (1, 0)}
[1753]
Attempt 4: Guessing 'sigma'
Solution found in 4 attempts: sigma
```

```
Target word: hound
Attempt 1: Guessing 'slate'
{}
[193, 198, 200, 238, 240, 242, 247, 253, 255, 274, 278, 279, 280, 285, 286, 287, 289, 293,
84, 385, 386, 387, 391, 393, 433, 434, 436, 437, 438, 442, 443, 444, 478, 482, 485, 487, 48
, 507, 511, 555, 556, 565, 567, 569, 571, 575, 576, 578, 579, 596, 602, 605, 606, 607, 610,
798, 813, 814, 818, 824, 825, 826, 827, 828, 830, 854, 879, 884, 886, 890, 913, 914, 917, 9
4, 989, 991, 995, 999, 1001, 1002, 1003, 1005, 1006, 1007, 1011, 1016, 1018, 1029, 1030, 10
1074, 1084, 1086, 1227, 1232, 1235, 1236, 1252, 1255, 1256, 1263, 1265, 1271, 1273, 1275, 1
 1357, 1358, 1405, 1408, 1411, 1413, 1415, 1451, 1452, 1453, 1457, 1458, 1466, 1470, 1472,
, 1493, 1496, 1503, 1518, 1522, 1586, 1592, 1593, 1605, 1607, 1611, 1615, 1616, 1620, 1623,
1, 2220, 2222, 2252, 2253, 2256, 2261, 2265, 2270, 2272, 2273, 2281, 2284, 2285, 2287, 2292
Attempt 2: Guessing 'biddy'
{'d': (2, 1)}
[381, 491, 571, 575, 602, 605, 607, 746, 798, 814, 890, 995, 1263, 1458, 1481, 1616, 2292]
Attempt 3: Guessing 'fjord'
{'o': (1, 0), 'd': (1, 0)}
[995, 1263, 1458, 2292]
Attempt 4: Guessing 'wound'
{'o': (1, 0), 'u': (1, 0), 'n': (1, 0), 'd': (1, 0)}
[995, 1263, 1458]
Attempt 5: Guessing 'mound'
{'o': (1, 0), 'u': (1, 0), 'n': (1, 0), 'd': (1, 0)}
[995, 1458]
Attempt 6: Guessing 'pound'
{'o': (1, 0), 'u': (1, 0), 'n': (1, 0), 'd': (1, 0)}
[995]
Max attempts reached. Solution not found.
```

Josh Stephenson – Hvordan redusere ordlisten

Example:

Let's look at the target word `HOUND`. The algorithm will start by guessing `SLATE` and then `CRONY` which will yield a green match on `N` in the 4th position and a yellow match on `O`. Once we filter out all the words that:

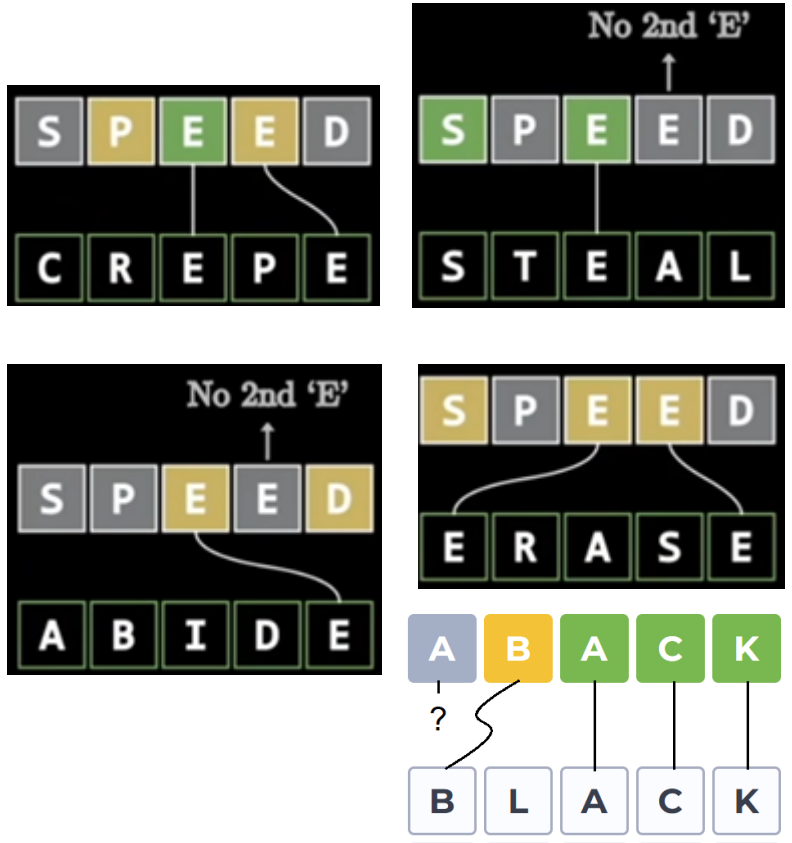
1. Don't have an `O` or have an `O` in the third position.
2. Don't have an `N` in the 4th position.
3. Have any of the gray letters: `{S,L,A,T,E,C,R,Y}`.

This will only leave 9 words in the answer list: `{BOUND, POUND, FOUND, DOING, MOUND, GOING, WOUND, HOUND, OWING}`. As you can see, most of these words have 4 of the same letters and most are in the same position, so if we were to continue with the algorithm as is, it would take 8 guesses to solve `HOUND`. That's why it makes more sense here to take the letters `{M, H, D, U, G, F, W, Z, J, Q}` (letters that are in the answer list but have not been matched as green or yellow yet) and find a word with as many of those letters as possible. In this case, the word `HUMID` does well, and after that the answer list is down to just one word: `HOUND`. With this improvement it only takes 4 guesses to solve it: `SLATE>CRONY>HUMID>HOUND`

- https://github.com/joshstephenson/Wordle-Solver

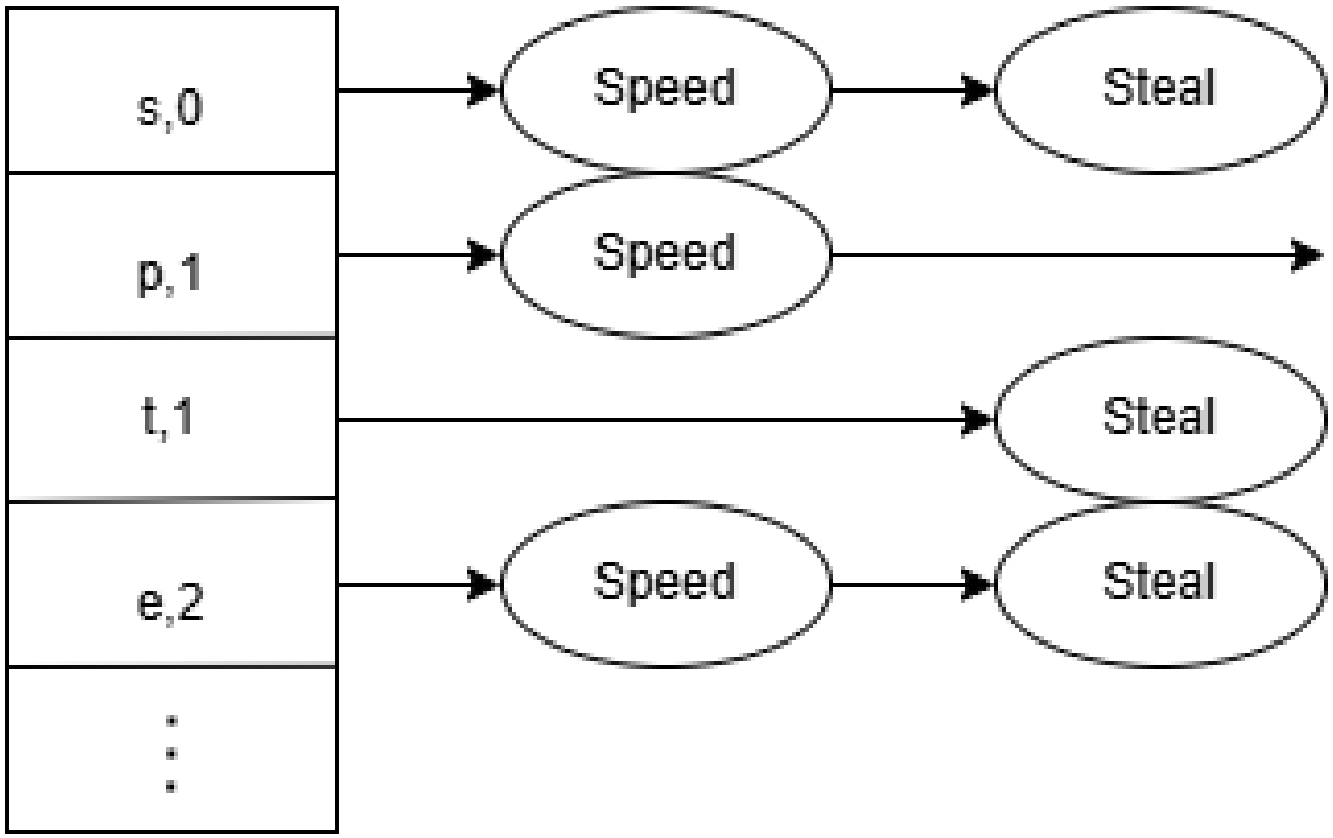
Hvordan prune?

- Gitt en bosktav (char,pos), hvis:
- Green
  - Fjerne alle ord som ikke har char på posisjon pos
- Yellow
  - Fjerne alle ord som har char på posisjon pos
  - og alle ord som ikke inneholder char
- Gray hvis andre chars ikke er grå
  - Fjerne alle ord som har char på posisjon pos
  - Og alle ord som inneholder char (antall grønne og gule – antall grå) ganger
- Else
  - Fjerne alle ord som inneholder char



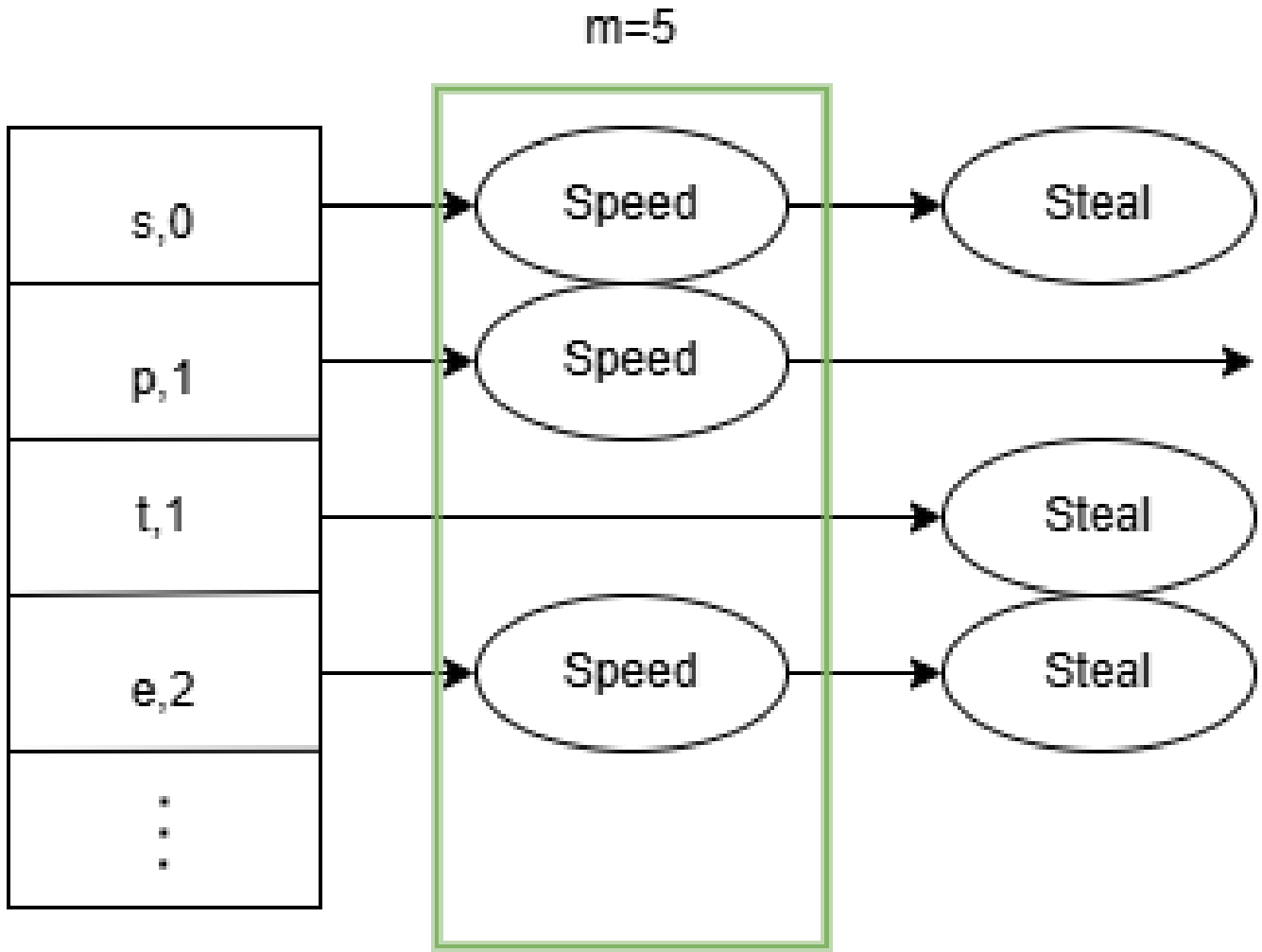
Inverted Index

Hver posting inneholder id-en til ordet, i stedet for strengen

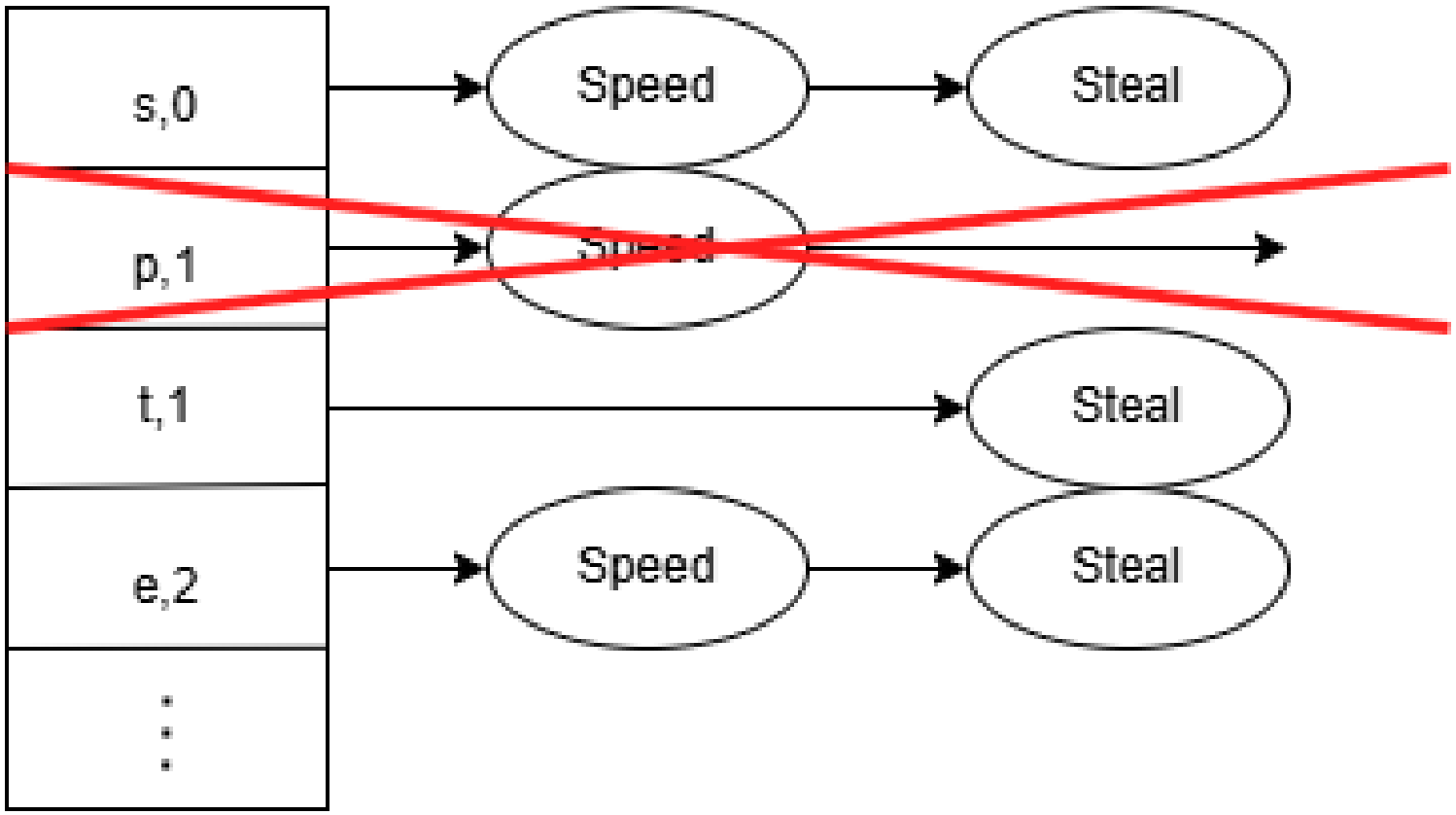


Word at a time 5-out-of-N merging

+ sjekker om ordet inneholder ugyldige bokstaver og mer...



Pruning



Takk for oss

https://github.com/SimonThalvorsen/wordle-for-soketekstudenter