

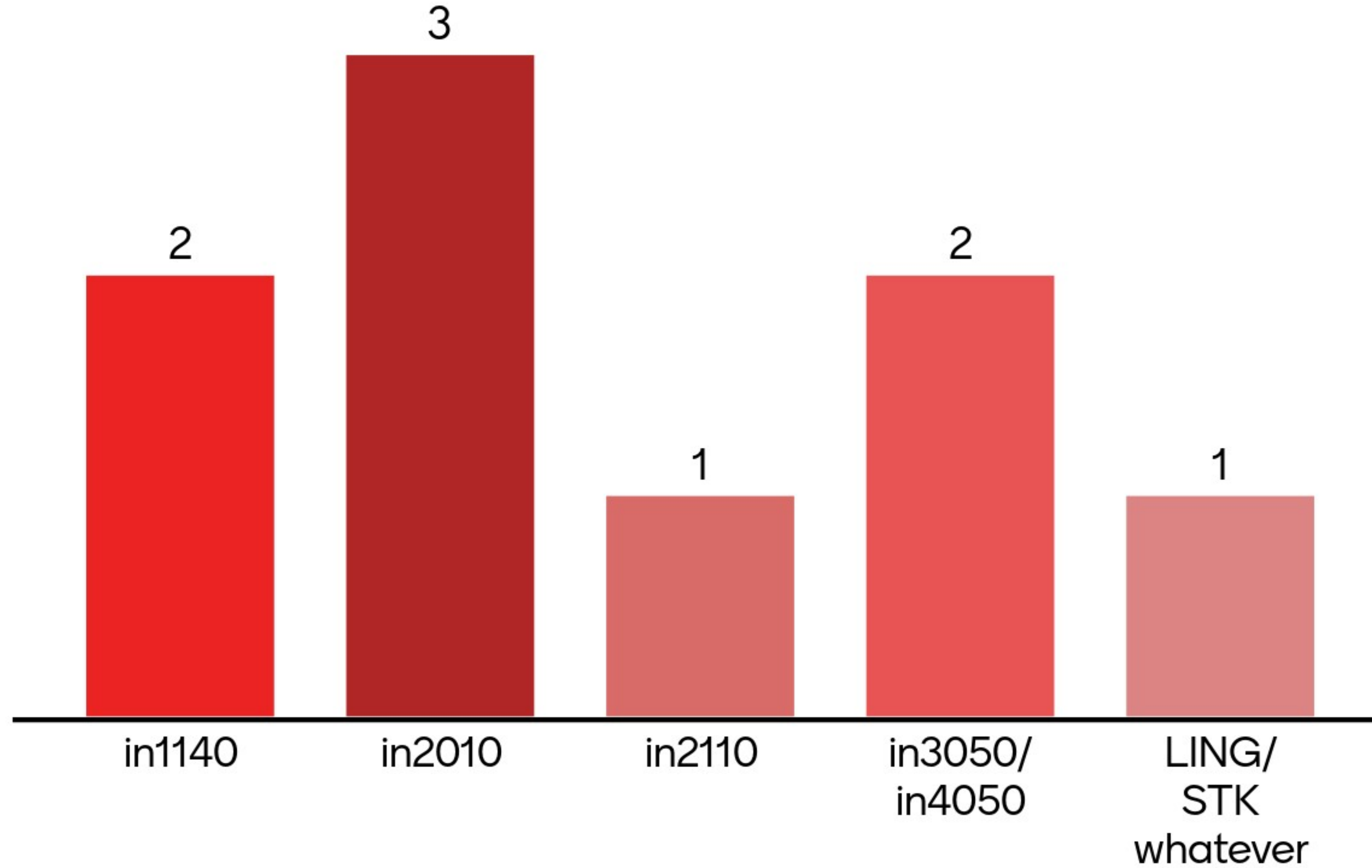
Søketek- rocchioBayes!!

2022-10-19

Poll om tidligere emner
Rocchio query expansion
Bayesian classification
Introuserere oblig D
Introducere oblig E
Minne om science fair
Lyntaler
Gjennomgå oblig C?
Assignment aid



Hvilke emner har du tatt tidligere?



Generell Rocchio

- Abstraksjon over dokumenter
- Vektorisering
- Sentroider

Rocchio query expansion

- Lag sentroide av queryen
- Sammenlign query-sentroiden med andre sentroider
- "Kast en stein i vannet"
- Skal gi de k mest like dokumentene
- (Anta at hvis et dokument er likt queryen så er det relevant)
- Man expander queryen.

Bayesian classification (Naïve Bayes)

- Beregn hvilken klasse et gitt dokument tilhører.
- E.g. gitt et stykke tekst, hvilket språk er det mest sannsynlig at teksten er på.
- Skal implementeres fra scratch i oblig E.

Bayes' regel, statistikk osv?

- *Ser an resultater fra pollen*
- Sannsynligheter
- Priors
- Posteriors

Oblig D

- betterranker, shinglegenerator
- Lag en ranker som bruker tf-idf-vekt
- Implementer en algoritme som kan generere shingles fra en streng
- Frist 28.10. Nest siste oblig.

Shingles

- Tenk "alle substrenger av en gitt lengde"
- For example, for $k = 3$, the string banana would be tokenized into the shingles {ban, ana, nan, ana}
- Lar oss gjøre approximate matching: organik kemmistry matches documents containing the terms organic chemistry

Oblig E

- Implementer naïve bayes classification fra bunnen av
- Handler mer om å forstå statistikken enn å progge
- Frist 11.11. Siste oblig.
- (Det står mye nyttig i boka for oblig E, se kap 13)
- ((^husk også at boka er open source: <https://nlp.stanford.edu/IR-book/pdf/13bayes.pdf>))

Science fair

- Kun for in4120. In3120 kan høre på.
- Frist for tema: 03.11
- Devilry: trenger ikke gjøre noe med science fair der

Pseudo-science fair? Lyntaler!!

- MAPS-event i dag 16:15 på Java
- Jeg snakker om VB Sieve (så det er søketekstrelevant å komme :D)
- Annika vil snakke om chatbots (blir garra 10/10)

SVMer

- "Support vector machines"
- Vanskelig(steste temaet)
- I forelesningen nå
- Maskinlæring, klassifisering
- Som naïve bayes bare mer komplisert



1-til-1-hjelp

Alle burde begynne på oblig D :)

