

IMPROVING BLOOM FILTER PERFORMANCE ON SEQUENCE DATA USING K-MER BLOOM FILTERS

FRANKO ŠESTO

GRGUR KOVAČ

IVAN MATAK

VODITELJ: DOC. DR. SC MIRJANA DOMAZET-LOŠO

SADRŽAJ

- BLOOM FILTER
- KORIŠTENI ALGORITMI
- USPOREDBA VELIČINE SETA SPREMLJENIH K-MERA I FPR MJERE OVISNO O DULJINI K-MERA
- ZAKLJUČAK

BLOOM FILTER

- KORISTI SE UMJESTO TRADICIONALNIH STRUKTURA PODATAKA ZA SPREMANJE ELEMENATA
- OSTAVLJA PUNO MANJI MEMORIJSKI OTISAK
- OMOGUĆUJE BRZO I EFIKASNO PRETRAŽIVANJE PRISUTNOSTI NEKOG ELEMENTA UZ MALU VJEROJATNOST POGREŠKE (*FPR*)
- FILTER JE ZAPRAVO BITOVNI VEKTOR
- KORISTI VIŠE HASH FUNKCIJA ZA SPREMANJE POJEDINOG ELEMENTA

KORIŠTENI ALGORITMI

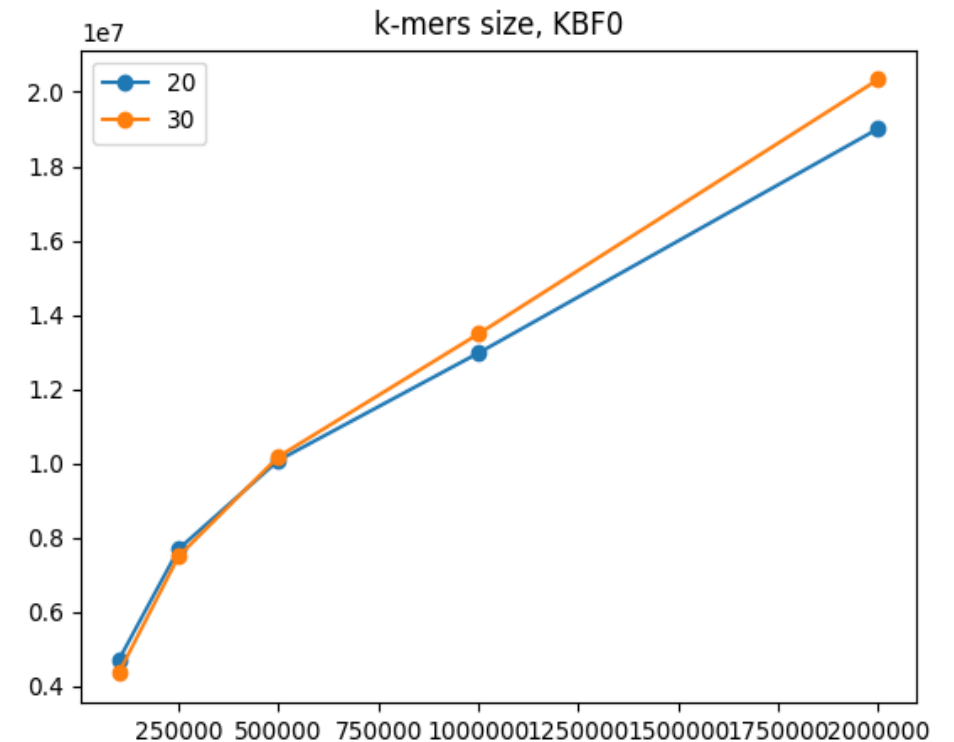
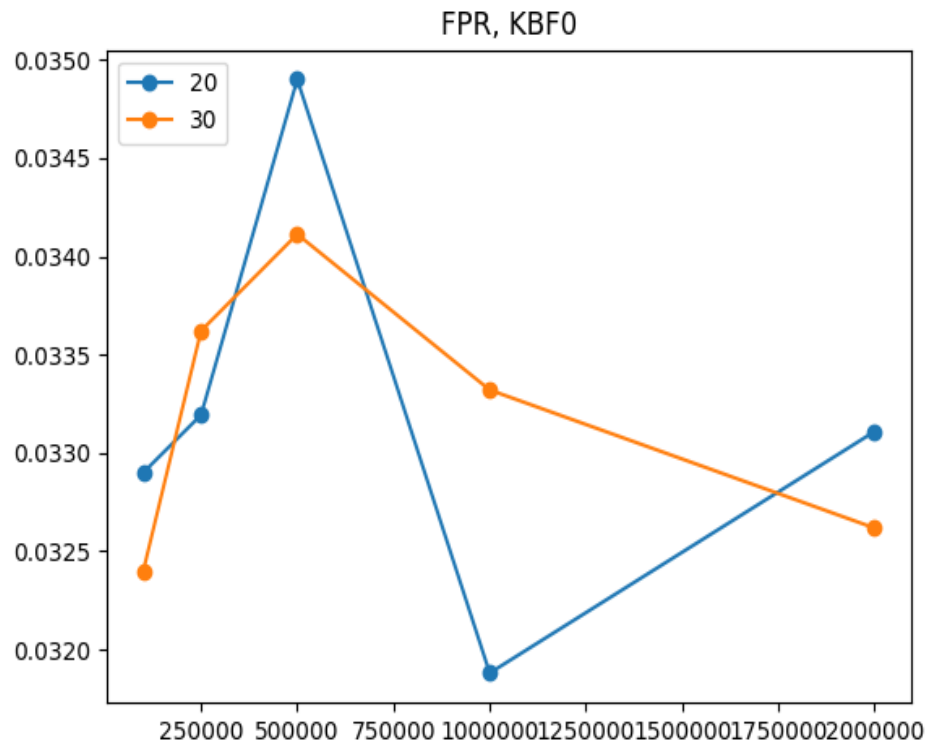
- KBF0
 - ZAPRAVO OBIČAN BLOOM FILTER
 - SLUŽI KAO REFERENCA ZA USPOREDBU
- KBF1
 - DODAJE DODATNU FUNKCIONALNOST PROVJERE SUSJEDA ZA 1 UDALJENIH LIJEVO I DESNO
 - AKO JE IJEDAN SUSJED SADRŽAN U BLOOM FILTERU VRAĆA TRUE
- KBF2
 - UVODI DODATNI SET RUBNIH K-MERA
 - POSTROŽUJE KRITERIJE U ODNOSU NA KBF1 TE ZAHTIJEVA PRISUTNOST BAREM JEDNOG SUSJEDA S OBJE STRANE U BLOOM FILTERU
 - IZNIMKA SU K-MERI KOJI SU NA RUBU PA NE MOGU IMATI SUSJEDA S OBJE STRANE

- KBF SPARSE STRICT I KBF SPARSE RELAXED

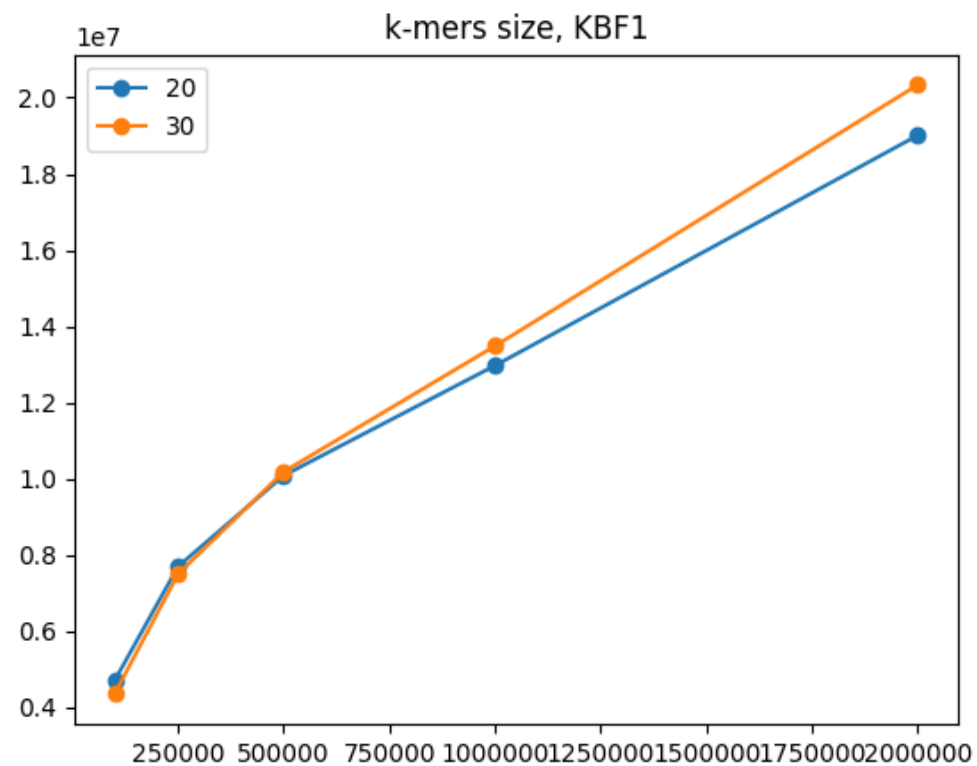
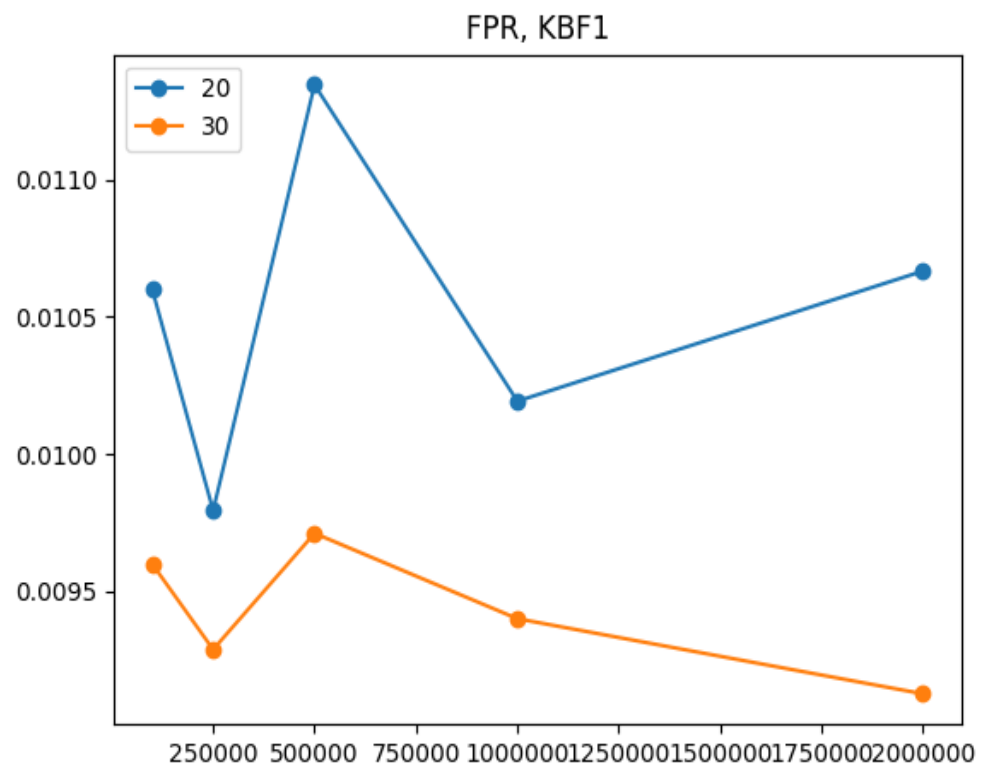
- KORISTE PREKLAPANJE K-MERA KAKO BI SMANJILI NEPOTREBNO SPREMANJE SVIH K-MERA
- NA TEMELJU PREKLAPANJA MOŽE SE ZAKLJUČITI O PRISUTNOSTI NEKOG ELEMENTA U FILTERU BEZ OBZIRA NA TO ŠTO NIJE U NJEMU
- KORISTE POHLEPNI ODABIR ZA SMANJENJE BROJA K-MERA

USPOREDBA VELIČINE SETA SPREMLJENIH K-MERA I FPR MJERE OVISNO O DULJINI K-MERA

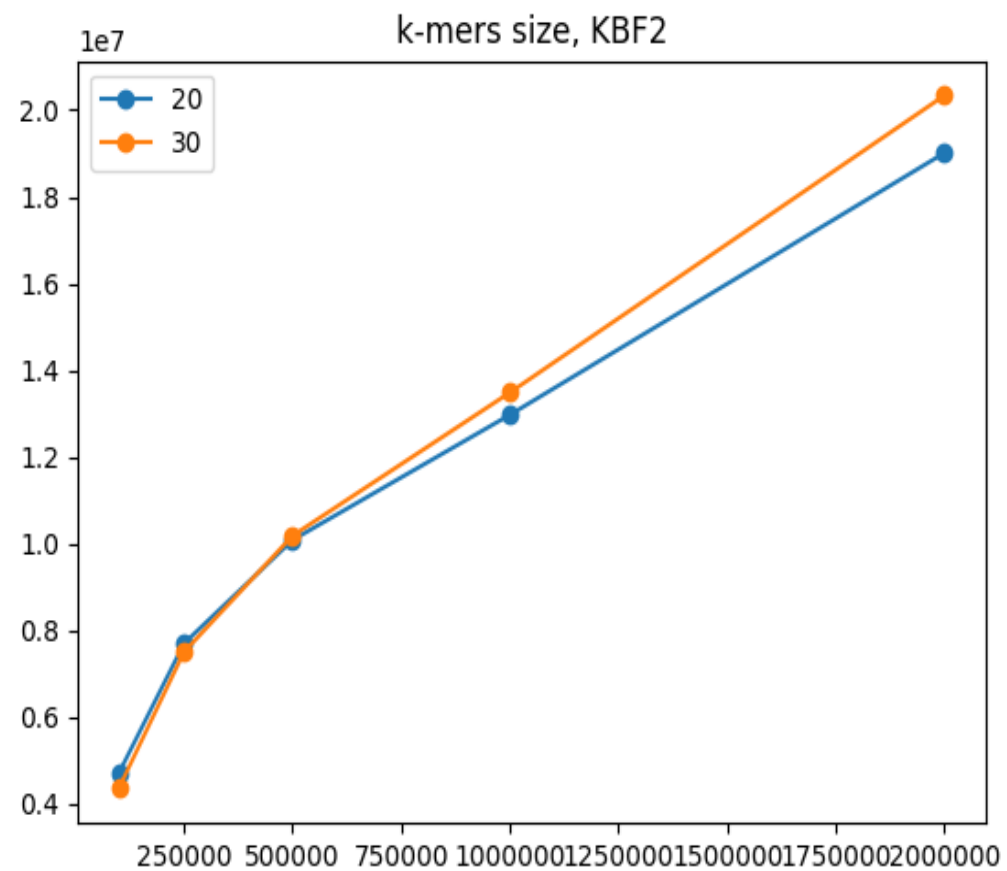
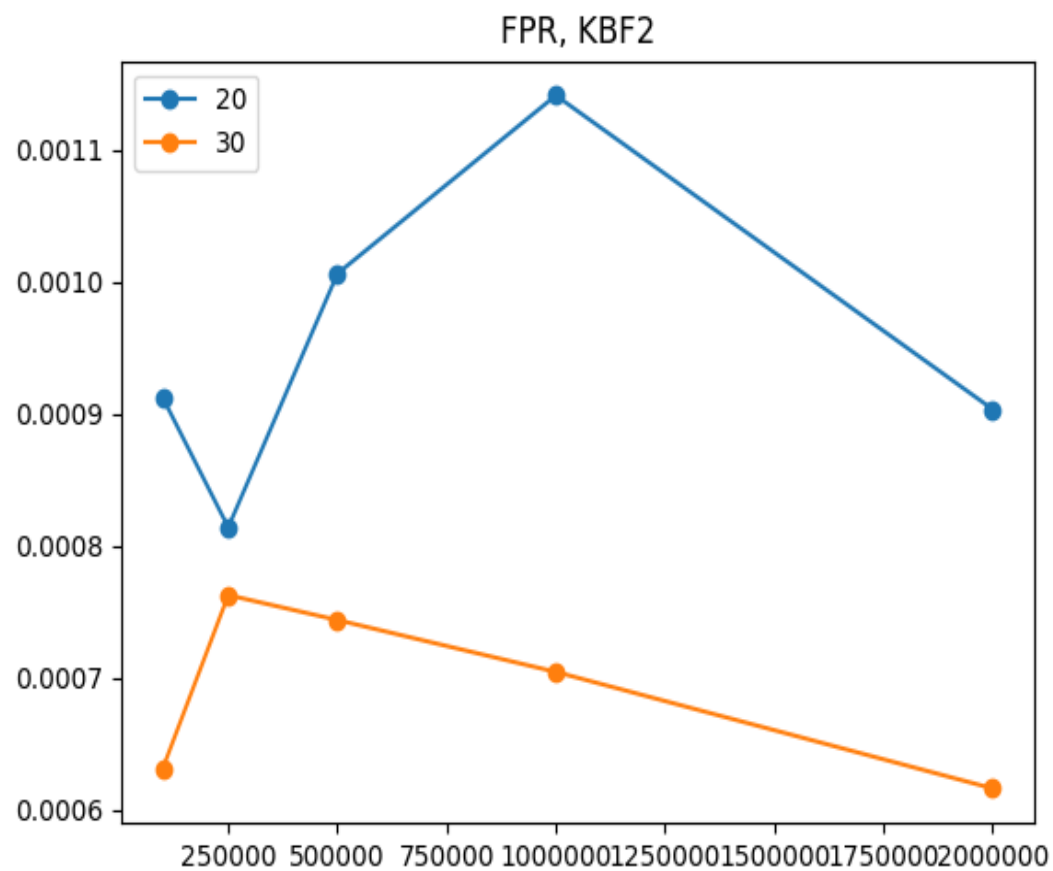
- KBFO



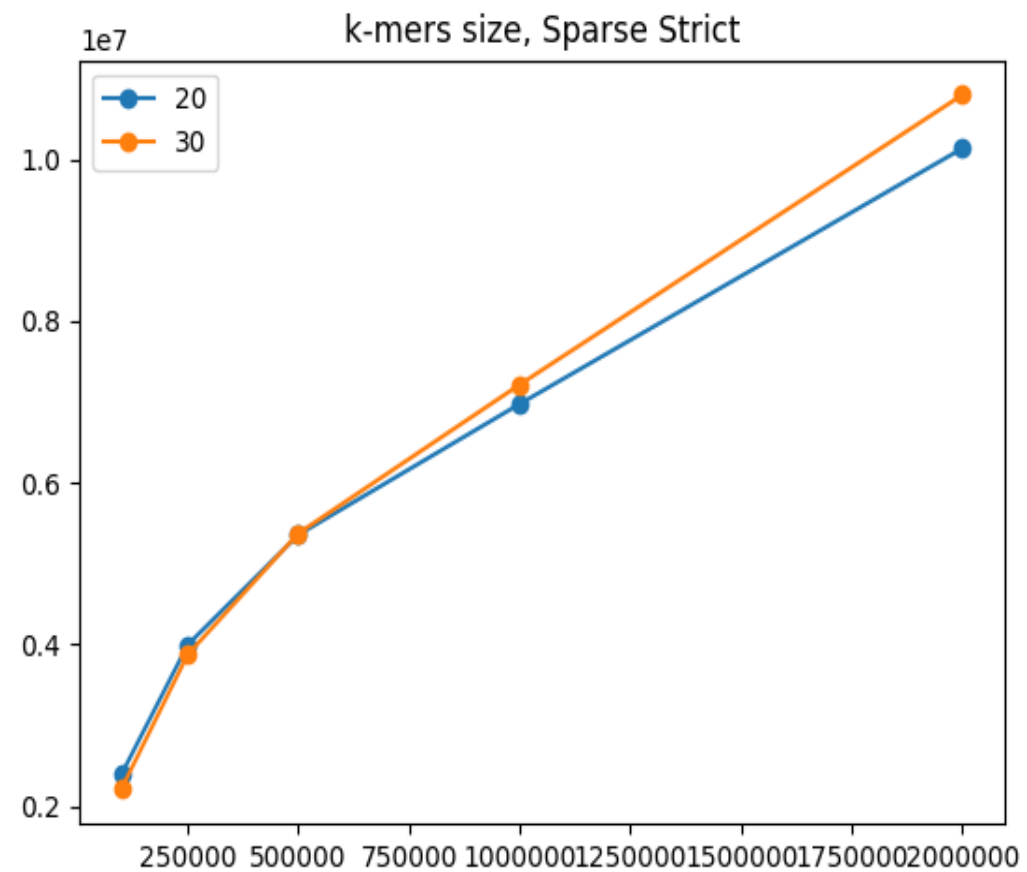
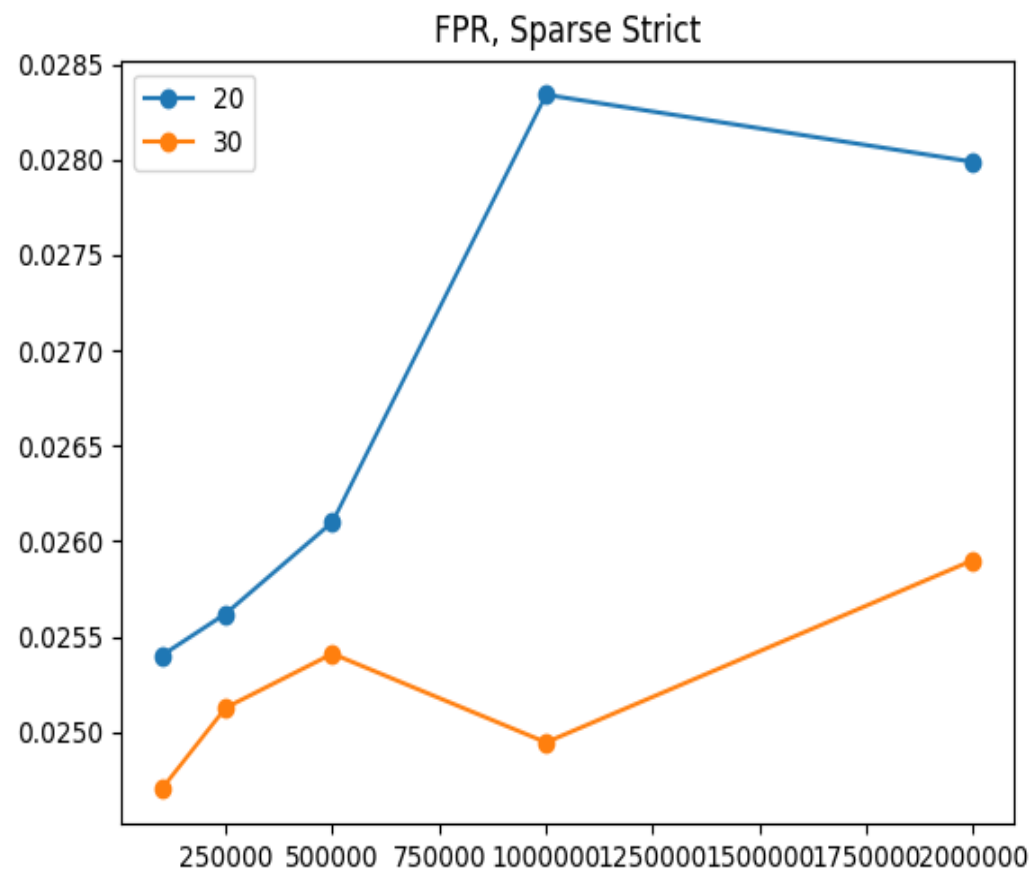
- KBF1



- KBF2

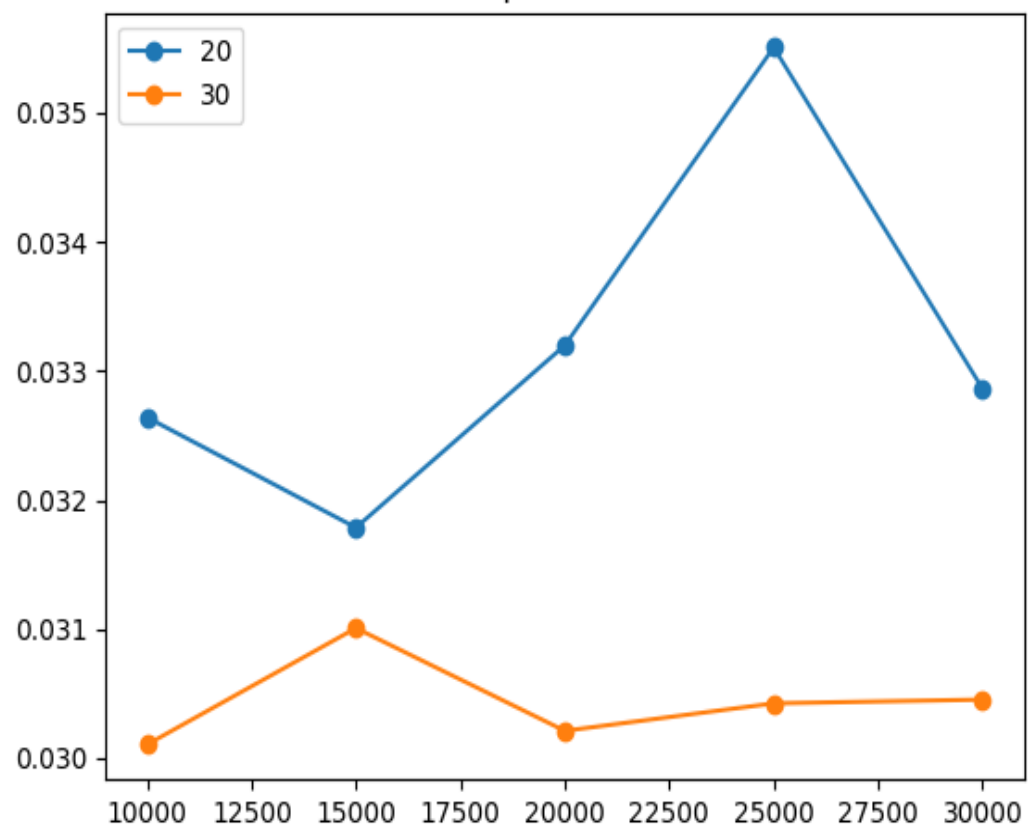


- KBF SPARSE STRICT

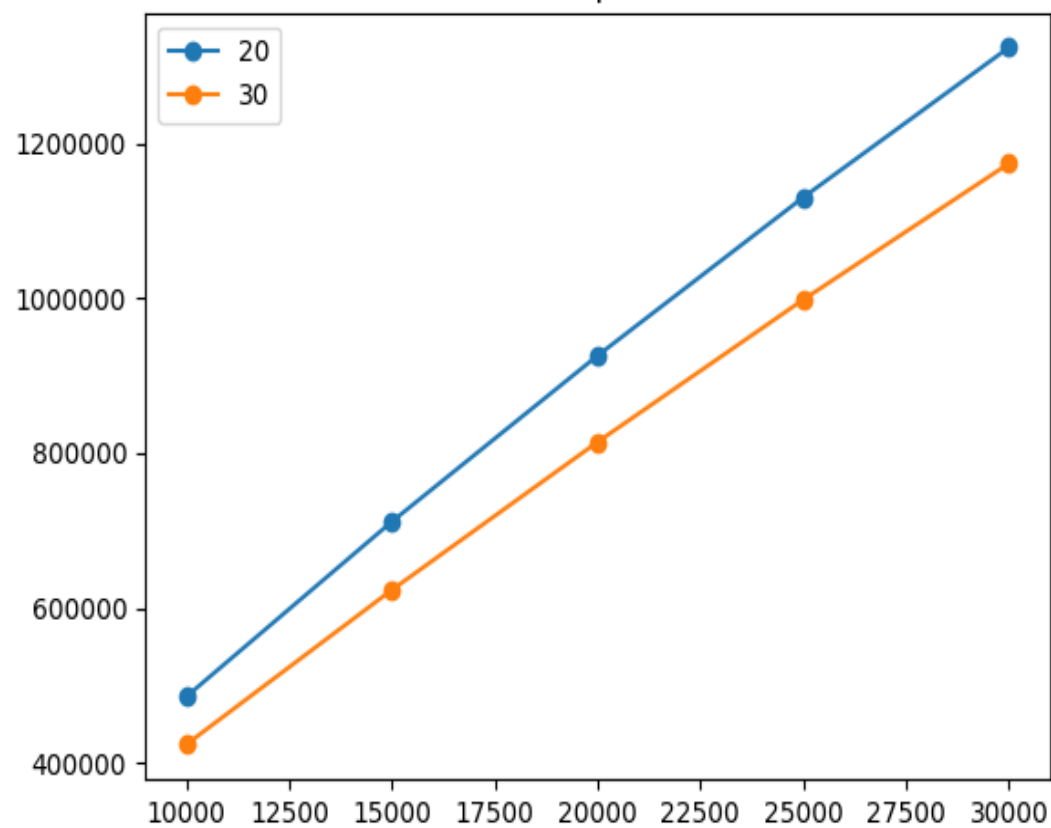


- KBF SPARSE RELAXED

FPR, Sparse Relaxed



k-mers size, Sparse Relaxed



ZAKLJUČAK

- POBOLJŠAN FPR
- MOŽEMO BIRATI SPARSE ILI KBF1 (ILI KBF2) IMPLEMENTACIJU OVISNO O ŽELJENOJ MEMORIJSKOJ POTROŠNJI ILI PRECIZNOSTI PROVJERE
- SMANJENJE FPR MJERE OMOGUĆUJE DALJNI NAPREDAK I RAZVOJ APLIKACIJA KOJE SE BAVE ANALIZOM SLJEDOVA