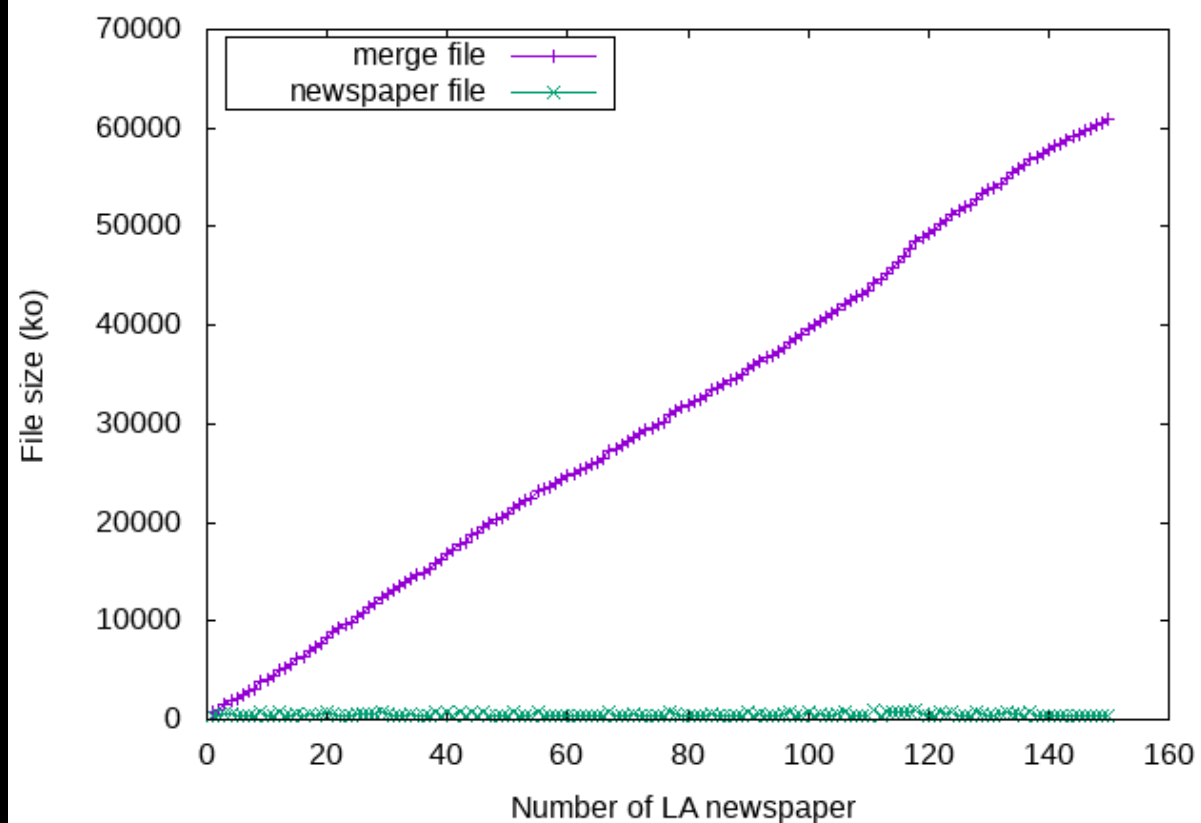


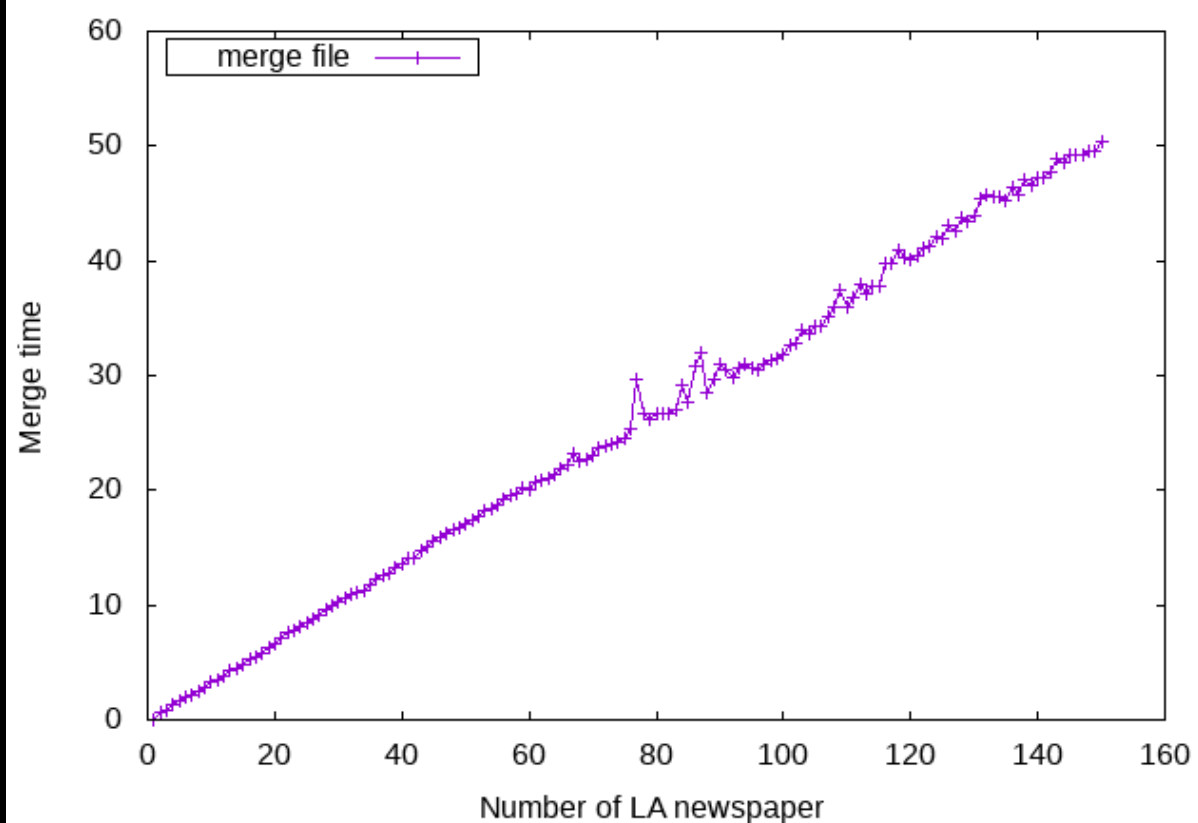
Axes de travail

- Tokenization
 - Stemming
- Inverted file
 - En mémoire
 - Merge
 - Binaire
- Query
 - Naïf (efficace)
 - Fagin (intéressant...)

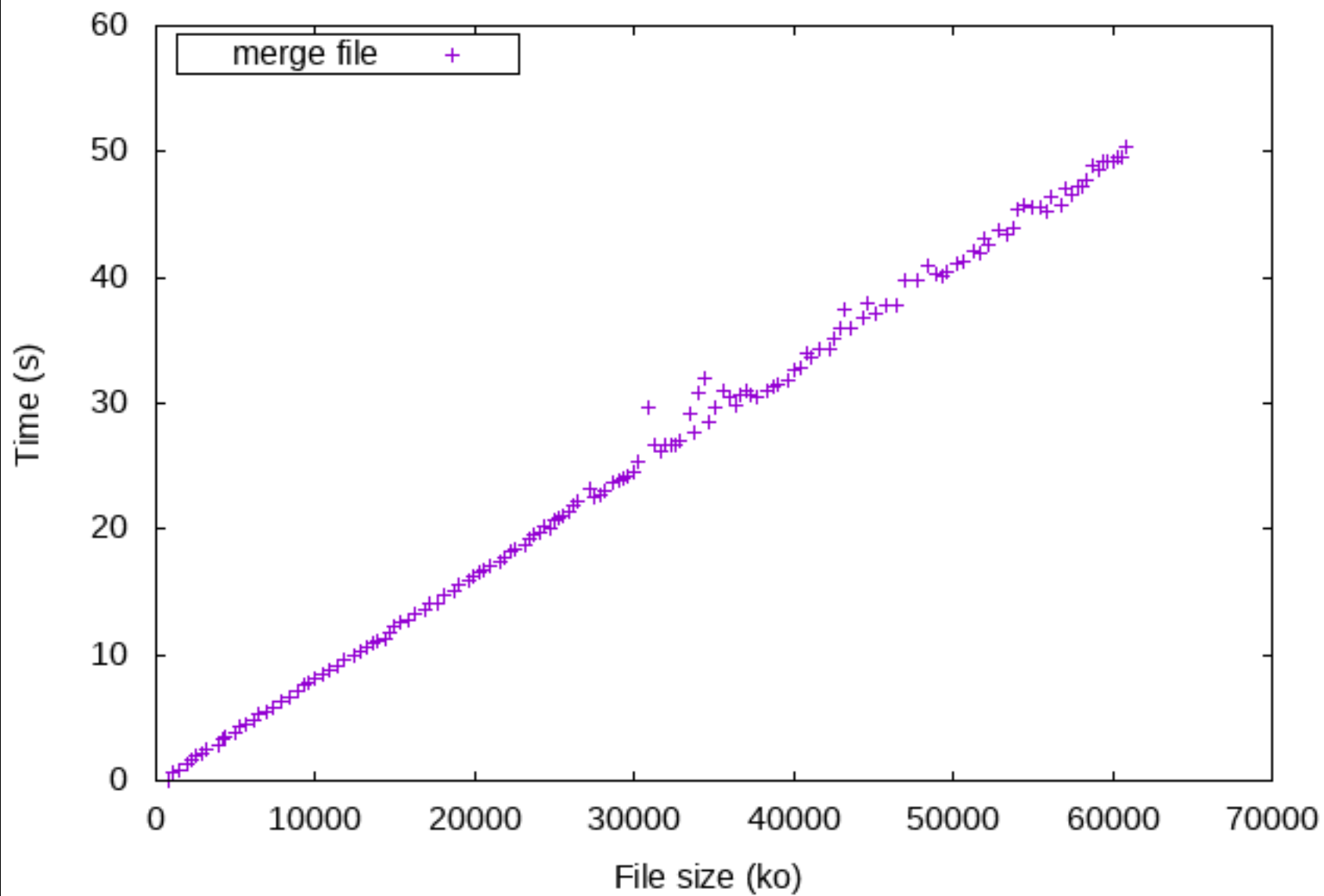
Size of the merge file depending of the number of already added newspaper

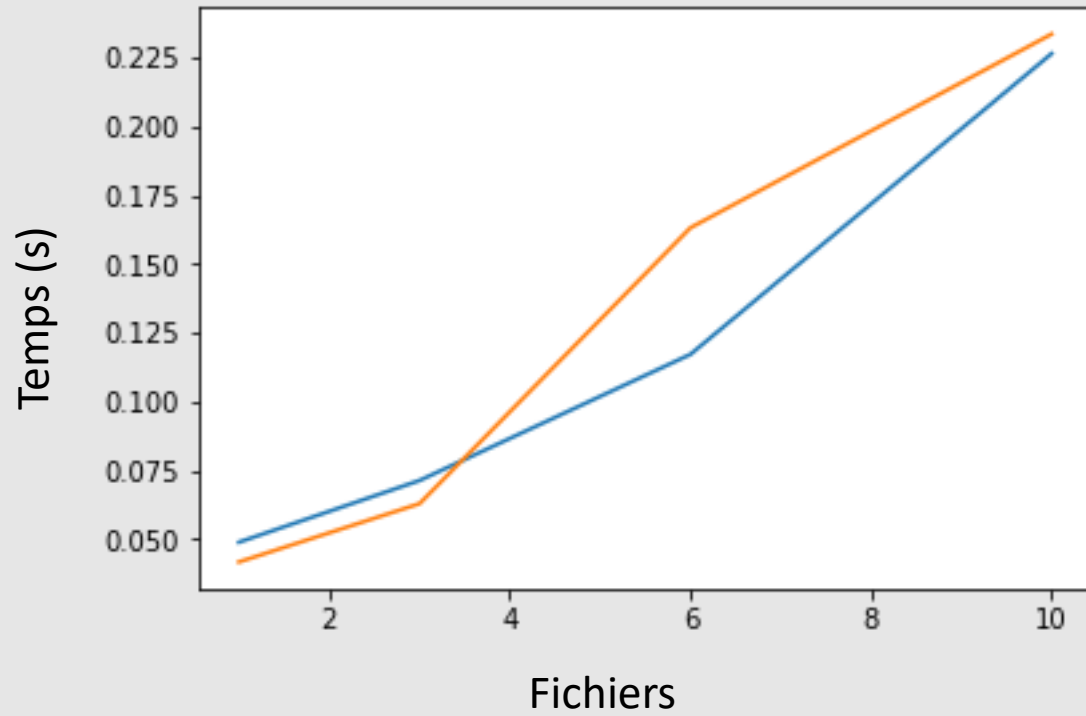


Time to add a new newspaper depending of the number of already added newspaper



Time of merge depending of the size of the merged file





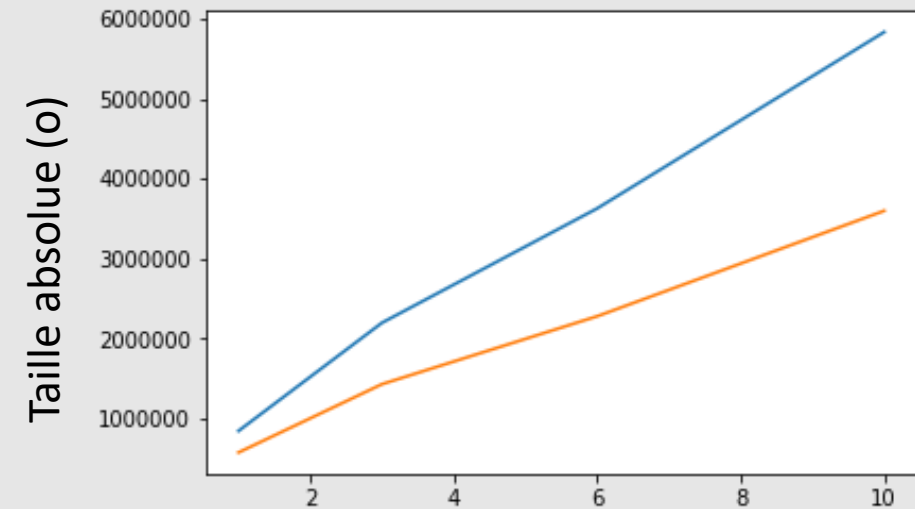
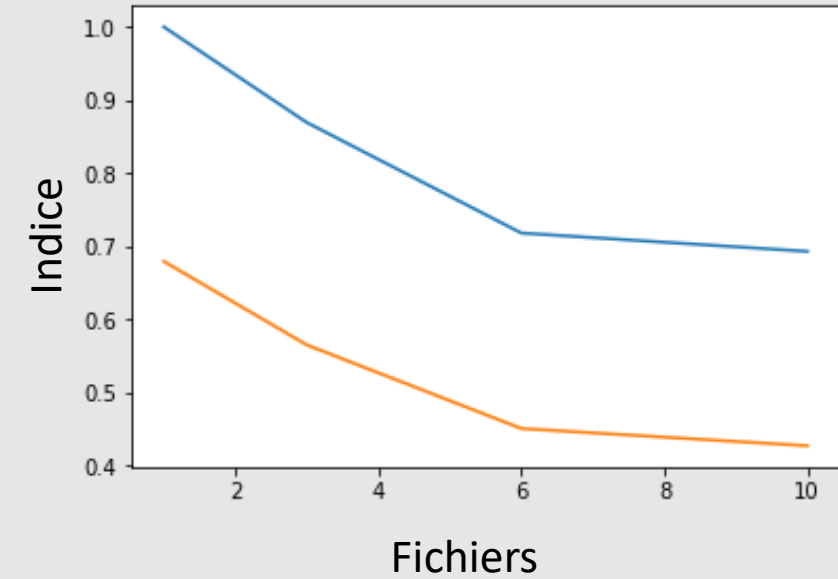
Temps pour lire une PL

Moyenne sur 10 lectures

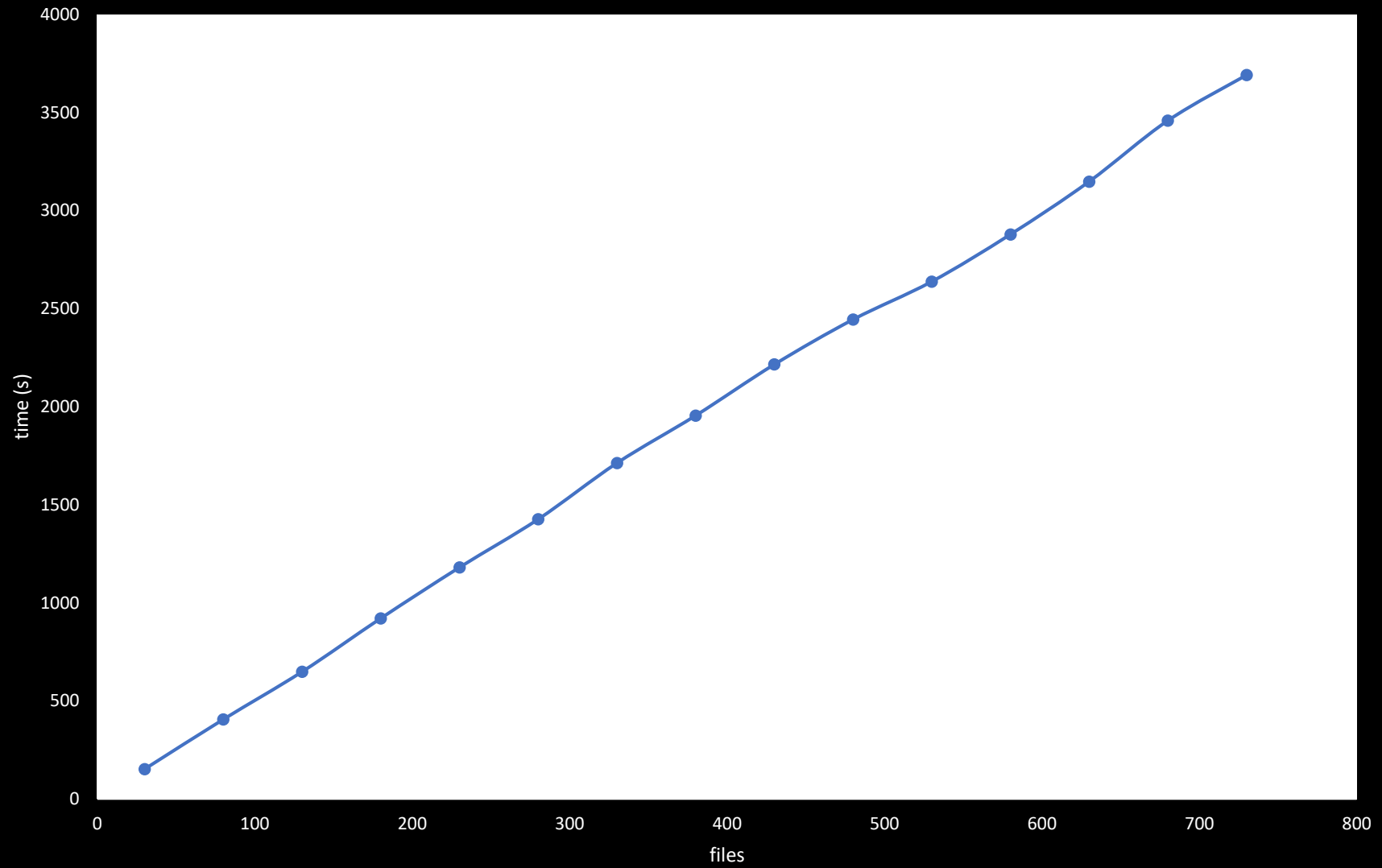
Bleu = naïf

Orange = smart

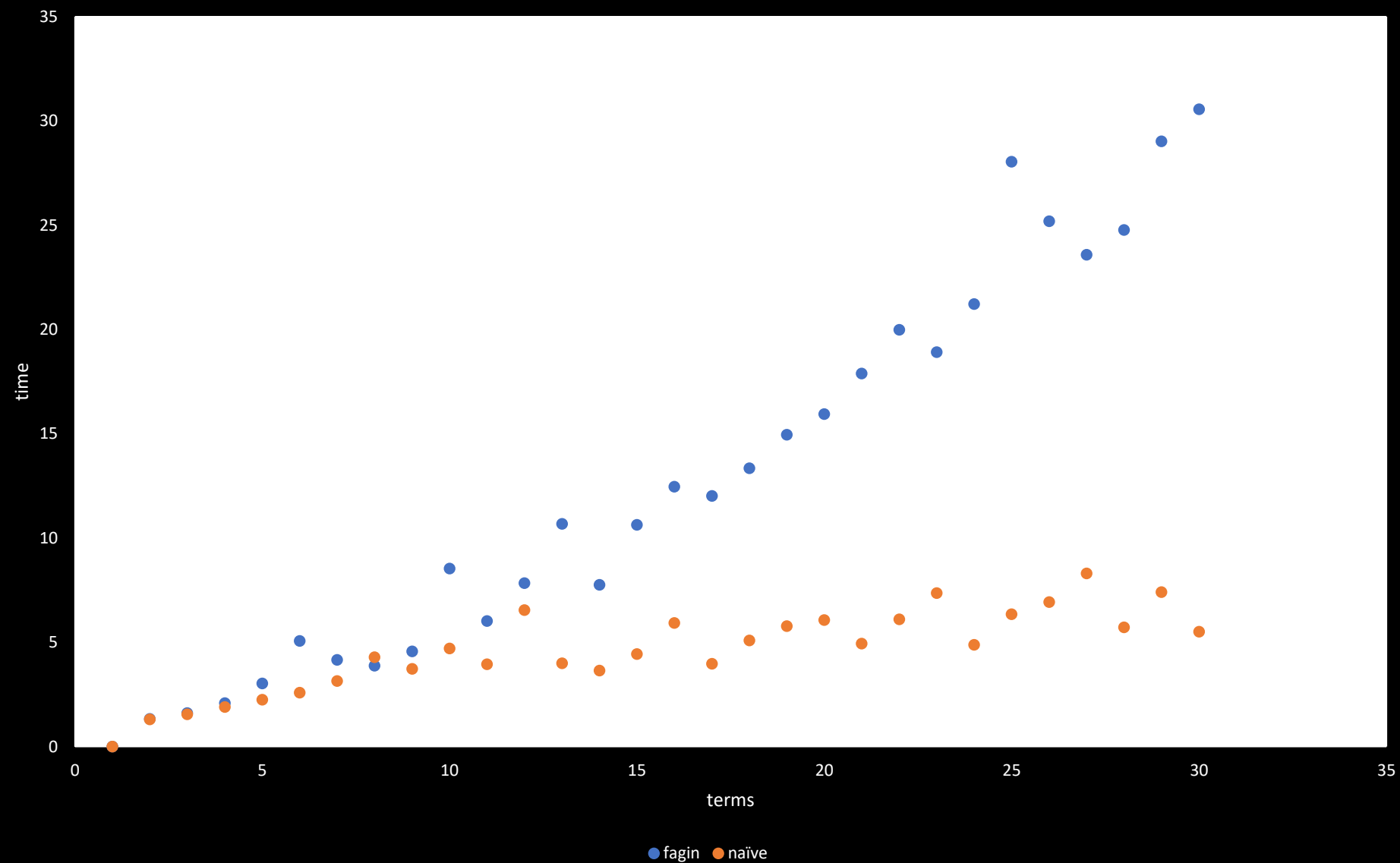
Taille relative d'un fichier dans un IF
Bleu = naïf
Orange = smart



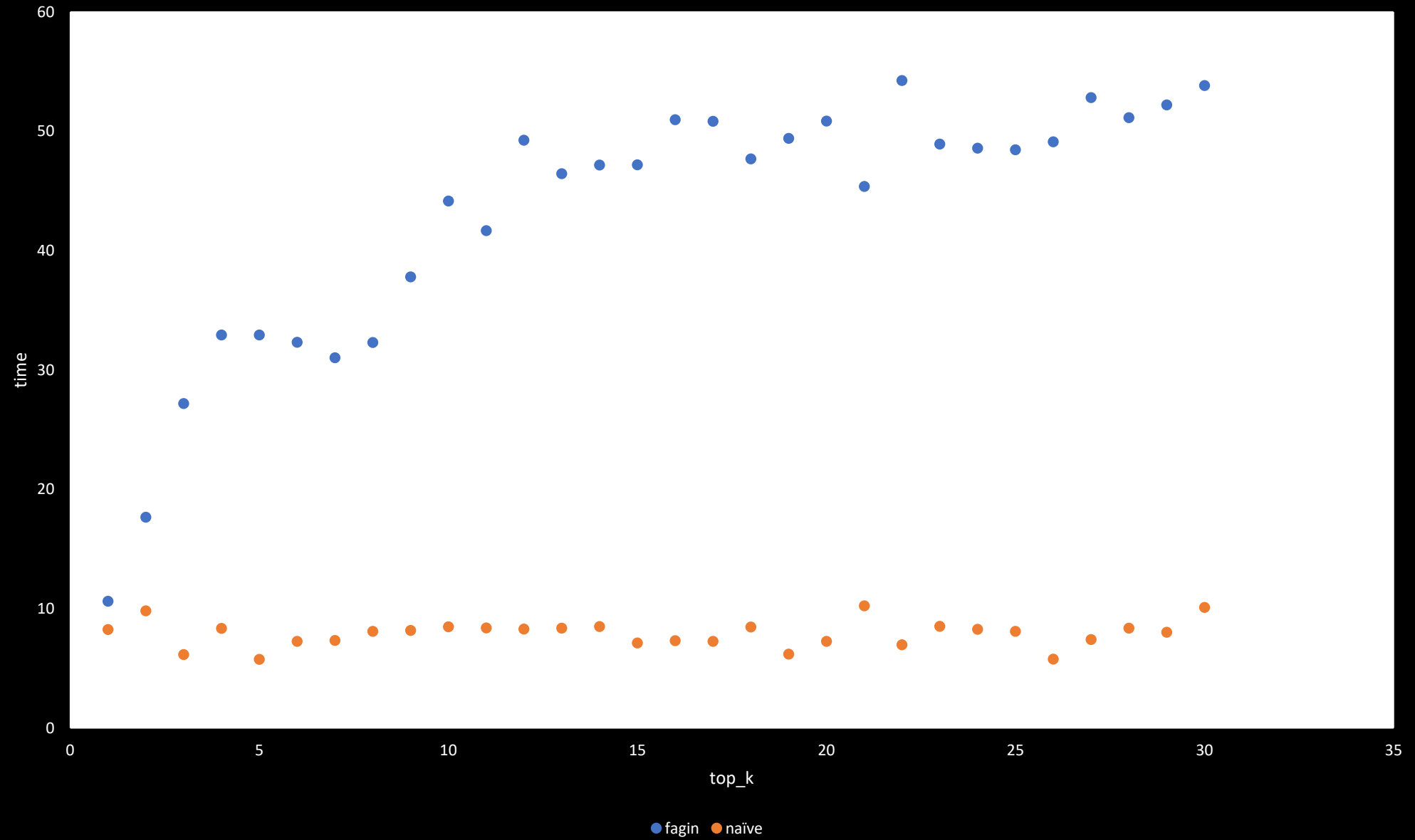
indexing time vs number of files



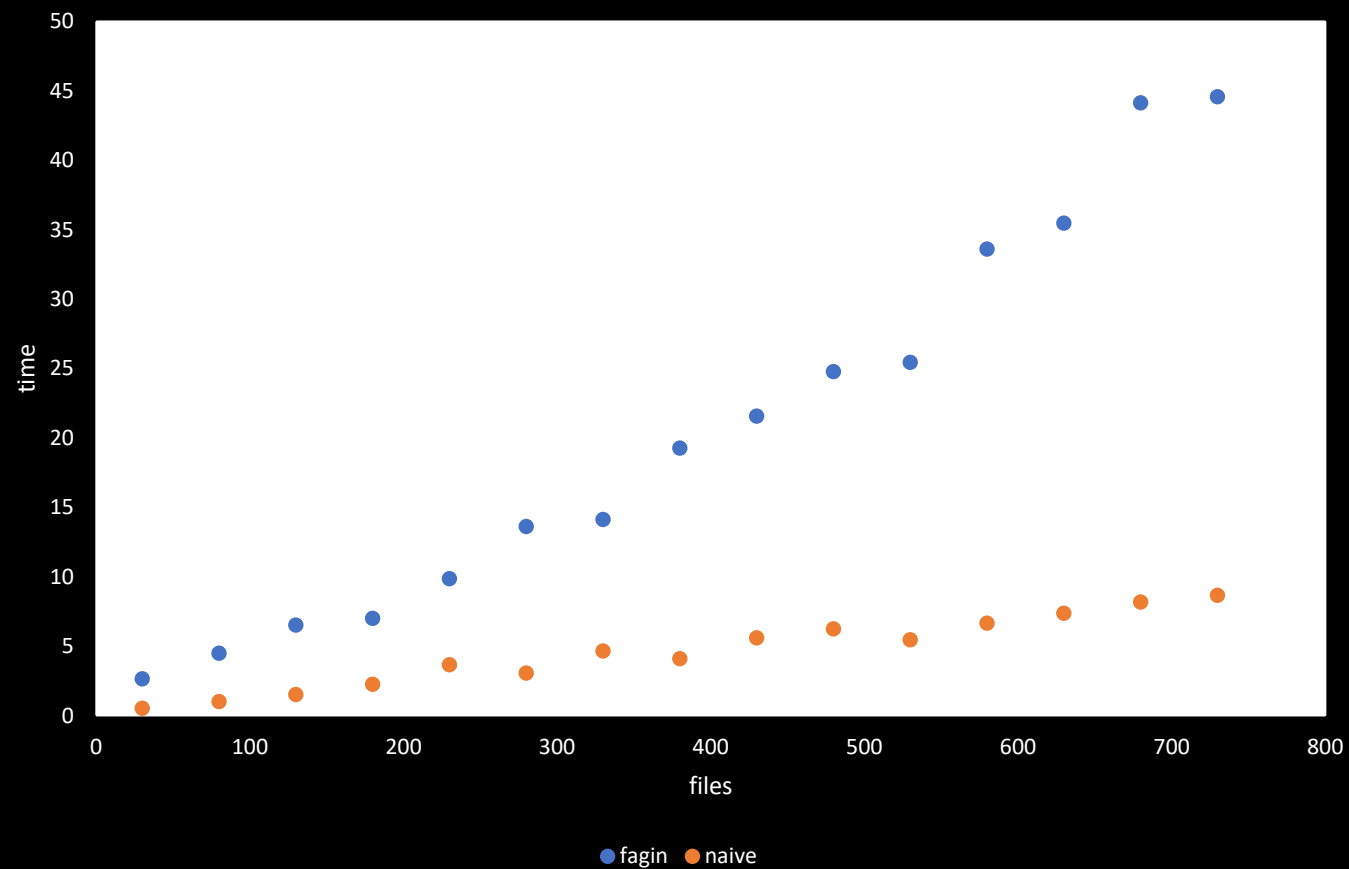
search time vs number of terms in query



search time vs number of results asked



search time vs corpus size



Naive only

