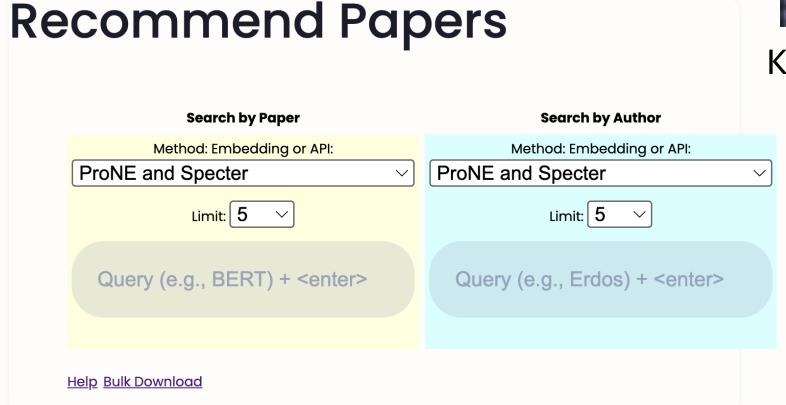
https://recommendpapers.xyz





Kenneth Church

Resources for Academic Search

- Web page
- API
- Bulk download of multiple embeddings, $Z \in \mathcal{R}^{n \times d}$
 - $S \in \mathcal{R}^{n \times n}$ is a similarity of n^2 papers for $n = 10^8$
 - S = ZZ^T where $10^2 < d < 10^4$ (hidden dimensions)
 - Some embeddings are based on BERT
 - and some are based on spectral clustering
- Claim: the embedding problem is bounded above and below by SVD
 - Consequence:
 - mini-batches cannot compute embeddings (without introducing losses)
 - because mini-batches are a FSM (constant space and linear time)



Kenneth Church