

Gabriel Martinica

ID: 40120255

For me this project was interesting. I talk about it more in the demo report. Noticing the differences between SPIMI and index naïve was not that new, I expected it. The biggest difference was the BM25 ranking which surprised me the most. This little equation is so powerful at ranking documents and I noticed as well how much preprocessing it needs. It does not use AI but still it is a good baseline. It was interesting to see as well that my implementation preferred shorter, term frequent documents and it ranked them higher. Makes sense but maybe it would be nice to not have document length matter that much, I feel a document can be long and still be very relevant. Maybe my algorithm decided to benefit lazy users 😊 If given more time, I would implement more efficient storage, faster indexing and try more with the  $k$  and  $b$  values.