Written and Developed by Manash Sarma

For any questions contact: SilvenLEAF@gmail.com

Problem and Objective:

Users want to easily search through all their files based on the inner content. They remember a piece from here and there but they forgot which file contained it. So they want a solution, to easily search across all the files and find out the one they are looking for. The solution must be fast and efficient but cost-effective.

Product Requirements:

1. User can connect their Dropbox account
2. After syncing their account, they can use the Full Text Search (FTS) across all their files
3. They can search through and filter out all files based on the inner content

Solution:

1. User will connect their Dropbox and then we’ll recursively get all files in it
2. We will extract the text content out of them
3. Then we will index them using ts-vector and store the file instance in our DB
4. Based on the user inputs and requirements we will filter out the files and show it to them

Mile Stones:

1. **(priority)** Create an API endpoint to connect and sync user’s Dropbox account
2. **(priority)** Extract all the files and upsert the file instance into our DB
3. **(priority)** Create a list API to show all the files, supporting FTS search.

There will be two types of FTS Search supported.

1. OR Search: At least one of the given texts must be in that file
2. AND Search: All of the given texts must be in that file
3. Support Normal Search, Sort and Pagination
4. Write automated tests, both **positive** testsand **negative** tests and test out the API endpoints
5. **(priority)** Create a well written API Documentation File for the Frontend Developer
6. Preferably add CI/CD using Github actions

Expected Timeline:

1. Can be at max 3 days. In reality can take less but just in case for safety and last minute changes.
2. In reality it was done in 2 days

Out of Scope and QUESTIONS:

1. **Question:** What happens if the user updates the file or deletes the file after syncing with our app?

**Solution:** We can either write a cron script that will sync the user’s account in a specific time periodically OR We can create a webhook with Dropbox so that whenever our users do any CRUD operations with any Dropbox files, it will upsert the specific record into our DB

1. **Question:** What happens if the user wants to search across all other connected users’ files? OR what if (s)he wants to search only his own files and ensuring that his files can not be search by other users?

**Solution:** We can make a secure auth flow surrounding that. Using private public flags and strict filtering and authenticating the user being the owner of that file

1. **Question:** What happens if the user doesn’t trust us? I mean (s)he is uncomfortable being their file contents being extracted by us? What if they want to make sure we don’t see a single thing about their file content?

**Solution:** We can hash the content in column 1 and in column 2 we can store the delimited hashed, ordered individual words and hash the search term and try finding a match in second column. But this approach will be inefficient. Can’t think of a better solution for this. Need to discuss with a Senior Engineer and check if we can do something about it.