

INDEXER NOTES

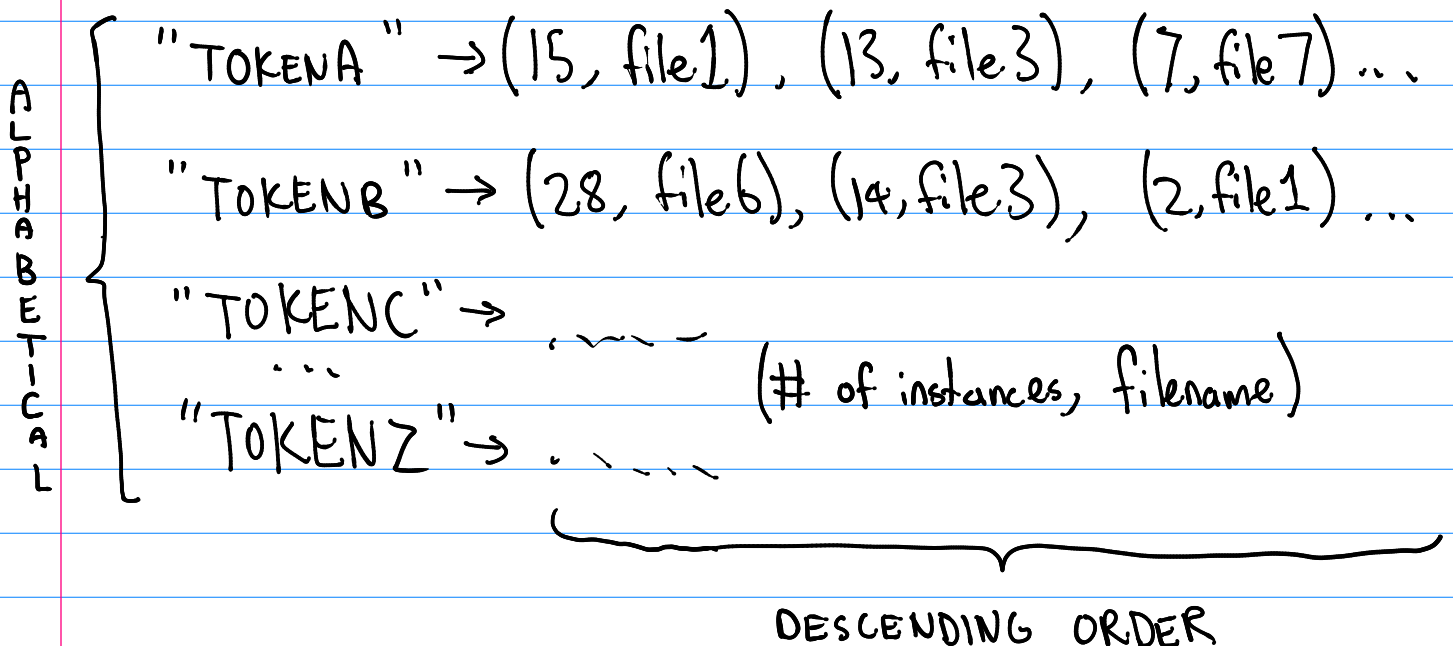
PROCESS:

1. TAKE A FILENAME/DIRECTORY AS INPUT
2. TOKENIZE THE CONTENTS OF EACH FILE
3. COLLECT INFO ABOUT EACH TOKEN IN TERMS OF THE TOKEN, WHAT FILES IT APPEARS IN, AND HOW MANY TIMES IT APPEARS IN THAT FILE
4. STORE THIS DATA INTO A FILE

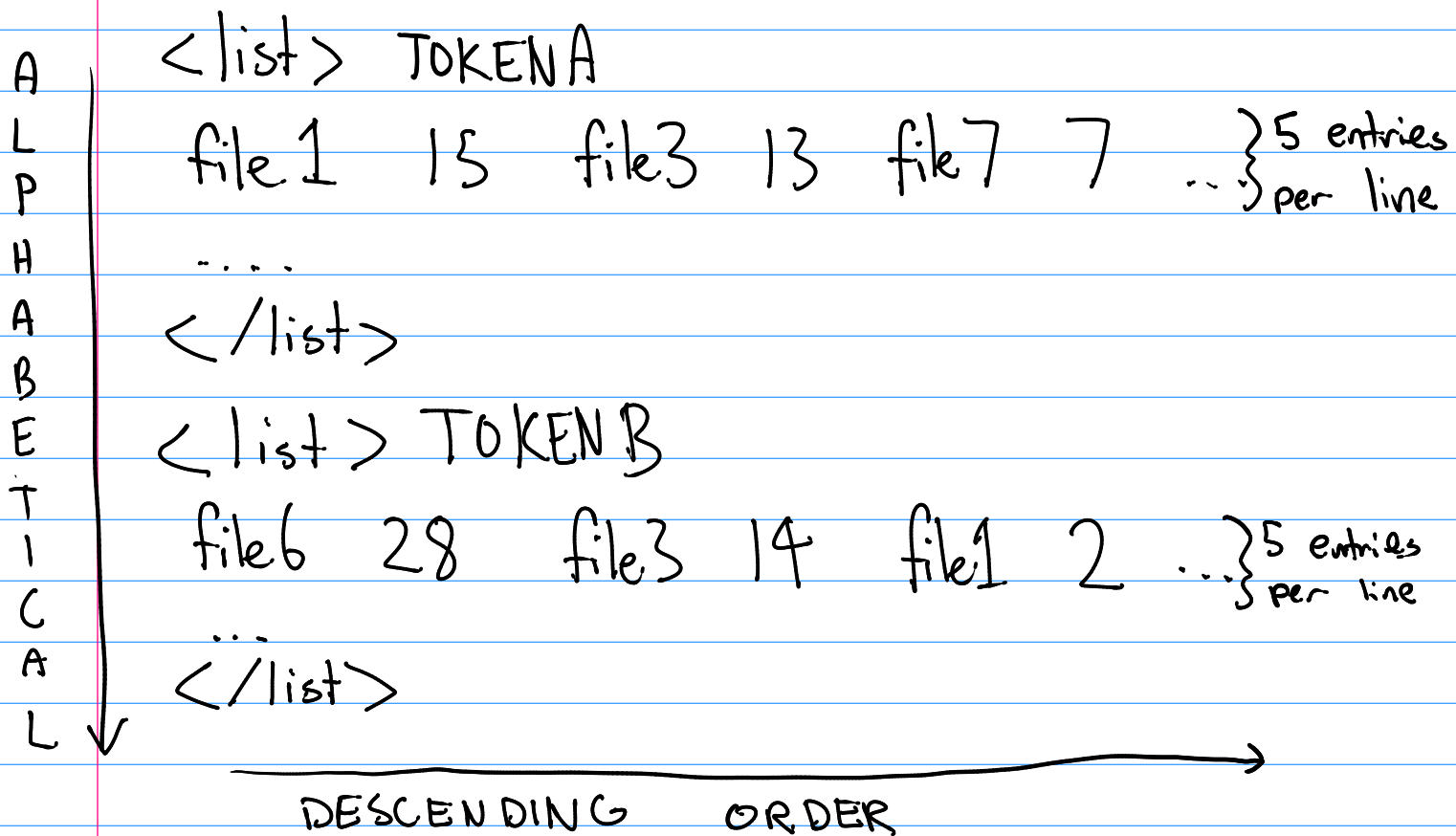
WE NEED

- A TOKENIZER!
- A DRIVER TO CALL THE TOKENIZER ON ALL FILES!
- A WAY TO WRITE MEMORY (THE TOKENIZER OUTPUT) TO FILE!

MEMORY STRUCTURE



FILE STORAGE STRUCTURE



ON MEMORY STRUCTURE:

↳ HASH TABLE OF LINKED LISTS?

→ needs varying number of keys

→ each token must be unique key

↳ what hash function would that even be