

Tokenizer Lab

Name: Ahmed

Lab: Tokenizer

Application Architecture:

The application consists of a Tokenizer class which reads a file, extracts words, and stores them uniquely using an `unordered_map`. Each unique word is assigned an incremental integer token ID.

The Tokenizer class provides functions to retrieve the token by word and retrieve the word by token.

Software Pattern Used:

- Modular Structure: The code separates concerns between tokenizing functionality and unit testing.
- Mapping Pattern: `unordered_map` is used for efficient token-to-word and word-to-token mapping.