

Text Concordance

APPLIED DATA
STRUCTURES

Defining the Problem

Objective

- provide users with word suggestions for terms that are used repetitively in a text.
- It also allows users to search for alternative word suggestions when they are unsure how to continue or complete a sentence.

Current Problem

A common challenge: frequent reuse of the same words due to limited vocabulary or uncertainty about appropriate word choice.

- repetition → reduces clarity, weakens expression, prevents the intended meaning

Significance for users

Helps users enhance their communication skills, shaping personal and professional outcomes.

used in workplace communication, interactions with family or friends abroad, and making typing quicker and seamless.

DATA STRUCTURES USED

01

Trie

Takes in two words, the current word and the preceding word, stores the words in a tree like system and flags when word is complete. → predicts next word.

02

Map

Takes words that are assigned keys, uses hash function, stored in set of lists with frequency, sorted.

03

Usage Of WordNet

Lexical database that is used for finding alternative words.

WHY NOT OTHER STRUCTURES?

01 Trie

A trie was chosen because autocomplete and next-word prediction depend on efficient prefix matching rather than full-word lookup. Tries support prefix search in $O(m)$ time, where m is the length of the word, independent of the number of stored words. This makes them ideal for real-time suggestions while typing.

Alternatives such as arrays or linked lists require linear scans, while hash tables only support exact-match lookup and cannot efficiently handle prefix queries. Balanced trees store words in order, but still perform poorly for prefix-based traversal compared to tries.

WHY NOT OTHER STRUCTURES?

02 Map

A map was chosen to store word frequencies because the problem naturally fits a key-value model, with each word acting as a key and its frequency as the value. When implemented using a self-balancing tree or hash-based structure, maps provide efficient insertion and lookup while allowing easy retrieval of the most frequent words. Arrays or linked lists are inefficient due to costly searches and updates, while using a trie alone complicates frequency handling and global ranking. The map provides both conceptual clarity and strong performance guarantees.

WHY NOT OTHER STRUCTURES?

03

Usage of WordNet

WordNet was chosen because it organizes words into sets based on meaning, not just surface similarity, allowing more accurate synonym suggestions. This makes it appropriate for assisting users with word variety without changing the intended meaning of a sentence. Simpler synonym lists lack structure, and machine-learning models would significantly increase system complexity without being aligned with the goals of a data structures course. WordNet offers a reliable, well-established linguistic resource that complements the core data structures without replacing them.

Main Features/ Interface of the Program

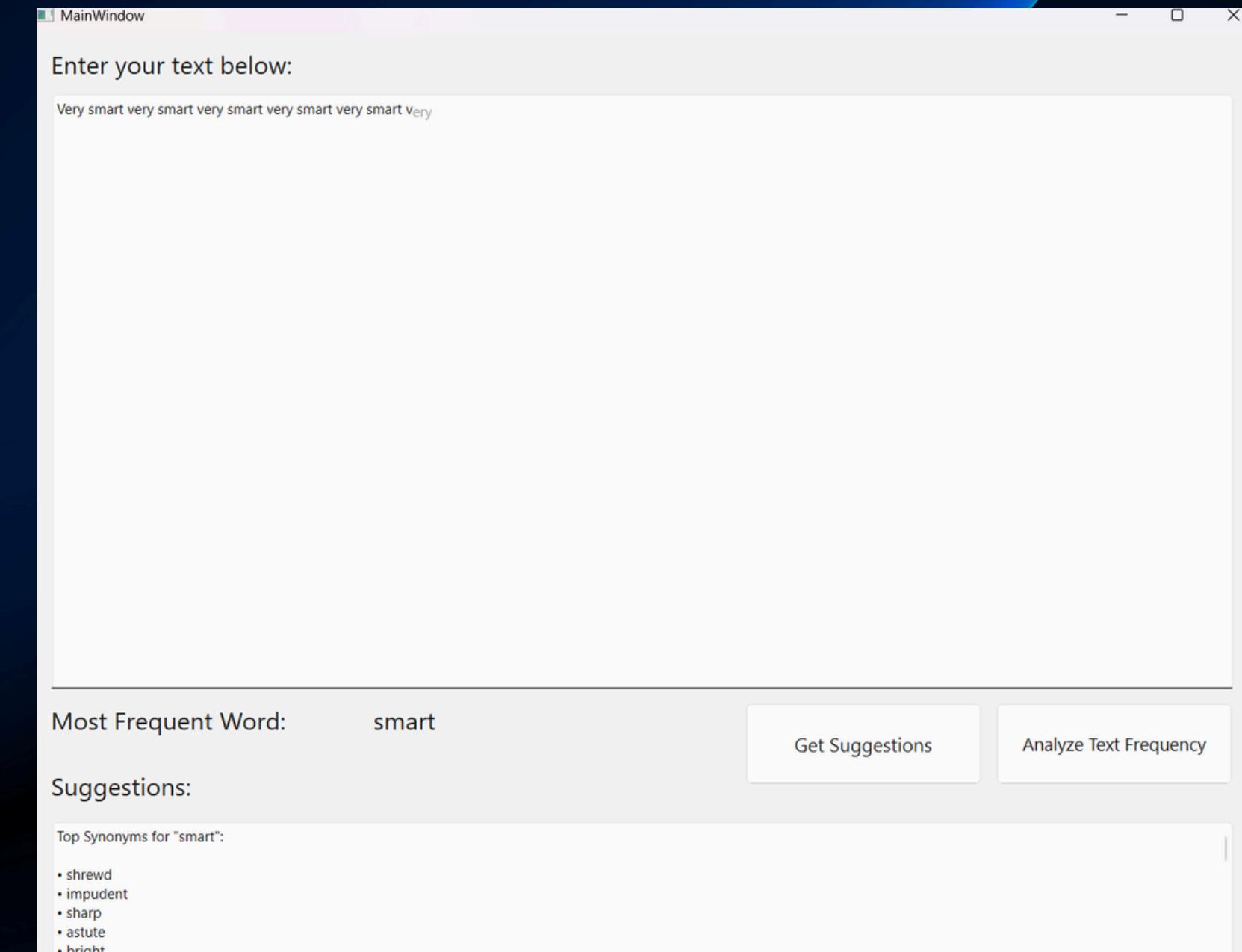
Analyze Text Frequency

Takes in the words of the text, sorts by frequency using map, and displays to the user the most repetitive word.

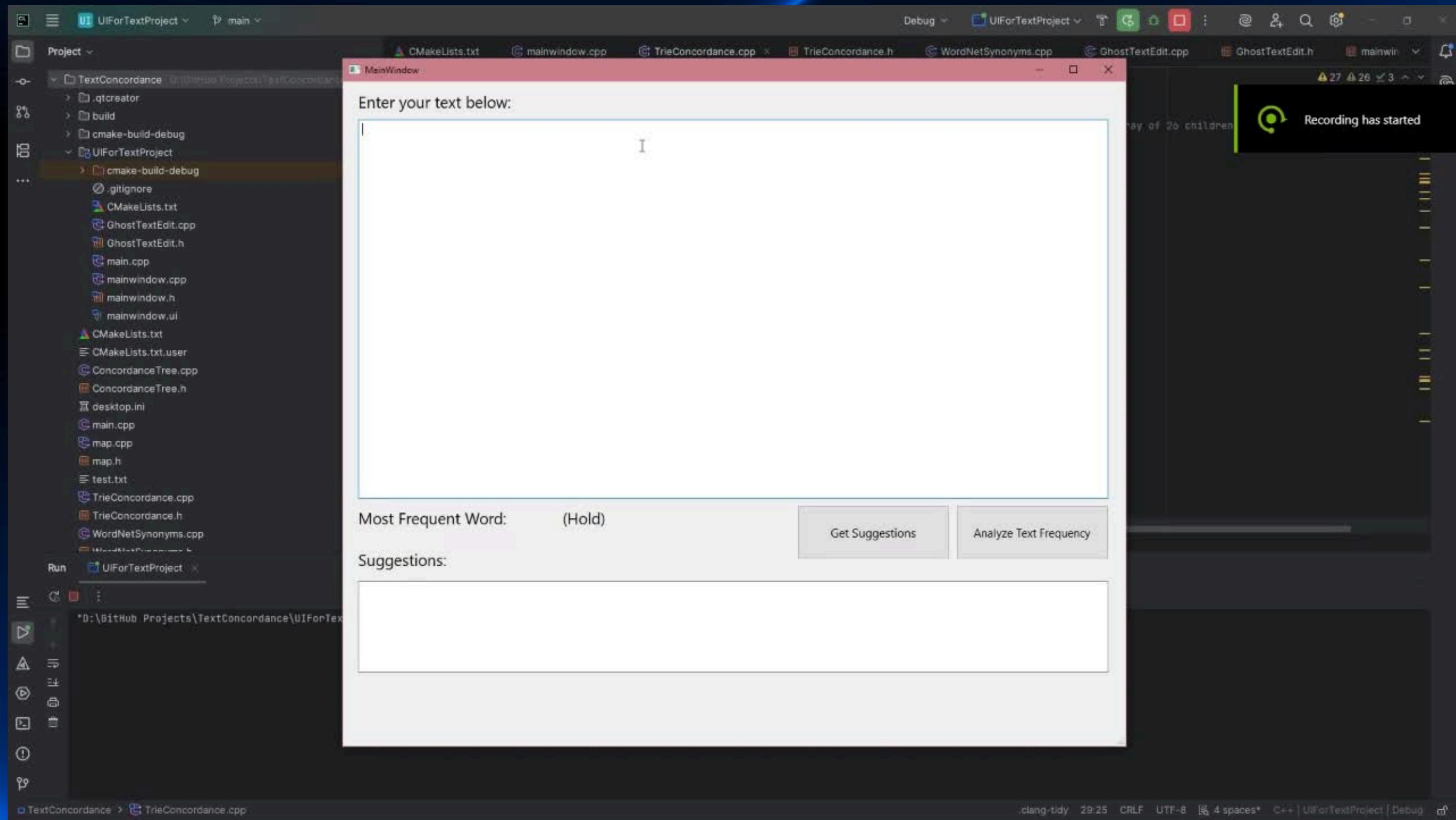
Get Suggestions

Uses WordNet to get a proper replacement/synonym for the most frequent word.

Autocomplete → shown in real time.



DEMO



Analysis and Critique

Our main points of limitations or possible improvement are summed up in the following points:

- Implementing a time- or keystroke-triggered refresh of the map and trie.
- Expanding the map for phrases in addition to words.
- WordNet does not provide synonyms for some words, so a backup API is needed.
- A professional makeover for the UI.
- The word count limit may be an issue for larger projects like books.

Marketing Usages

01 Educational impact

Helps users become aware of repetitive patterns in their writing and encourages step by step vocabulary improvement

02 Productivity & efficiency

Reduces time spent rewriting text and offers suggestions while typing, not after.

03 Professional communication

- Word repetition is a known issue in:
 - Emails, Reports, and Proposals
- This tool helps produce more concise, varied, and professional text.

Learning Outcomes

- Optimization and complexity analysis (comparing different data structures to find the most suitable ones)
- API integration
- Collaborative development and task delegation
- Input cleaning (ETL)

THANK YOU