Hotel Price Analysis (Variant 5)

FANTASTIC FIVE

Subhram Satyajeet – 110127932

Kashyap Bharatbhai Prajapati – 110126934

Sachreet Kaur – 110122441

Aditya Kunal Bhate – 110125765

Samiksha Arora – 110128455

Agenda

- Web Crawler
- Individual Contribution
- Architecture
- Modules Insights
- Demo
- Questionnaire

Individual Contribution

Subhram Satyajeet – Edit Distance/Spell Checker, Best Deals Recommendation based on price.

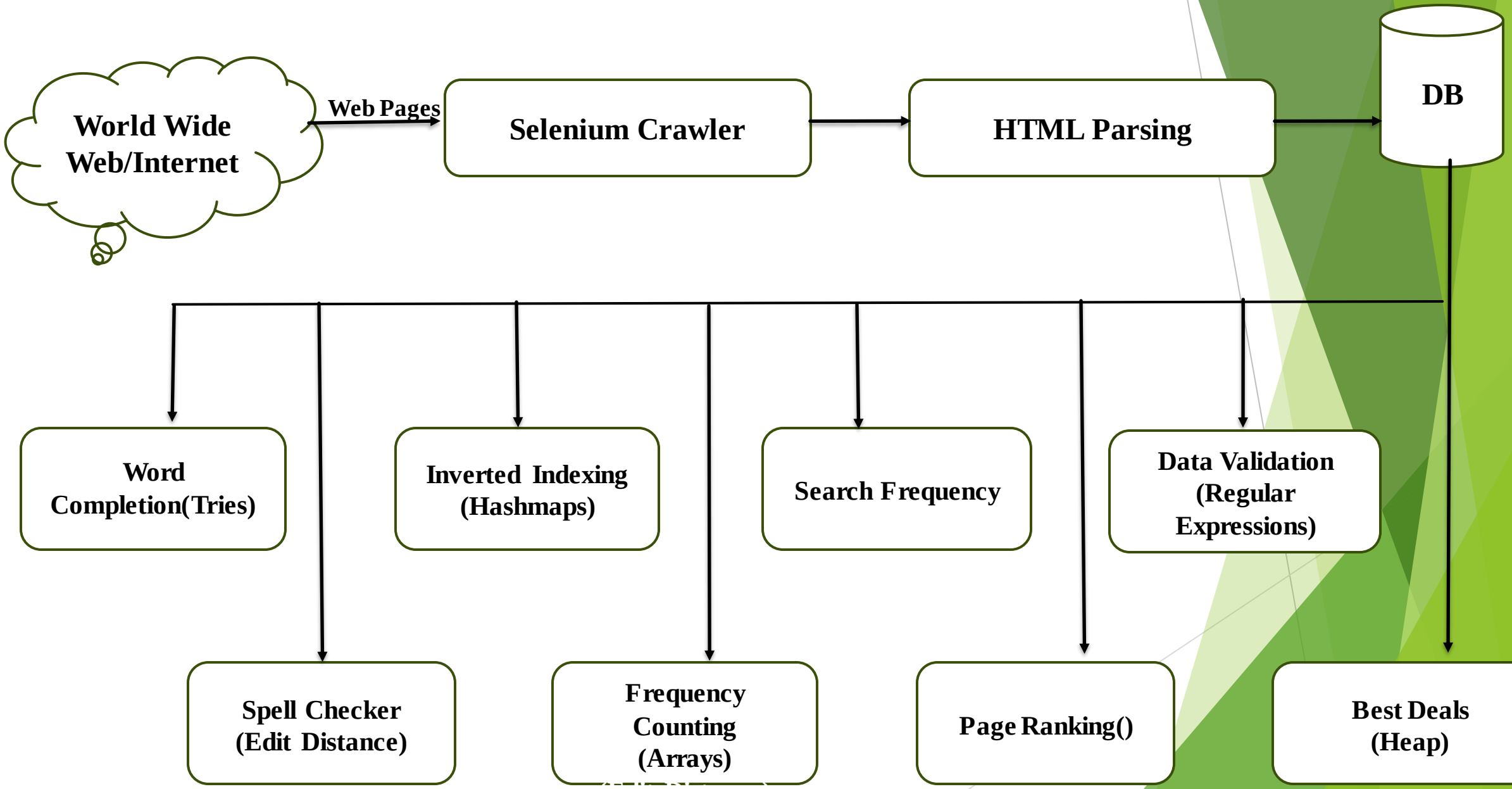
Kashyap Bharatbhai Prajapati – Web scraping/crawling, HTML Parsing.

Sachreet Kaur – Word Completion and suggestion using Trie data structure, HTML Validation.

Aditya Kunal Bhate – Inverted Indexing and Pattern matching using KMP.

Samiksha Arora – Page Ranking, Frequency Count

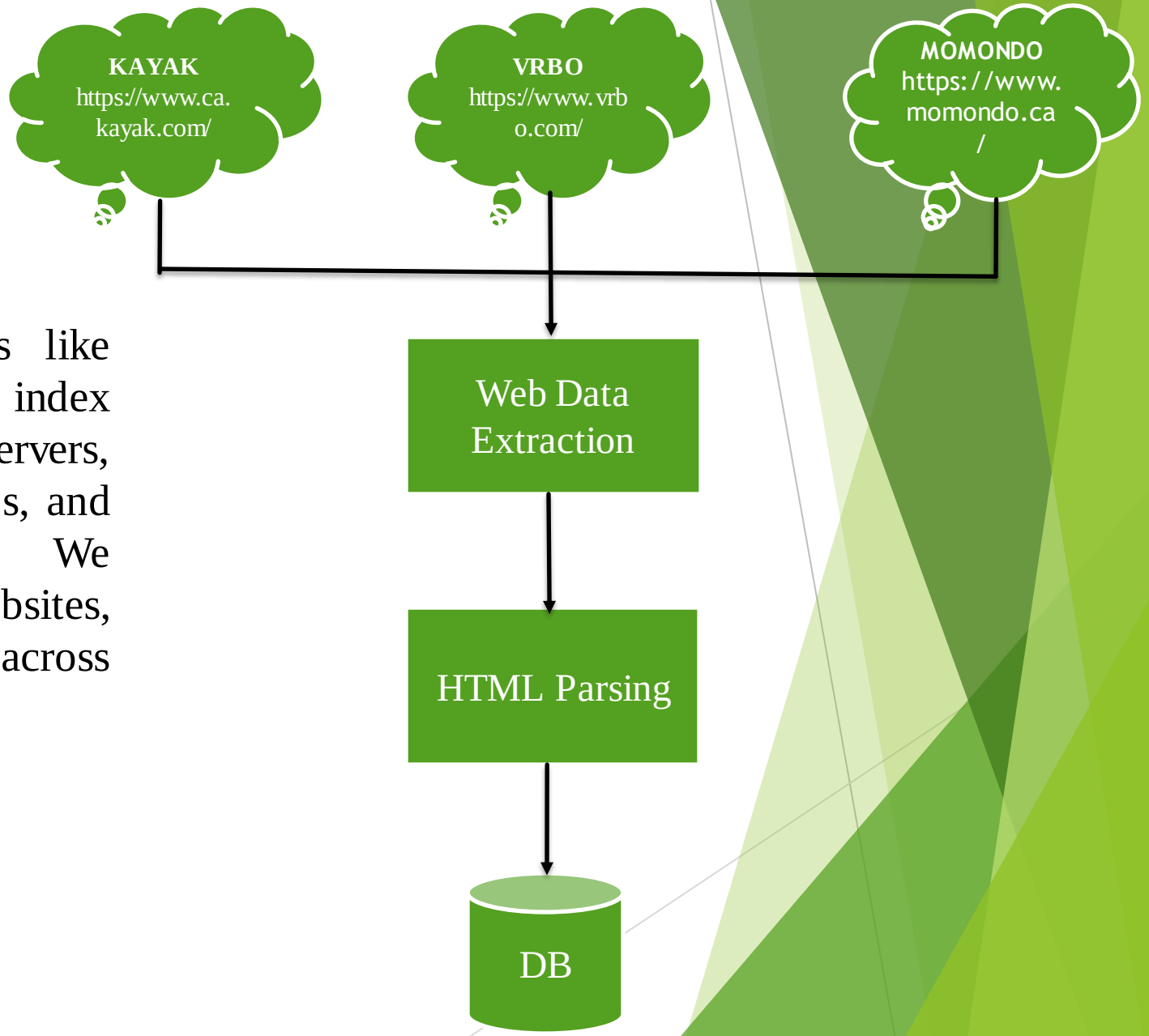
Architecture



Modules Insights

Web Crawler:

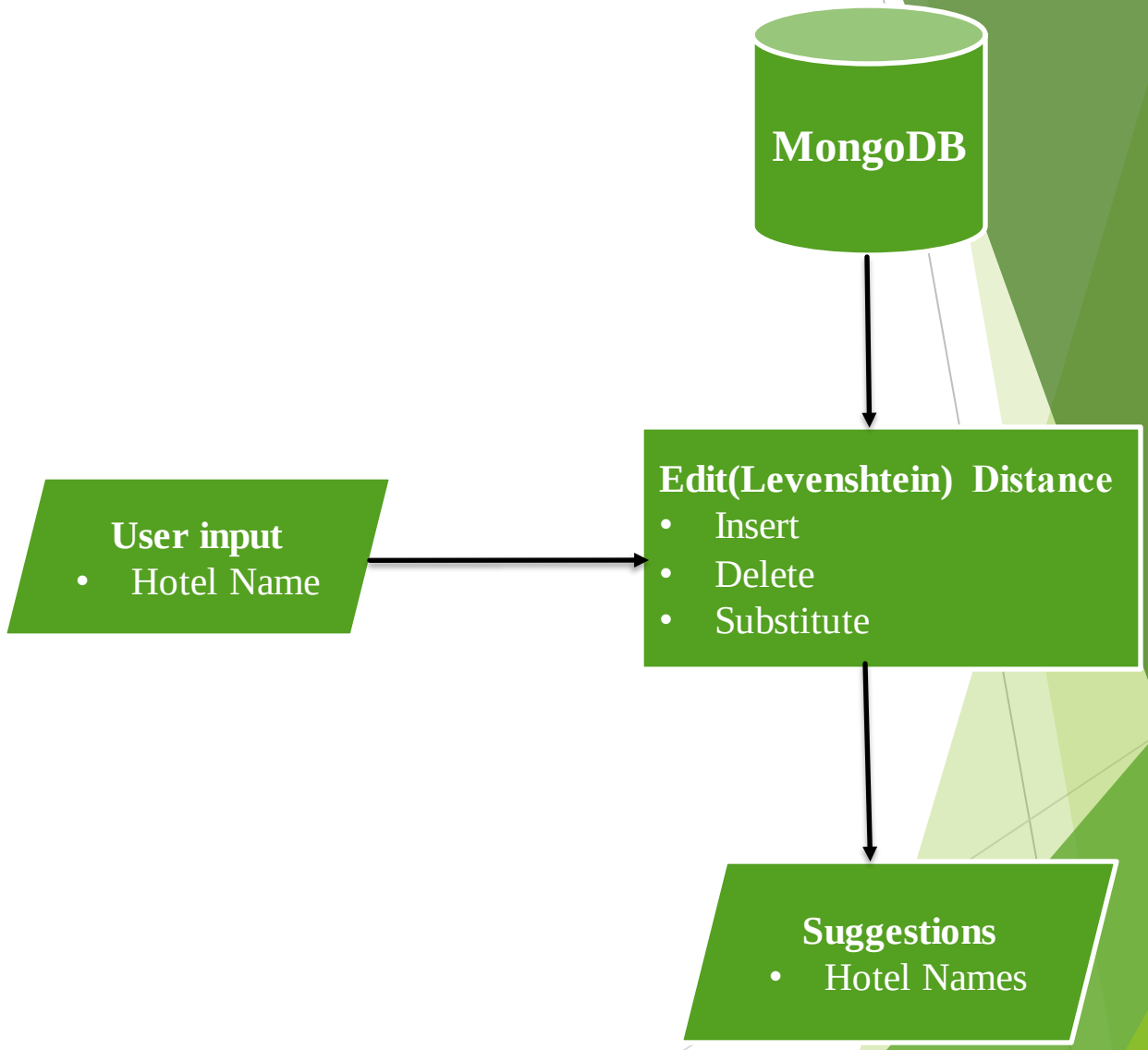
Web crawlers, essential for search engines like Google and Bing, systematically explore and index web pages by initiating requests to web servers, accessing and parsing content, extracting URLs, and recursively navigating through pages. We implemented this feature to crawl the hotel websites, extract important information, and store them across the Mongo DB database.



Module Insights

Spell Checker:

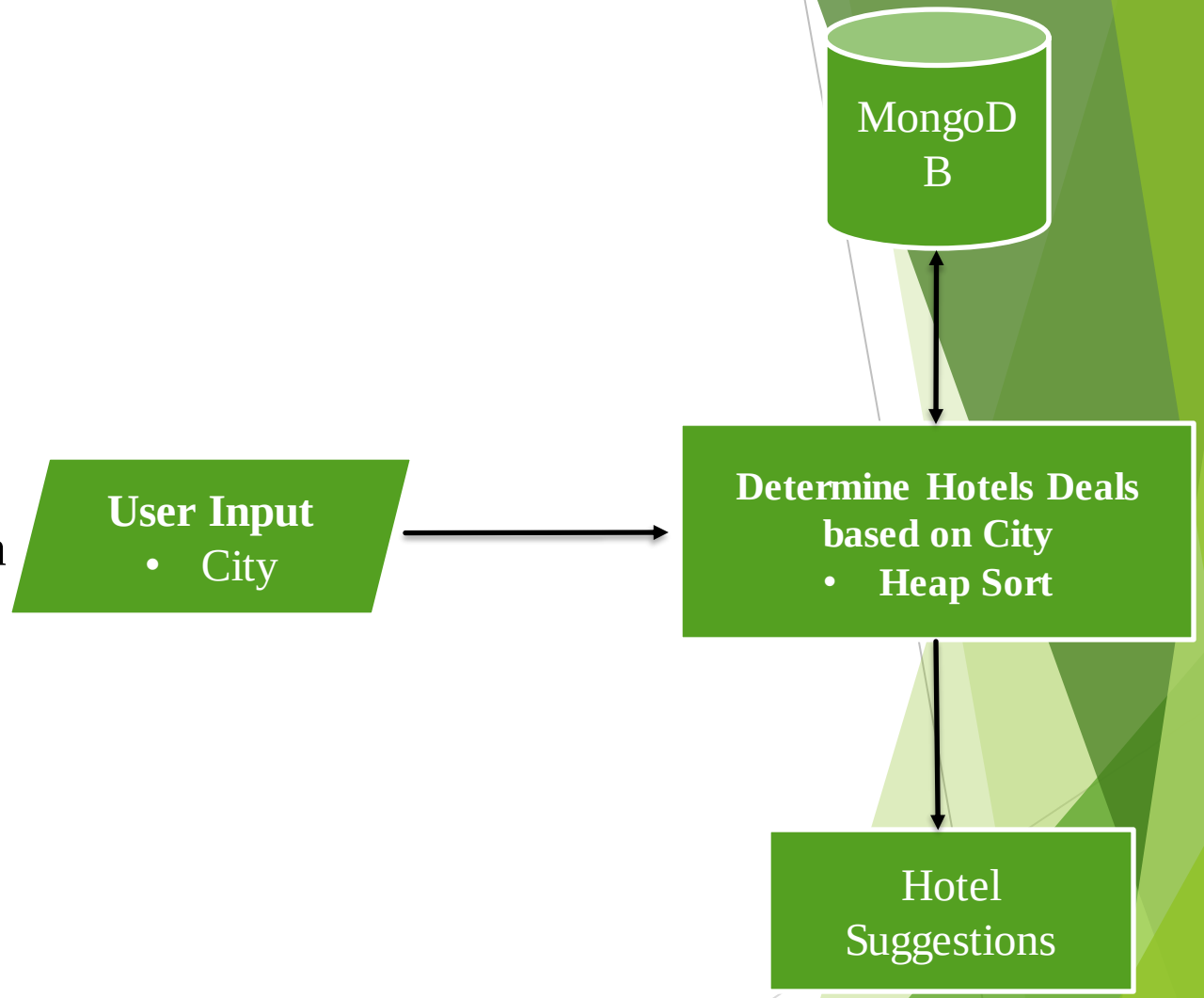
Spell Checker functionality is based on edit distance algorithm using dynamic programming techniques. This module takes user defined hotel as an input and check the matching possible hotel names based on the edit distance and displays the hotel names with minimal alphabetical changes.



Module Insights

Best Deal Suggestions:

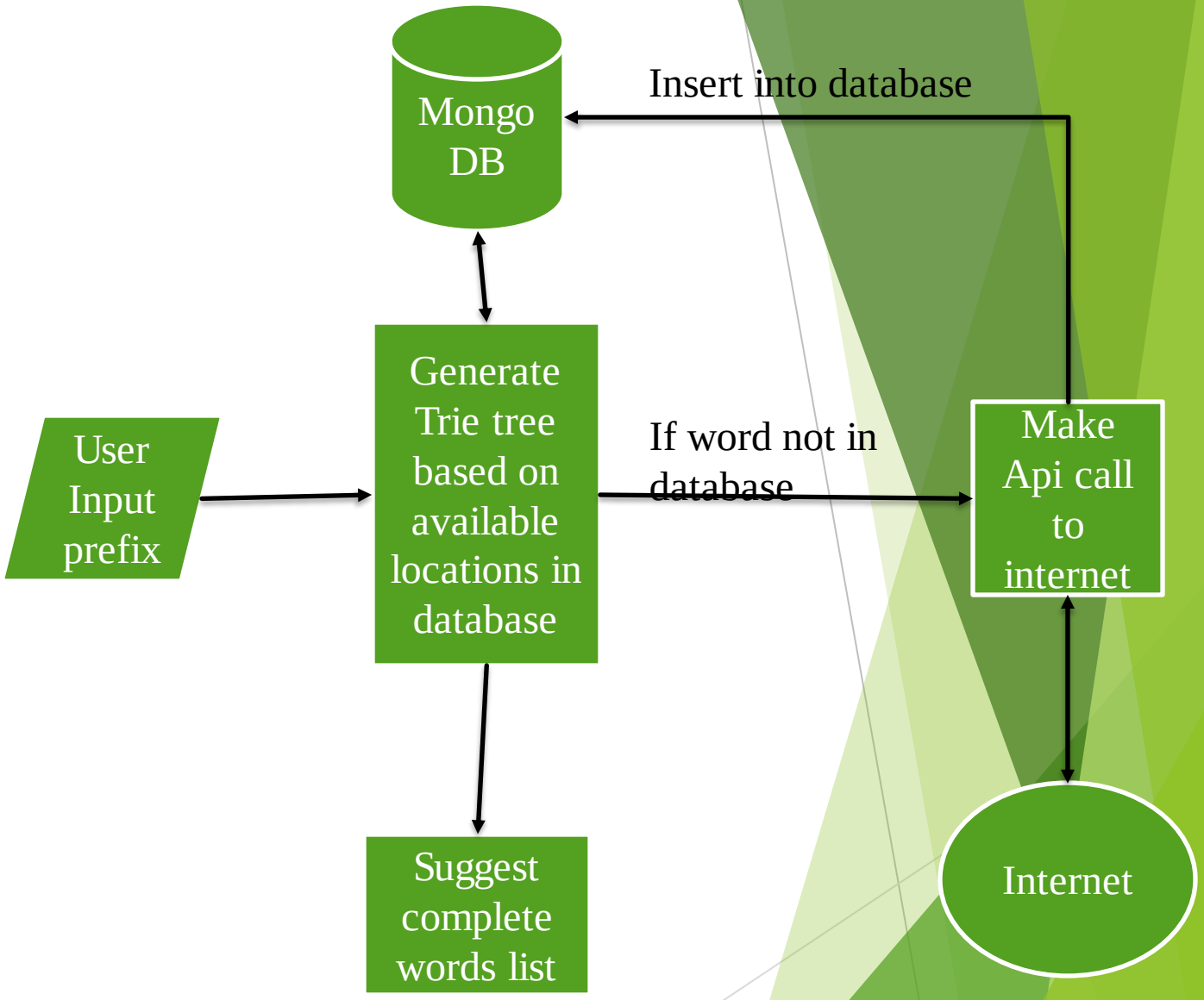
This functionality reads crawled hotel data from the MongoDB database and suggests the best deals of hotels among all three websites based on the user's selection of city.



Module Insights

Word Suggestions/Completions:

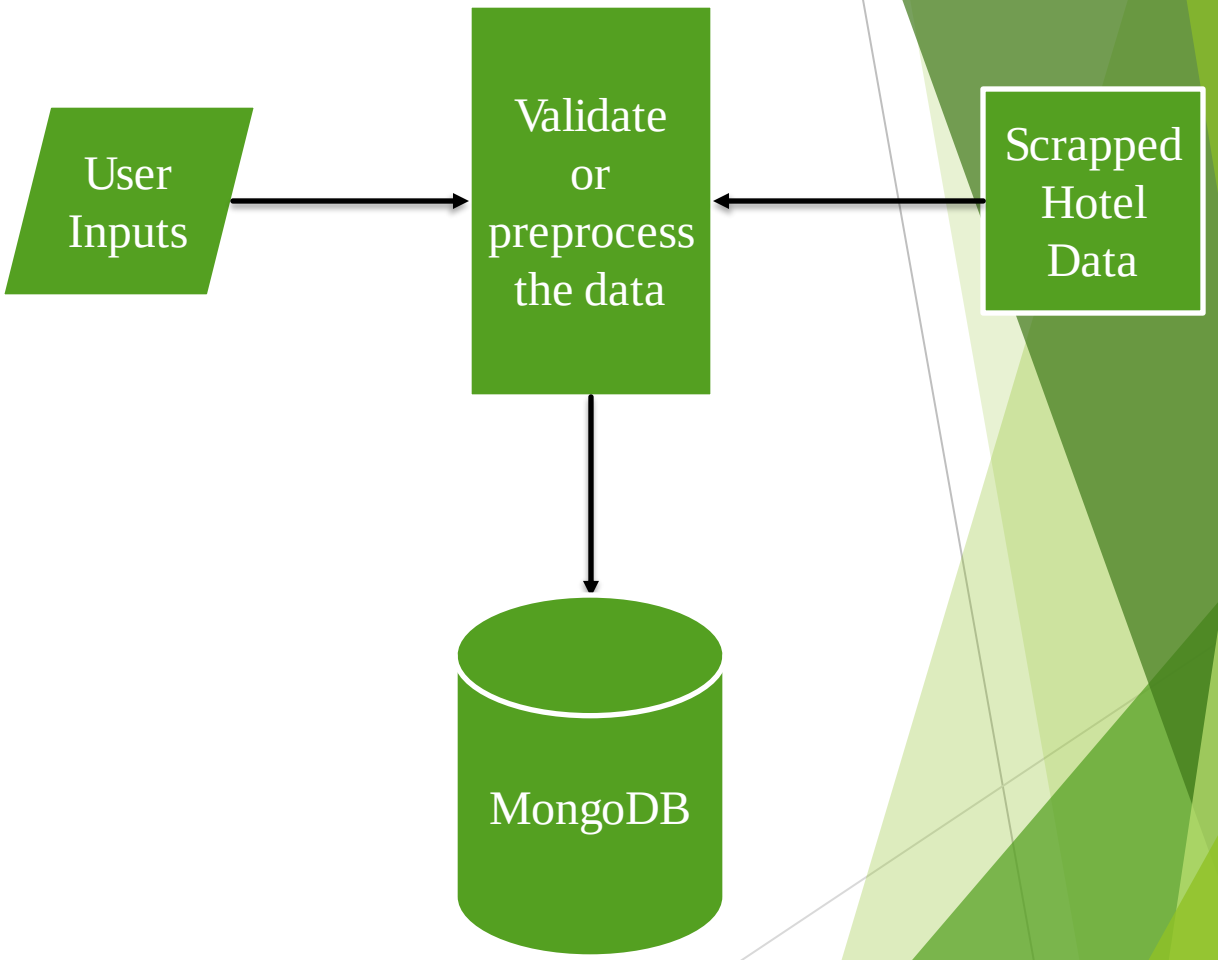
This functionality suggests possible suggestions based on the prefix strings entered by the user. It is based upon Trie tree data structure.



Module Insights

Data Validator:

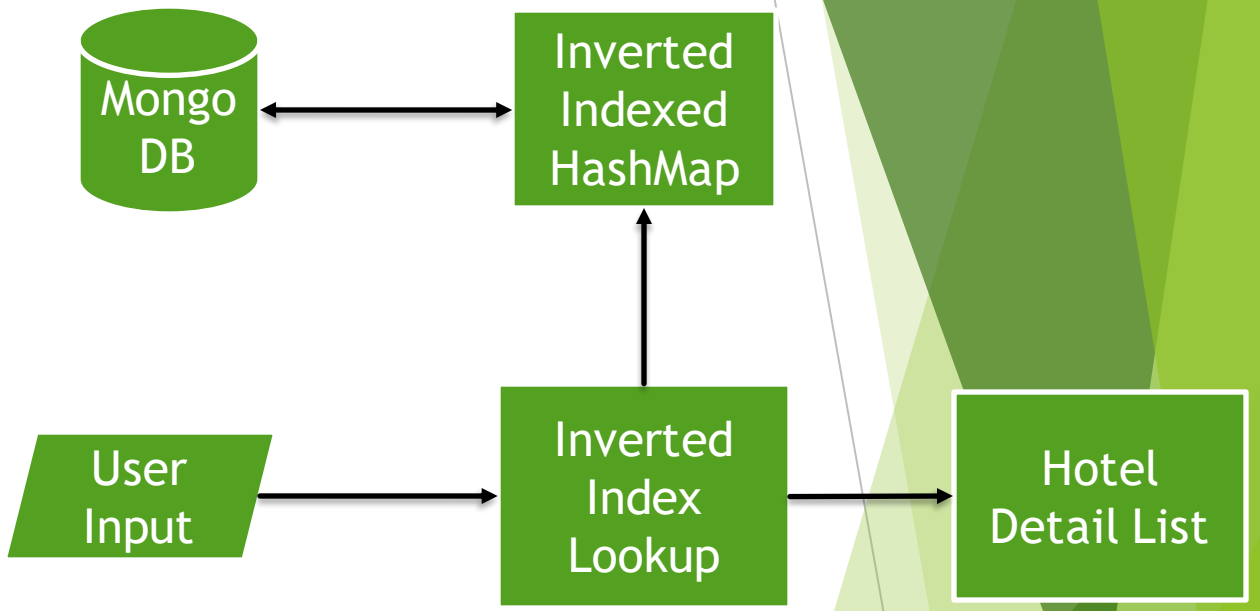
This functionality has been used to validate the characters provided by the user in the input prompts and also before inserting the hotel data into the database so that special or unwanted characters are excluded before insertion to data.



Module Insights

Inverted Indexing:

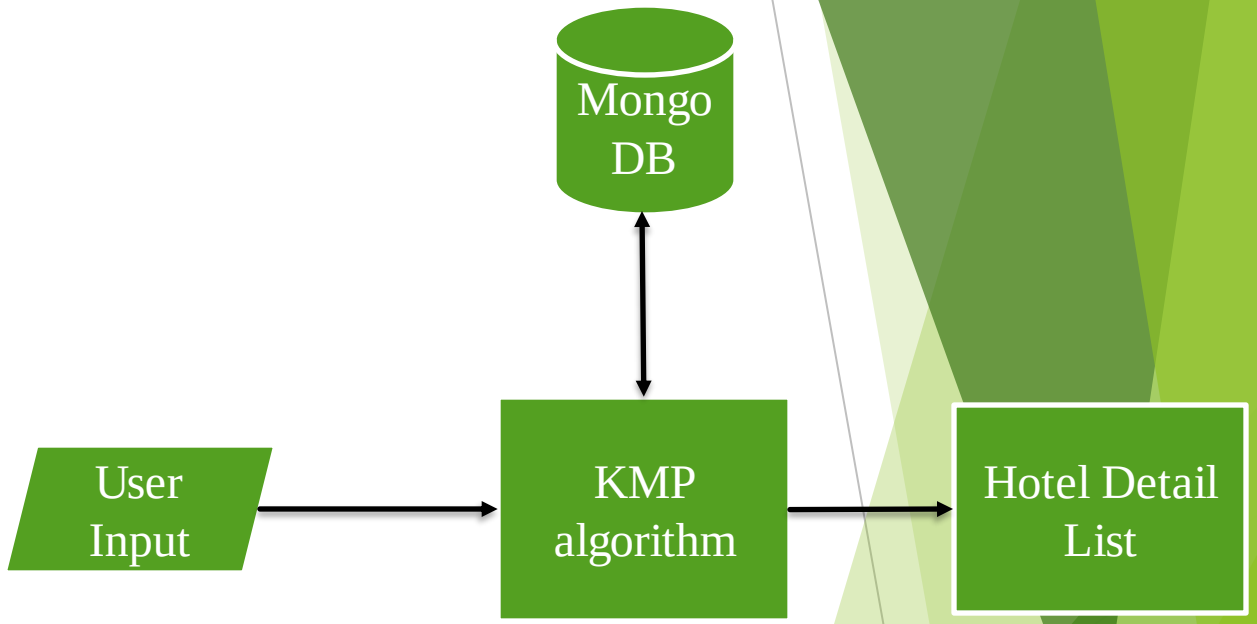
This functionality helps to optimise search within the database. It groups the data having common strings within them and assigns them a particular index so that if a search is performed it only has to scan the group instead of complete data.



Module Insights

Pattern Matching:

This functionality helps to search for specific pattern from hotel description (for e.g. some extra comments entered by the user for hotel facilities). It is based on **Knuth-Morris-Pratt** algorithm.



Module Insights

Page Ranking:

Assessing web page importance via algorithms analyzing content, links, user engagement for search results. Tasks involve web crawling, data processing, and implementing algorithms like PageRank, considering factors such as content quality, keywords, backlinks, and user behavior to rank pages within search engines.

Demo

Questionnaire

”

Thank you