Word Evaluator Research

VocabVersus

Thomas van der Molen

|  |  |
| --- | --- |
| **Project Information** | |
| Project members | Thomas van der Molen |
| Project Name | VocabVersus |

Table of Contents

[Topic 3](#_Toc131613585)

[Goal 3](#_Toc131613586)

[Summary 3](#_Toc131613587)

[Questions 4](#_Toc131613588)

[Word Set Storage 5](#_Toc131613589)

[Persistence 5](#_Toc131613590)

[Storage Options 5](#_Toc131613591)

[Relational Database 5](#_Toc131613592)

[Non-relational Database 5](#_Toc131613593)

[Word evaluation 5](#_Toc131613594)

[Database Query 5](#_Toc131613595)

[You know.. SQL 5](#_Toc131613596)

[Apache Lucene 5](#_Toc131613597)

[Cloud Search 5](#_Toc131613598)

[Evaluator Performance 5](#_Toc131613599)

[Storage 5](#_Toc131613600)

[Response Time 5](#_Toc131613601)

# Topic

Vocab Versus requires a service where **correct words can be stored** and later be checked against to **see if a word submitted by a player is correct**.

For this service, large amounts of data in the form of word lists (which contains all correct words for a given topic e.g. English, periodic table, etc.) have to be persistently stored to be re-used in games. Secondly, the service will have to be able to look through a given word list and find if a word submitted by a user is contained within the list.

This whole process will have to happen with many player submitions and return the result in real-time, as to give the player feedback on their submission.

# Goal

The goal of this research will be to explore the different methods of **storing large amounts of text** in the form of word lists, with the ability to (very) **quickly search** through the text and find a match for a given word.

# Summary

Shortly summarize the findings here

# Questions

There are multiple questions that will guide the research, these questions are based on the topics explained in the [introduction to this topic](#_Topic_1).

|  |  |
| --- | --- |
| **Main Question** | |
| What is the best method of storing and evaluating large amounts of grouped words to use for a game of Vocab Versus | |
| **Sub Questions** | **Research Methods** |
| What storage solutions exist for storing large amounts of textual data | [Literature study](https://ictresearchmethods.nl/Literature_study)  [Community research](https://ictresearchmethods.nl/Community_research) |
| What specialized methods exist for finding matching parts of text within a large amount of text | [Available product analysis](https://ictresearchmethods.nl/Available_product_analysis) |
| What combination of persistent storage and word matching technique gives the most optimal execution time | [Prototyping](https://ictresearchmethods.nl/Prototyping)  [Benchmark test](https://ictresearchmethods.nl/Benchmark_test) |

# Word Set Storage

## Persistence

Why persistent, need it for a while, and there will be a lot

## Storage Options

### Relational Database

Think MSSQL or MySQL

### Non-relational Database

This is many different things, key-value, graph, but most importantly Document

# Word evaluation

## Database Query

## You know.. SQL

## Apache Lucene

You know lucene, prob the best, EVEN allows for fuzzy evaluation which is nice :). Also ElasticSearch and Solr are based on it, so could look into that as sub options

## Cloud Search

Apparently azure search and amazon cloudsearch and Agolia are a thing

# Evaluator Performance

Explain that a couple of solutions will be built and compared via Postman response time to see which is the fastest

## Storage

How expensive are different options in terms of storage capacity required

## Response Time

How FAST are they