

# Final Presentation

## Spring 2017

# Vocabulary in Reading

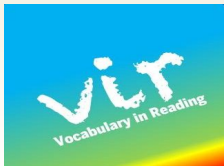
**Team Members: Camilo Rivera and Charles Benitez**

**Product Owner: Seyedjafar Ehsanzadehsorati**

**Mentors: Eric Dwyer and Mohsen Taheri**

**Instructor: Masoud Sadjadi**

**School of Computing and Information Sciences  
Florida International University**



# Project definition

- <http://vocabinreading.com>

Type your text

Enter Text

Process Text

Insert Image File

Insert PDF File

File: Process Text

File: Process Text

Number of words	
High frequency	0
Medium Frequency	0
Low Frequency	0
Academic Word List	0

Proceed to Enhanced Text

Vocabulary in Reading

VIR Dashboard Overview

Welcome to Vocabulary in Reading! Please upload your text below.

Upload text  
Copy and paste!

Upload Word  
What's up .doc?

Upload PDF  
Portable document format

OCR  
Optical character recognition

View Details

View Details

View Details

View Details

VIR Dashboard

Categories

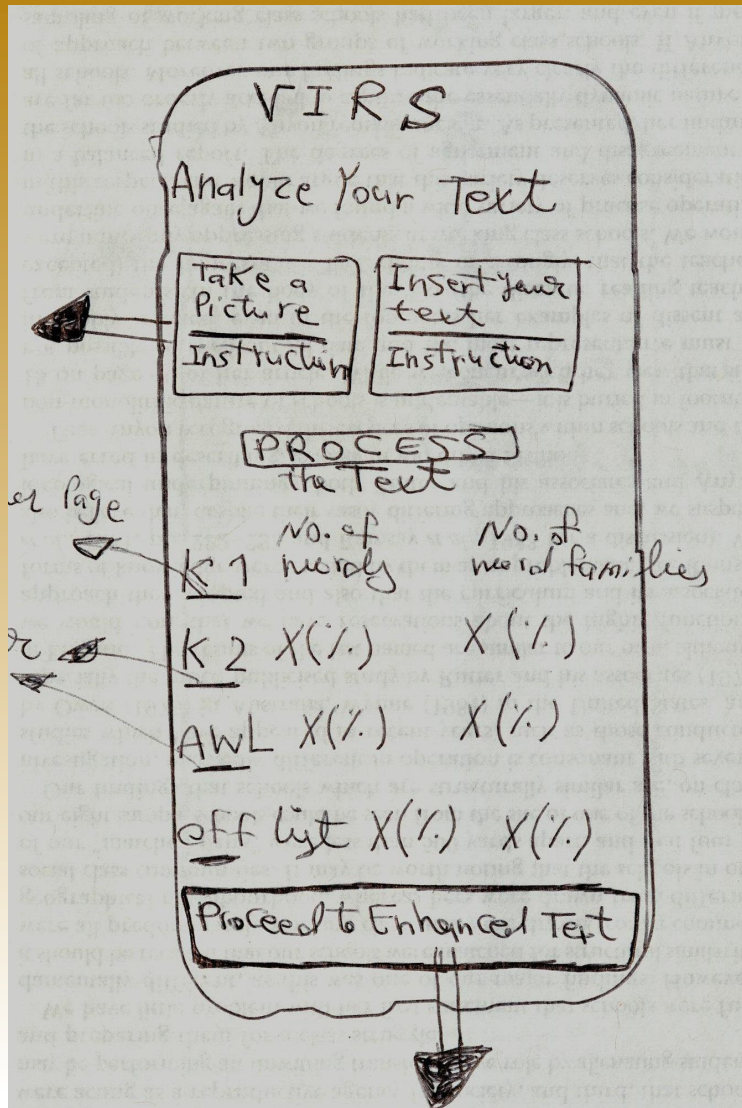
Dictionary

Frequency Chart

Edit Database

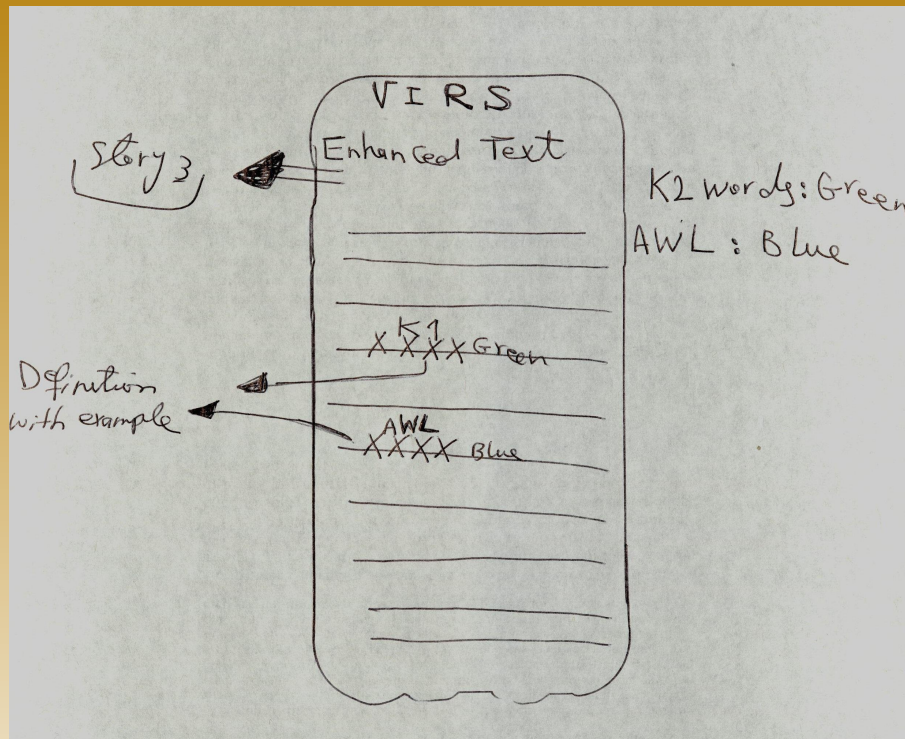
Backup/Import Database

# Requirements: Use Cases

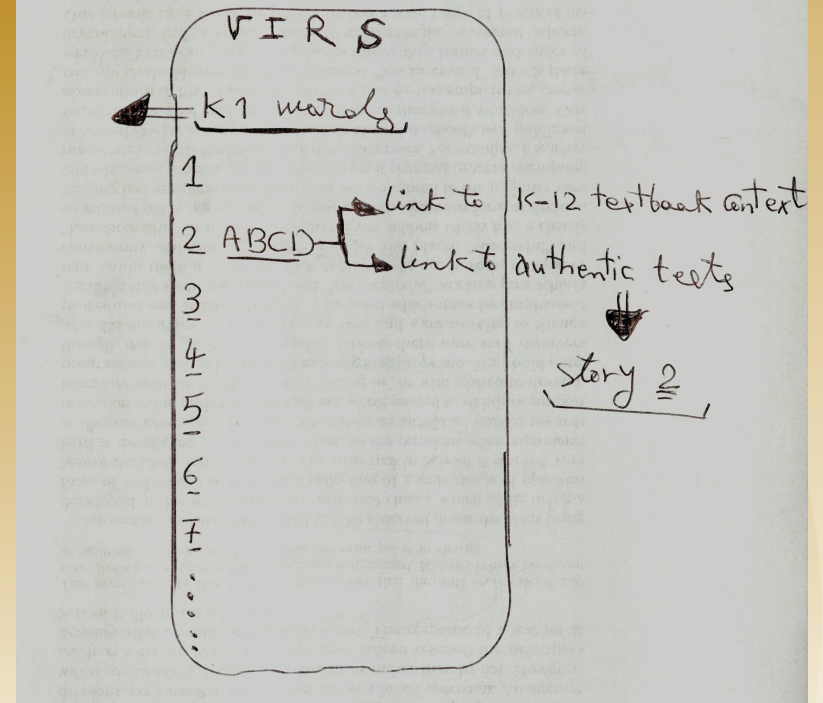


- The following figures are the original sketches provided to the team by Project Owner Seyedjafar Ehsanzadehsorati
- They illustrate the various use cases to be implemented in the then-hypothetical VIRS application.

# Requirements: Use Cases



We will have the same page for  
AWL - K2 - Offlist



# System Design: Architecture

- The VIR webapp utilizes an MVC (Model-View-Controller) architecture.
- The VIR Admin Panel uses a Client-Server architecture.
- Below is the deployment diagram for the system as a whole.



# Class Diagram



# User Stories

1. Words by Category
2. Dictionary
3. Upload/Scan PDF
4. Landing Page GUI
5. VIR Admin Page

# User Story #1 Words By Category

- Core of our project is the capability of categorizing text input by the user through pdf file or by directly typing in the text.
- We used a text-scan algorithm first scanning the words in the submitted text.
- Each word is stored as an element in the input array.
- We compare the words in the input array with the words in the database which are also in an array.



# Words By Category continued

```
// Scanning text and comparing with K1
for( var i = 0; i < o.words.length; i++ ){ // Array holding all high freq words
    for( var j = 0; j < words.length; j++ ){ // Array holding text inputted by user
        if(o.words[i].word === words[j]) {
            o.wordCount++;
            o.textWords.push({words: o.words[i].word, color: 'high'});
        }
    }
}
};
```

- Above is the algorithm in JavaScript(AngularJS) code
- If the word is found, it is pushed into an array of that specific type of word (in this case, High-Frequency aka K1).

# Words By Category continued

```
130 $awlc = $_POST['awlc'];
131 $loc = $_POST['loc'];
132 $med = $_POST['med'];
133 $hi = $_POST['hi'];
134
135 //print_r($med_array);
136
137 //Output loop : POST check marks act as binary flags for each category
138 foreach ($out_arr as $tok){
139
140     if(helper($tok, $awl_array) && $_POST['AWL']){echo "<a href=\"define.php?word=$tok\"><font color = \"\$awlc\">$tok </font>";}
141     else if(helper($tok, $hi_array) && $_POST['hi']){echo "<a href=\"define.php?word=$tok\"><font color = \"\$hi\">$tok </font>";}
142     else if(helper($tok, $med_array) && $_POST['med']){echo "<a href=\"define.php?word=$tok\"><font color = \"\$med\">$tok </font>";}
143     else if(helper($tok, $lo_array) && $_POST['lo']){echo "<a href=\"define.php?word=$tok\"><font color = \"\$lo\">$tok </font>";}
144     else{ echo "<a href=\"define.php?word=$tok\">$tok</a> ";}
145 }
146 }
147 ?>
```

```
<?php

function helper($needle, $haystack){
    foreach ( $haystack as $i ){
        //echo "Looking for $needle...";
        if( $i == $needle ) { return 1; }
    } return 0; }

//get input
ob_start();
echo $_POST['input'];
$out = ob_get_contents();
ob_end_clean();

//$out_arr = array();
$out_arr = preg_split('/[\s,.:?!+/,', $out);
//print_r($out_arr);

//get arrays

do { $awl_array[] = $row_awl['word'];
} while ($row_awl = mysql_fetch_assoc($awl));
```

- Above is the algorithm implementation written in PHP

# Words By Category continued

Vocabulary in Reading ✉ 🔔 👤 Mark Rosenberg

VIR Dashboard

Categories

Dictionary

Frequency Chart

Edit Database

Backup/Import Database

Output:

A secret always has a strengthening effect upon a newborn friendship as does the shared impression than an external figure is to blame the men of the Crown have become united less by their shared beliefs we observe than by their shared misgivings—which are in the main externally directed In their analyses variously made of Alastair Lauderback George Shepard Lydia Wells Francis Carver Anna Wetherell and Emery Staines the Crown men have become more and more suggestive despite the fact that nothing has been proven no body has been tried and no new information has come to light Their beliefs have become more fanciful their hypotheses less practical their counsel less germane Unconfirmed suspicion tends over time to become wilful fallacious and prey to the vicissitudes of mood—it acquires all the qualities of common superstition—and the men of the Crown Hotel whose nexus of allegiance is stitched after all in the bright thread of time and motion have like all men no immunity to influence

Input:

A secret always has a strengthening effect upon a newborn friendship, as does the shared impression than an external figure is to blame: the men of the Crown have become united less by their shared beliefs, we observe, than by their shared misgivings—which are, in the main, externally directed. In their analyses, variously made, of Alastair Lauderback, George Shepard, Lydia Wells, Francis Carver, Anna Wetherell, and Emery Staines, the Crown men have become more and more suggestive, despite the fact that nothing has been proven, no body has been tried, and no new information has come to light. Their beliefs have become more fanciful, their hypotheses less practical, their counsel less germane. Unconfirmed suspicion tends, over time, to become wilful, fallacious, and prey to the vicissitudes of mood—it acquires all the qualities of common superstition—and the men of the Crown Hotel, whose nexus of allegiance is stitched, after all, in the bright thread of time and motion, have, like all men, no immunity to influence.

☒ **AWL** Lime ▼

☒ **Low Frequency** Blue ▼

☒ **Medium Frequency** Red ▼

☒ **High Frequency** Gold ▼

- Output of algorithm written in PHP

# User Story #2 Dictionary

- We searched for an open source English dictionary:

<https://sourceforge.net/projects/mysqlenglishdictionary/>

- We imported it into our MySQL database.
- We then wrote SQL queries which returned either the definition of specific word, or all the definitions at once.
- Using SQL queries facilitates the retrieval of specific definitions without having to manually search both databases every time the page loads.

# Dictionary Continued

Vocabulary in Reading

VIR Dashboard

Categories

Dictionary

Frequency Chart

Edit Database

Backup/Import Database

## Category Viewer

Academic word list - Low - Medium - High

Showing words 1 to 50 of 1390 total words.

First page - Previous page - Next page - Last page

abandon  
abandonment  
abnormal  
abnormally  
abstract  
abstraction  
abstractly  
academia

Untitled Document - Google Chrome

localhost/admin/define.php?word=abstraction

Definition for word **Abstraction** a.

"The act of abstracting, separating, or withdrawing, or the state of being withdrawn; withdrawal."

[Go back](#) [Close](#)

accommodation

Vocabulary in Reading

VIR Dashboard

Categories

Dictionary

Frequency Chart

Edit Database

Backup/Import Database

Showing words 101 to 200 of 176023 total words.

Define First - Prev - Next - Last

Word	Part of speech	Definition
Abate	v. t.	To bring down or reduce from a higher to a lower state, number, or degree; to lessen; to diminish; to contract; to moderate; to cut short; as, to abate a demand; to abate pride, zeal, hope.
Abate	v. t.	To deduct; to omit; as, to abate something from a price.
Abate	v. t.	To blunt.
Abate	v. t.	To reduce in estimation; to deprive.
Abate	v. t.	To bring entirely down or put an end to; to do away with; as, to abate a nuisance, to abate a writ.
Abate	v. t.	To diminish; to reduce. Legacies are liable to be abated entirely or in proportion, upon a deficiency of assets.
Abate	v. t.	To decrease, or become less in strength or violence; as, pain abates, a storm abates.
Abate	v. t.	To be defeated, or come to naught; to fall through; to fail; as, a writ abates.
Abate	n.	Abatement.
Abatement	n.	The act of abating, or the state of being abated; a lessening, diminution, or reduction; removal or putting an end to; as, the abatement of a nuisance is the suppression thereof.
Abatement	n.	The amount abated; that which is taken away by way of reduction; deduction; decrease; a rebate or discount allowed.
Abatement	n.	A mark of dishonor on an escutcheon.
Abatement	n.	The entry of a stranger, without right, into a freehold after the death of the last possessor, before the heir or devisee.
Abater	n.	One who, or that which, abates.
Abatis	n.	Alt. of Abatis
Abattis	n.	A means of defense formed by felled trees, the ends of whose branches are

- Displayed above are “define.php” which defines specific words and “dict.php” which provides all stored definitions.

# Dictionary Continued

```
router.get('/dictionary/:word', (req, res, next)=>{
  console.log("Word in route is: " + req.params.word);
  var theWord = req.params.word;
  var queryAwl = 'SELECT * FROM dict where word = "' + theWord + '"';

  connection.query(queryAwl, (err, results, field)=>{
    if(err) console.log(err);

    var tempJSON = {};

    results.forEach((item, index) => {
      tempJSON = {
        "id" : results[index].id,
        "word" : results[index].word,
        "wordtype": results[index].wordtype,
        "definition": results[index].definition
      };
    });
  });
});
```

- Algorithm Above written in JavaScript(ExpressJS) with a MySQL query.



# User Story #3 Upload/Scan PDF

- We implemented a system for selecting pdf files and scanning and extracting the text from them in order to process and analyze the text.
- We used the ng-file-upload (<https://github.com/danialfarid/ng-file-upload>) library for selecting and submitting a file and the pdf.js open source library for reading pdf files (<https://mozilla.github.io/pdf.js/>)
- We implemented algorithm using these libraries for extracting the text and used our same scan algorithm for categorizing it.



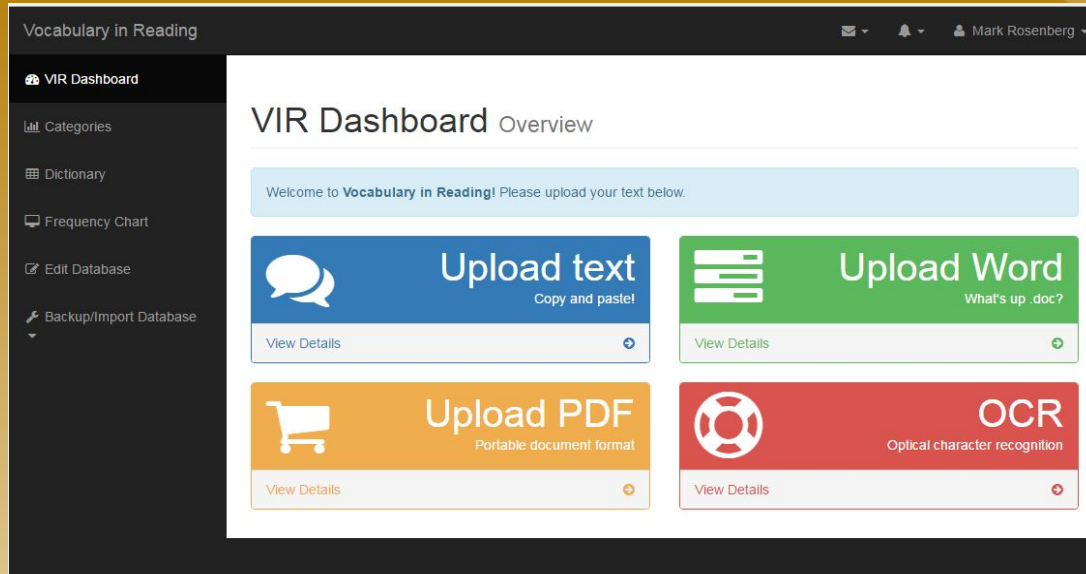
# User Story #4 Landing Page GUI

- For the Web App, we wanted a look and feel that would be very convenient and easy for mobile phone users.
- We went with big buttons and simple vertical layout.
- We also wanted to reach the widest audience possible so we went with a site.

The interface is titled "Type your text" and includes a text input field labeled "Enter Text" and a "Process Text" button. Below this, there are two columns for file insertion: "Insert Image File" with a JPG icon and "Insert PDF File" with a PDF icon. Each column has a "File:" label and a "Process Text" button. At the bottom, there is a table titled "Number of words" with four rows: "High frequency", "Medium Frequency", "Low Frequency", and "Academic Word List", each with a corresponding "0" value. A "Proceed to Enhanced Text" button is located at the very bottom.

Number of words	
High frequency	0
Medium Frequency	0
Low Frequency	0
Academic Word List	0

# User Story #5 VIR Admin Panel



- The VIR Admin Panel allows for automated management of the backend of the VIR app.
- The site mirrors the functionality of the application as a website.
- It allows for the database to be exported and manipulated on the fly

# User Story #5 VIR Admin Panel

The screenshot shows the 'Vocabulary in Reading' admin panel. The left sidebar contains navigation links: VIR Dashboard, Categories, Dictionary, Frequency Chart, Edit Database, and Backup/Import Database (selected). The main content area is titled 'Backup/Import Database Import Data'. It features a message about uploaded files, a warning about data truncation, and a preview of the AWL table data.

Vocabulary in Reading

VIR Dashboard

Categories

Dictionary

Frequency Chart

Edit Database

Backup/Import Database

Import Data

Export Data

Backup/Import Database Import Data

The following files have been uploaded to the server. If no files are shown, the files were not accepted by the system. In that case, please [try again](#).

**WARNING:** It is **YOUR** responsibility to check that the files are correct before attempting to update the database, including checking if the columns are correct. The previews **WILL** be the data going into the database. Good luck.

Hint: Selecting **TRUNCATE** will **delete all existing data in the database** for that specific list before adding the uploaded data. Not selecting it will append the uploaded data to the end of the existing data.

AWL File Uploaded : AWL.csv  
TRUNCATE before updating AWL ? ☒

Previewing AWL table . . .

test0
test1
test2
test3
test4
test5
test6
test7
test8
test9

Update SQL Database

Back

The screenshot shows the 'Vocabulary in Reading' admin panel. The left sidebar contains navigation links: VIR Dashboard, Categories, Dictionary, Frequency Chart, Edit Database, and Backup/Import Database (selected). The main content area is titled 'Backup/Import Database Export Data'. It features a message about selecting databases to export and a list of available databases.

Vocabulary in Reading

VIR Dashboard

Categories

Dictionary

Frequency Chart

Edit Database

Backup/Import Database

Import Data

Export Data

Backup/Import Database Export Data

Please select the databases you wish to export.

AWL

Low Frequency

Medium Frequency

High Frequency

- SQL queries generated automatically based on provided Excel data

# Test Suites and Test Cases

- Test case 1: Test database to make sure words are being stored correctly and reliably: PASS
- Test case 2: Categorizer algorithms correctly detect and prevent garbage input: PASS
- Test case 3: Categorizer algorithms correctly strip input of invalid symbols and punctuation: PASS
- Test case 4: All words are correctly categorized and color coded according to specifications: PASS

# Summary

- This has been our presentation guiding you through the most important parts of our senior project entitled Vocabulary in Reading. We hope you all enjoyed it!
- Contact Information
  - Camilo Rivera: [crive150@fiu.edu](mailto:crive150@fiu.edu)
  - Charles Benitez: [cbeni024@fiu.edu](mailto:cbeni024@fiu.edu)
- Questions?
- Thank You!

