

Documentation of AACR Lexical Tools

Created by Marisol Mercado Santiago, Ph.D.

Automated Analysis of Constructed Response Research Group at CREATE for STEM Institute

Principal Investigator: Kevin Haudek, Ph.D.

June 2020

Version 1

(Note: diagrams created with online tool <https://app.diagrams.net/> and Visual Studio)

Acknowledgements:

This material is based upon work supported by the National Science Foundation (Grants 1323162 and 1347740). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the supporting agencies.

Table of Contents

Introduction	2
Installation Instructions (Windows only)	3
File formatting and input requirements	5
How to make it work	6
Understanding the output	9
Class Diagram	10
Sequence Diagram (Part I)	11
Sequence Diagram (Part II)	12
Use Case Diagram	13

Introduction

AACR Lexical Tools was developed to support the AACR research group's needs to examine how balanced are scores assigned to categories, in addition to exploring if lexical diversity measurements may inform the development of machine learning models of student responses. It was programmed in C# by Marisol Mercado Santiago, a member of the AACR research group, CREATE for STEM Institute, Michigan State University.

The code is open to others to contribute or learn from it.

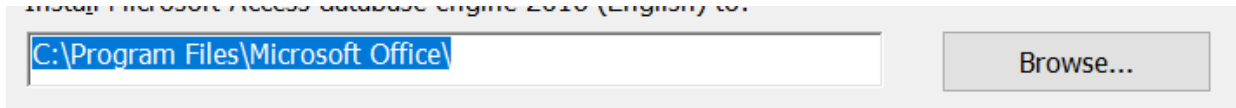
Note: We are not responsible for the software outputs or accuracy of the calculations.

Installation Instructions (Windows only)

(1) Please download and install the Microsoft Access Database Engine 2016 Redistributable:

<https://www.microsoft.com/en-us/download/details.aspx?id=54920>

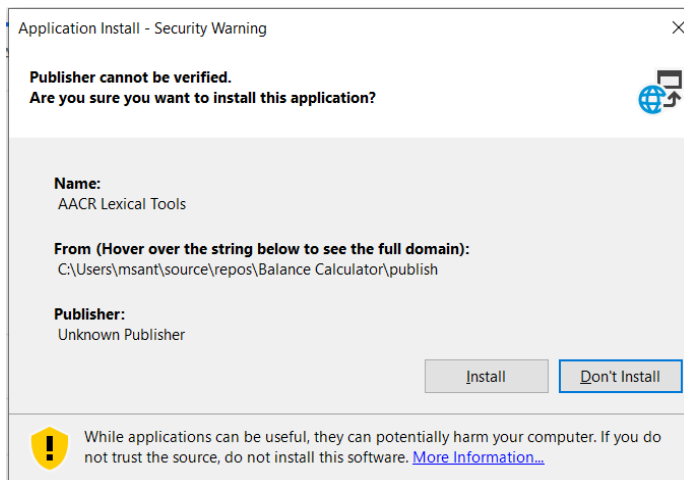
Install it on your default Program Files directory:



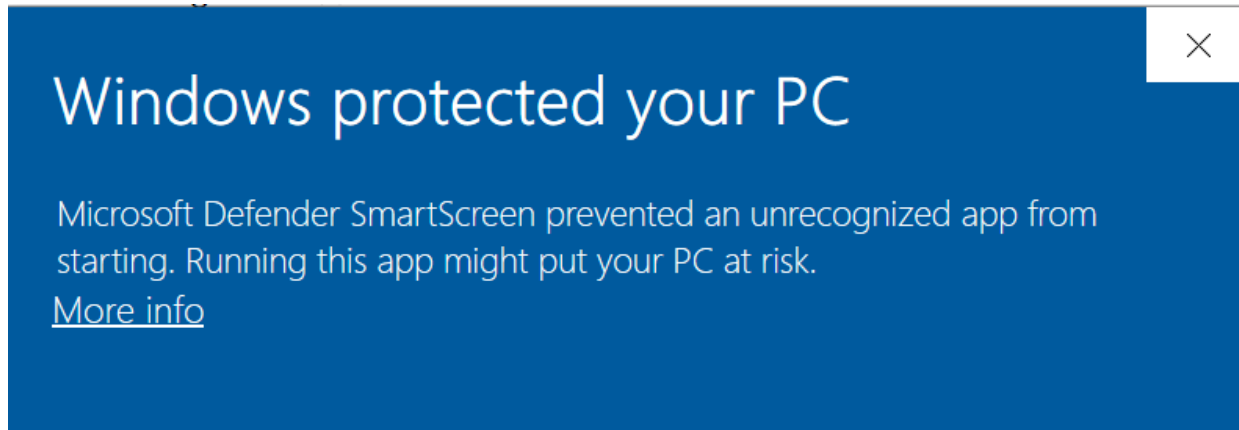
Install it for 64-bit if your computer is a 64-bit system. If not, install the one for a 32-bit system.

(2) Download the zip file “AACR Lexical Tools 1.2”. Extract the zip file and run “setup”:

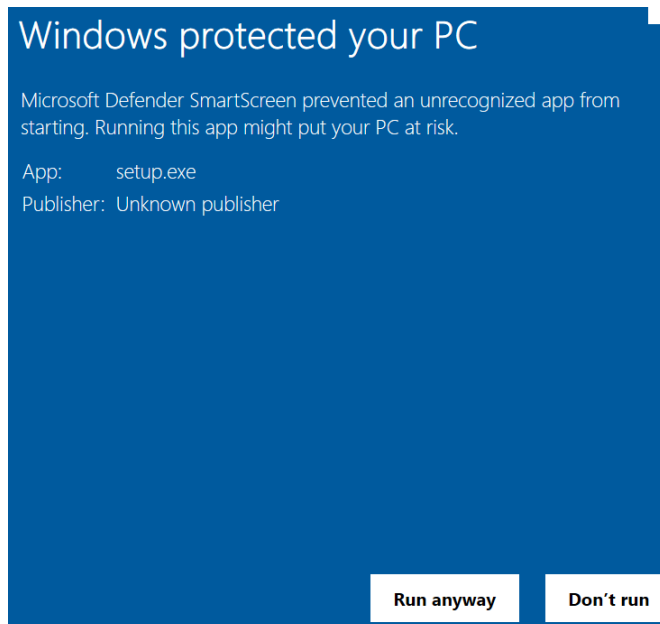
(3) Click on **Install**:



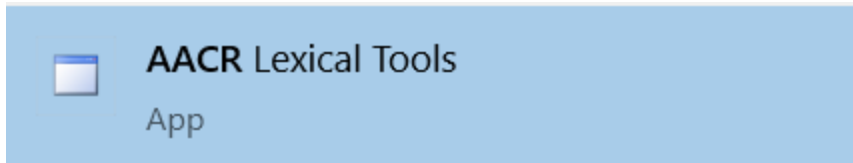
NOTE: If Windows show you this message, please click on “More info”



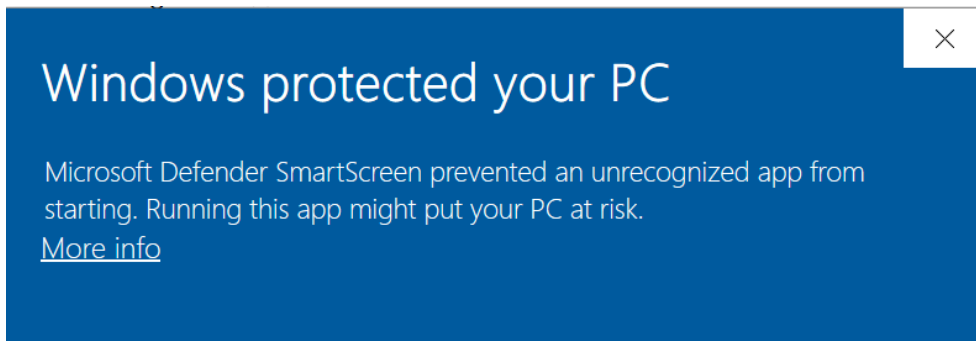
...then click on “Run anyway”:



(4) The app will be installed in your Windows computer.



(5) On the first run, Windows may show you the “defender” message again. Please, click on “More info” and then on “Run anyway”:

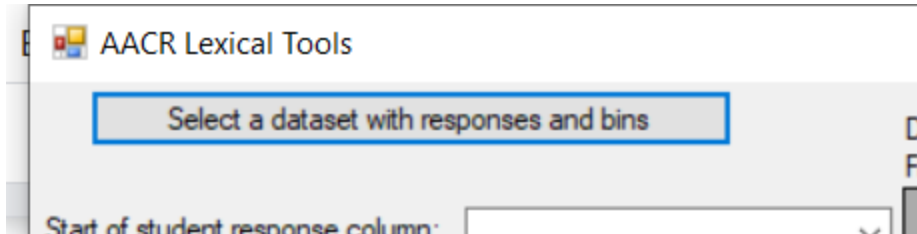


File formatting and input requirements:

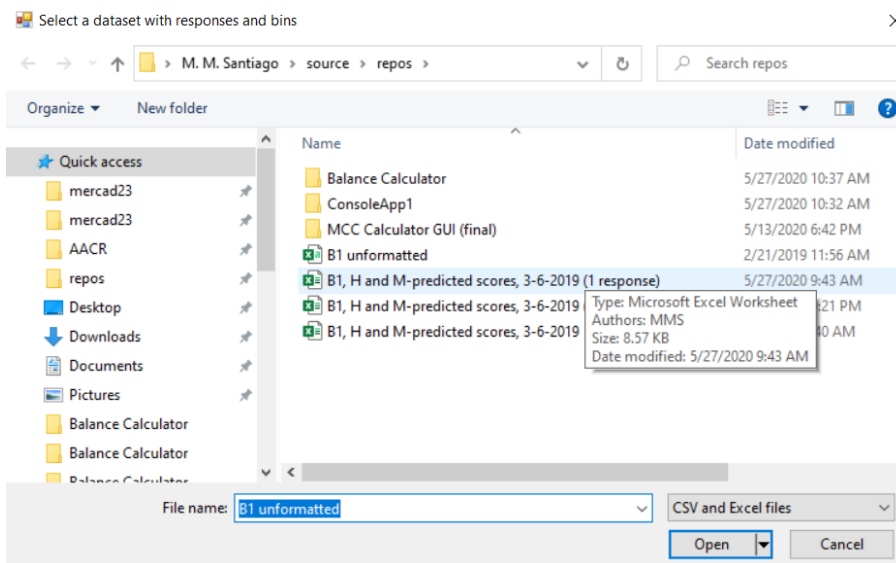
- It only accepts **Excel** and **CSV files**.
- **Files must have column headers.**
- It will ignore all-empty rows (a row without content), but it won't calculate values correctly if you have a mix of empty scores across rows and cells. (You will need to input a properly-formatted file.)
- If your Excel file has multiple sheets, it **will only read and load the first sheet**.

How to make it work:

(1) Click on the button “**Select the dataset with responses and bins**”:



(2) A file explorer window will appear. Please navigate to where your file is located, click on it, then click on “**Open**”:



(3) You'll see a **preview** of your data at the right:

AACR Lexical Tools

Select a dataset with responses and bins

Data loaded. Please select response and score columns below.

Start of student response column:

End of student response column:

Start of score columns:

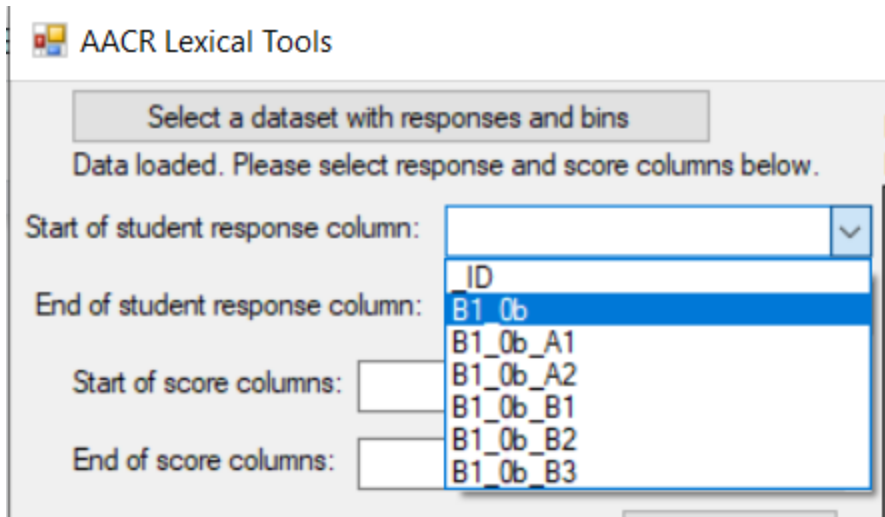
End of score columns:

Calculate

Data preview
File: C:\Users\msant\source\repos\B1 unformatted.csv

_ID	B1_0b	B1_0b_A1	B1_0b_A2	B1_0b_A3
The hand dry...	1	1	0	
The claim th...	1	1	1	
That bathroo...	0	0	0	
The automati...	1	1	0	
The student ...	1	1	0	
The student ...	1	1	0	
Public bathr...	1	1	0	
x				
x				
The claim is...	1	0	1	
The student'...	1	1	0	
x				
x				
there is bac...	1	0	1	
The claim is...	1	1	0	
The student ...	1	1	0	
x				
x				
Dryers in pu...	1	1	0	
The claim is...	0	0	0	

(4) Select the appropriate column(s) with student responses by using the first two drop-down menus:



AACR Lexical Tools

Select a dataset with responses and bins

Data loaded. Please select response and score columns below.

Start of student response column:

End of student response column:

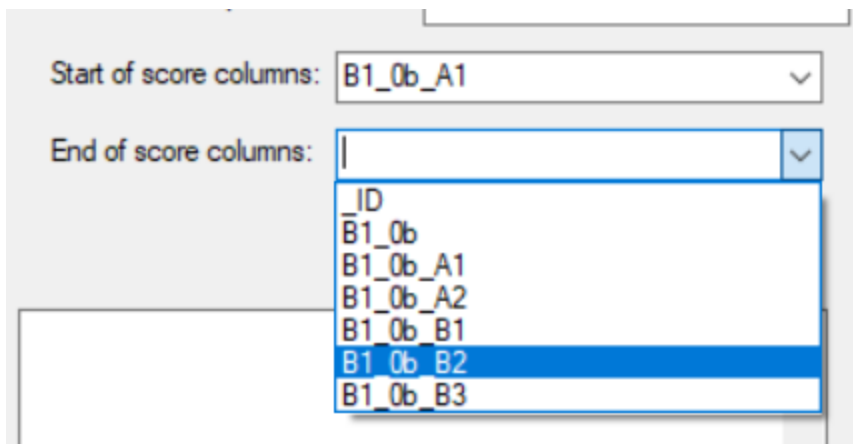
Start of score columns:

End of score columns:

Dropdown menu options: ID, B1_0b, B1_0b_A1, B1_0b_A2, B1_0b_B1, B1_0b_B2, B1_0b_B3

(The app can concatenate student responses across multiple columns if needed.)

(5) Choose the score bins using the third and fourth drop-down menus. These will be the bins that the app will use to calculate the balance of each one. The app can process one or multiple bins, holistic or analytic.



Start of score columns:

End of score columns:

Dropdown menu options: ID, B1_0b, B1_0b_A1, B1_0b_A2, B1_0b_B1, B1_0b_B2, B1_0b_B3

(6) Once you selected your response and score columns, please click on “Calculate”

Understanding the output:

The first text box will display the calculated Balance measure for each score column selected in the 3rd and 4th drop-down menus.

It will also show the following lexical diversity measurements: Type-to-token ratio, Brunet's Index, Guiraud's R, Honore's Statistic, Herdan's C, Carroll's Corrected TTR, Somers' S, Dugast's Uber Index, Yule's K, and vocd-D.

The second text box where it says "--Information of..." will display

- k (number of levels),
- n (number of responses), and
- c[i] (counts for each level) of each bin.

These are variables which are used in the [Balance formula shown here](#).

The third text box shows distinct stemmed words and the count of each word (how many times each one appears in the whole dataset).

AACR Lexical Tools

Select a dataset with responses and bins

Data loaded. Please select response and score columns below.

Start of student response column: B1_0b

End of student response column: B1_0b

Start of score columns: B1_0b_A1

End of score columns: B1_0b_B1

Calculate

Balance of B1_0b_A1 bin: 0.76216269391232
Balance of B1_0b_A2 bin: 0.991284030972042
Balance of B1_0b_B1 bin: 1.03658076272128

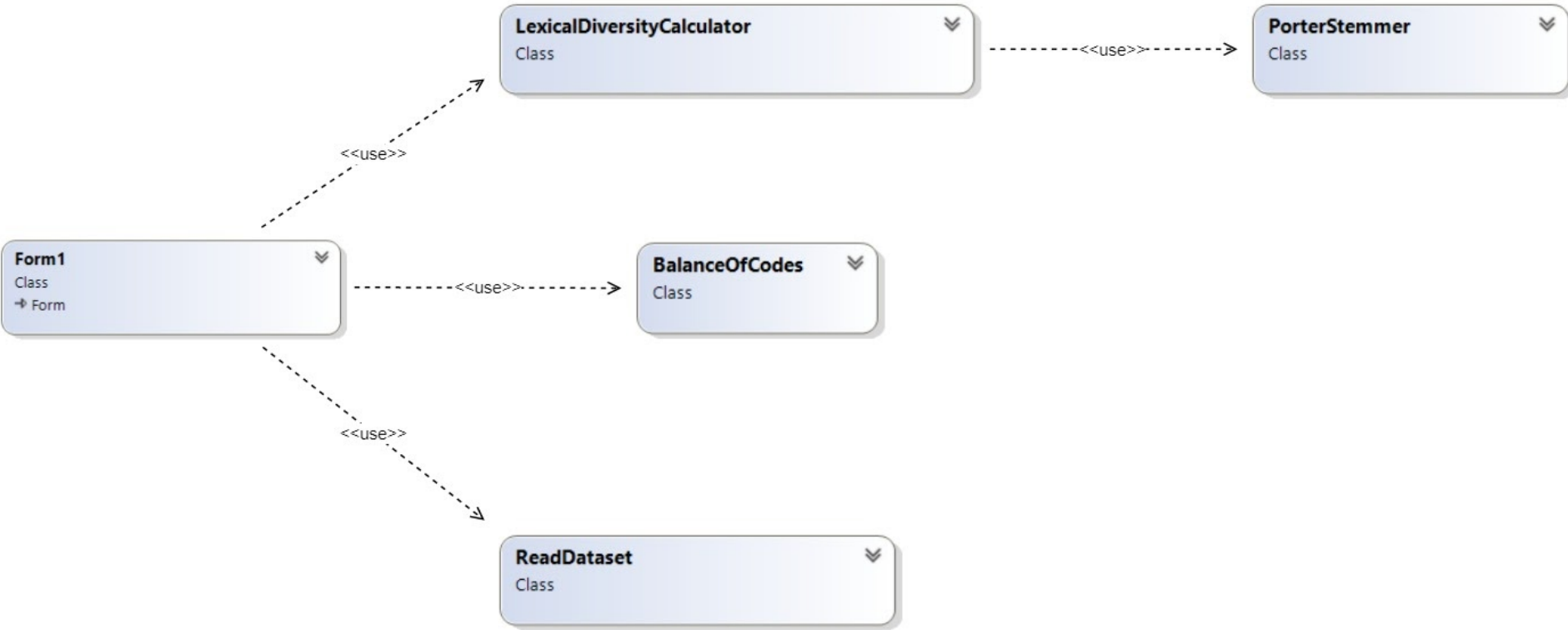
Type-to-token ratio: 0.063
Brunet's Index: 26.077
Guiraud's R: 4.9759
Honore's Statistic: 1683.4299
Herdan's C: 0.6836
Carroll's Corrected TTR: 3.5185
Somers' S: 0.8245

--Information of B1_0b_A1 bin: --
k: 2
n: 891
counts of 0: 68
counts of 1: 483

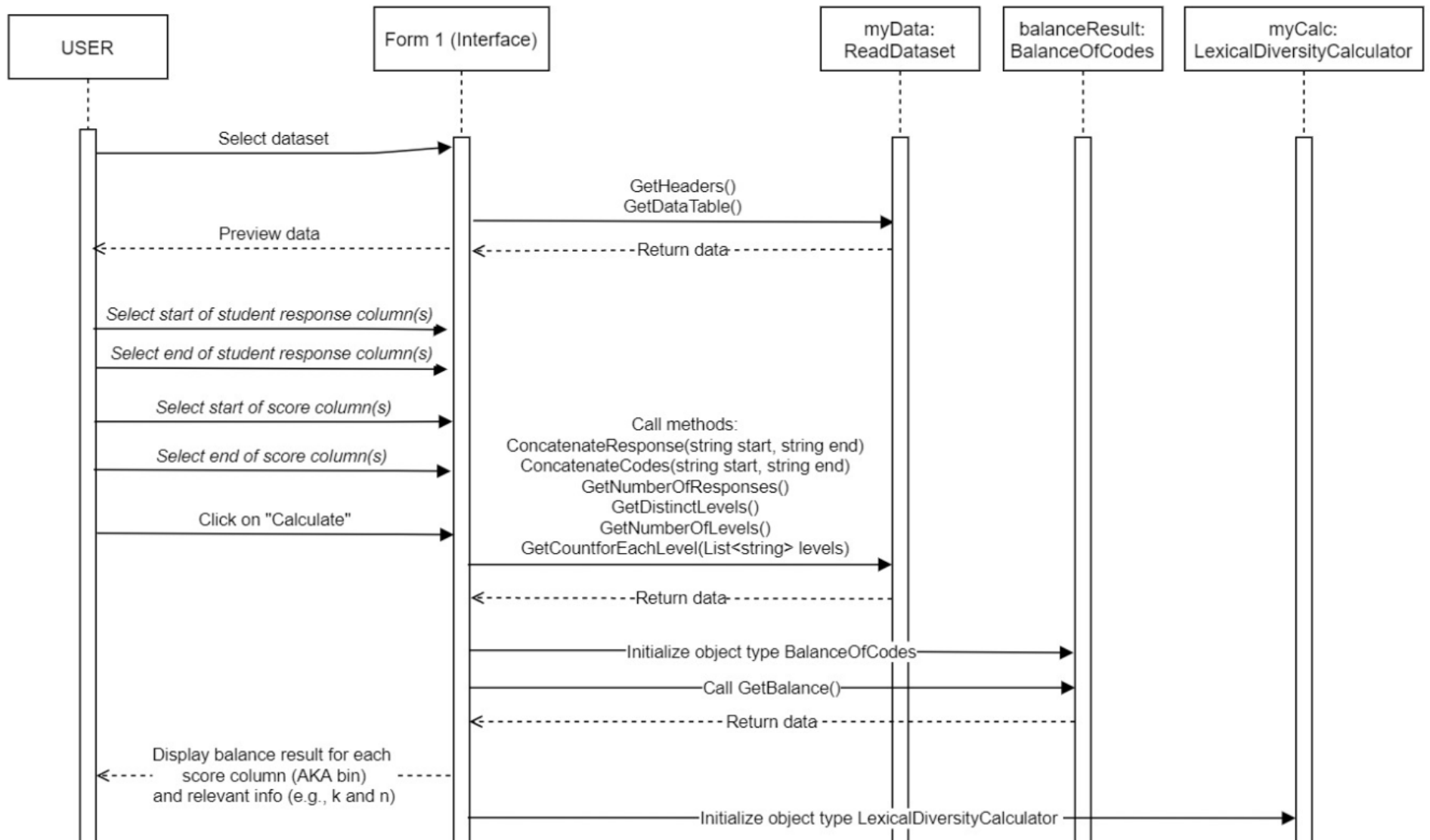
--Information of B1_0b_A2 bin: --
k: 2
n: 891
counts of 0: 179
counts of 1: 372

--Distinct stemmed words:
hand - 471
dryer - 454
that - 413
are - 376
x - 305
is - 298
danger - 292
automat - 223
in - 217

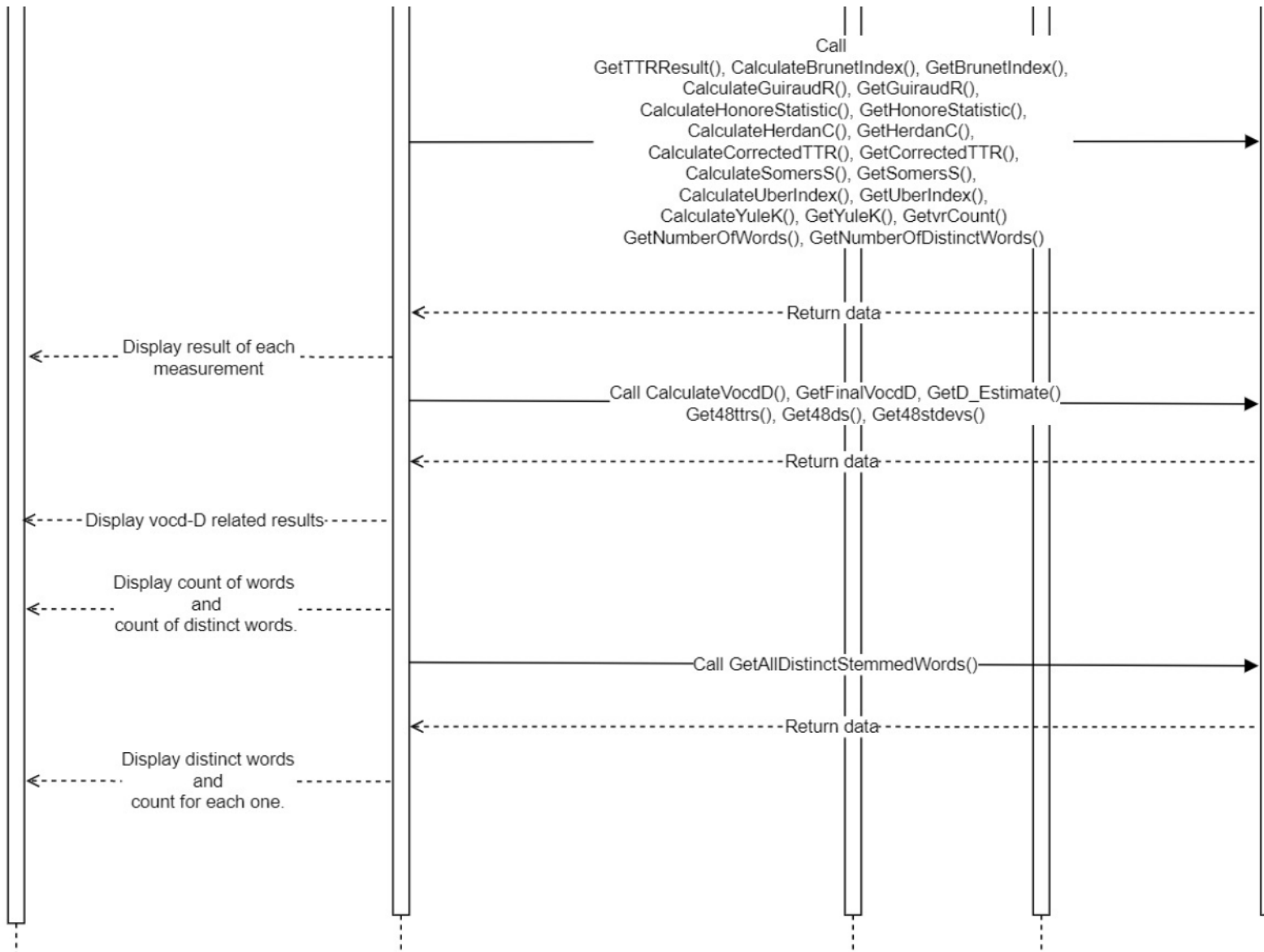
Class Diagram



Sequence Diagram (Part I)



Sequence Diagram (Part II)



Use Case Diagram



Acknowledgements:

This material is based upon work supported by the National Science Foundation (Grants 1323162 and 1347740). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the supporting agencies.