

ChemBox Project Documentation

Tom Schneider

Date: March 24, 2024

Center Number: 29065

Candidate Number: 7638

Ellesmere College

Contents

1	Analysis	2
1.1	Introduction	2
1.2	Prospective Users	3
1.3	Specific Objectives	3
1.3.1	Chemical Equations and Calculations - ChemCalculator	3
1.3.2	Chemical Equation Balancer - ChemBalancer	6
1.3.3	Visualisation of Chemical Molecules - ChemEditor . . .	6
1.4	Current and Proposed Systems	7
2	Documented Design	8
2.1	Introduction to the Documented Design	8
2.2	Project Hierarchy	8
2.3	Structure of the GUI	9
2.4	Algorithm Design for ChemCalculator	12
2.5	Algorithm Design for ChemBalancer	13
3	Testing	28
4	Evaluation	28

1 Analysis

1.1 Introduction

Technology gives us the benefits of saving time and doing work more efficiently. The use of software and technology in chemistry does not only help increase accuracy and decrease human error, but also reduces the time spent performing repetitive tasks by hand. ChemBox is a software project with the aim of creating an interactive, user-friendly and intuitive toolbox for automating and simplifying complex and repetitive tasks that come up on a daily basis for students, educators and professionals in the field of chemistry. The application features a range of different tools that should help chemists work more efficiently and also carry out their work more accurately. ChemBox is split into three distinct modules with different functionalities.

The first module is the "ChemCalculator". The aim of this part of the program is to help the user carry out calculations by filling in equations and formulae. Although substituting numbers into predefined equations is a rather trivial task, it leaves a lot of room for human error when it comes to things like converting between units or applying mathematical laws correctly. This module should help the user with the most important and at the same time most trivial tasks in chemistry.

The second module, "ChemBalancer", is for balancing chemical equations, as the name already suggests. Chemical equations come up in every lab experiment, calculation or research problem. While balancing short equations made up of very few different elements is arguably a rather easy task, it can get quite tricky when you have to work with a large number of different elements, complex ions or just very long equations. Making just a tiny mistake when balancing an important equation can cause a big set back as it can take long to find small errors like mixing up a 2 with a 3.

The third and last module is the "ChemEditor". Visualising molecular structures can play a vital role in understanding a substances chemical properties or understanding interactions with other substances. Drawing molecules out by hand is pretty straight forward. It is knowing when a bond is valid and which atoms bond together and which don't that is the tricky part. Having a tool that can help you make sure the chemical molecule you want to draw can even exist, can be a great help not only for beginner level chemists but also for more experienced chemists.

1.2 Prospective Users

Chembox will provide valuable tools to a diverse user base, spanning from students to professional chemists. The intuitive and straight forward design will allow users with varying backgrounds and degrees to use ChemBox for their own specific needs.

In the early stage, the main users of this system will be pupils and staff attending Ellesmere College, but it could be a goal to make the software available open source to anyone online.

Engaging with pupils at the college during the early stages allows for a valuable user feedback loop. This direct interaction with the user group will provide insights into the software's usability, identify potential improvements, and address any specific requirements that may arise within the college context.

1.3 Specific Objectives

Through being an A-Level Chemistry student myself, I have learned a lot about using chemical equations and performing calculations as well as balancing chemical equations and visualising chemical substances and molecules. I was able to identify a number calculations that processes that come up on a regular basis and divide them into non-negotiable and nice-to-have objectives.

1.3.1 Chemical Equations and Calculations - ChemCalculator

The first module of the program is for performing calculations which are based on chemical formulae. Where appropriate the program should allow the user to choose from a range of different units for each calculation, so the user doesn't have to calculate the conversions like the one from cm^3 to dm^3 for example.

Required functionalities:

1. Standard moles calculation:

$$\text{moles} = \frac{\text{mass}}{\text{molar mass}}$$

2. Calculation to find the concentration:

$$\text{concentration} = \frac{\text{moles}}{\text{volume}}$$

3. Avogadro's number calculations. The user should be able to give a number of different inputs, including mass, moles, molecular weight and the number of atoms. After giving two independent inputs, the program should be able to calculate the rest of the values using Avogadro's number.

The equation the calculator will be based on is:

$$\text{number of atoms} = \text{Avogadro's number} \times \text{moles}$$

This should be paired with a mole calculator for the possibility use the following formula:

$$\text{number of atoms} = \text{Avogadro's number} \times \frac{\text{mass}}{\text{molar mass}}$$

4. Atom Economy calculation:

$$\text{Atom Economy} = \frac{\text{Mr of desired product}}{\text{Sum of Mr of all reactants}} \times 100$$

5. Percentage Yield calculation:

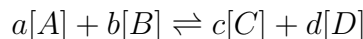
$$\% \text{Yield} = \frac{\text{Actual yield}}{\text{Theoretical yield}} \times 100$$

6. Calculation for the Specific Heat Capacity and Enthalpy changes:

$$q = mc\Delta T$$

(q = energy change) (m = mass) (c = specific heat capacity)
(ΔT = temperature change)

7. Equilibrium Constant calculation for a reversible reaction:



$$K_C = \frac{[C]^c [D]^d}{[A]^a [B]^b}$$

(K_C = Equilibrium Constant) (Upper case letter = Concentration) (Lower case letter = Moles in Equation)

8. Rate Equation and Rate constant calculation:

$$\text{Rate} = k[A]^m[B]^n$$

Non-essential objectives:

1. Gibbs Free Energy calculation:

$$\Delta G = \Delta H - T\Delta S$$

(ΔG = Gibbs Free Energy) (ΔH = enthalpy)
(T = temperature) (ΔS = entropy)

2. Acid calculations - pH and $[H^+]$

$$pH = -\log[H^+]$$

$$[H^+] = 10^{-pH}$$

3. Acid dissociation constants K_a and pK_a

$$K_a = \frac{[H^+][A^-]}{[HA]}$$

$$pK_a = -\log K_a$$

$$K_a = 10^{-pK_a}$$

1.3.2 Chemical Equation Balancer - ChemBalancer

A substantial part of the project will be the ChemBalancer which will be the module that balances chemical equations. This system must be able to take complex unbalanced equations and convert them into a balanced version. It must be able to handle subscript numbers, brackets and complex ions.

1.3.3 Visualisation of Chemical Molecules - ChemEditor

The third part of the program will be the ChemEditor module, which can be used for visualising the structures of chemical molecules. This module will require a user-friendly and easy to use interface, with the main focus on the canvas. The user should have the option to choose from a range of different elements what he wants to add to the canvas. In a tool bar, the user should also be able to choose the bond order (single, double, triple) and the charge on each atom. When clicking on an atom, there should be an option to add a bond to another atom or delete the atom.

When the user constructs their molecules, ChemEditor will have to conduct real-time checks to ensure that atoms do not exceed their valence electrons and that it is chemically possible to have a molecule with the given structure. The required objectives for this module are:

1. Tool bar:

In the tool bar on the top end of the application, there has to be a list of buttons for choosing the element, which must include the most common elements (Carbon, Hydrogen, Sulphur, Chlorine, etc.). There also has to be the option to choose the bond order (single, double or triple bond) as well as choosing the option to form a dative bond. Another essential option in form of buttons should be removing atoms and bonds as well as being able to save the drawn structures as a document. A possible non-essential enhancement would be getting extra information about atoms upon highlighting as well as getting information like the molar mass and the empirical formula of a molecule after highlighting.

2. Canvas:

The canvas is the area in which the user can draw their molecules. There are a number of essential features that must be included here.

- (a) The user must be able to draw atoms by clicking on the canvas.

- (b) Upon selecting an existing atom on the canvas, depending on the chosen action type, the user should have different options:
 - i. When the chosen action type is "Draw", a number of greyed out atoms and bonds to those atoms should be drawn, out of which the user can choose where he wants to place his next atom.
 - ii. When the chosen action type is "Bond", the program should draw a bond from the selected atom to every existing atom on the canvas, with which a bond would be possible. The colour of those bonds needs to be different to the colour of the actual existing bonds, to avoid confusion.

1.4 Current and Proposed Systems

The current standard is to work out chemical equations or draw molecules on paper. This might make sense for simple equations or small molecules, but it gets less efficient and more difficult as complexity increases. Although there are some software solutions for very specific tasks, there isn't one intuitive and easy to use application that combines the different tasks in one place.

Naturally, drawing molecules with pen and paper feels best and is the preferred choice by most people. This project is not here to replace that, it should merely pose as a help for chemists when working certain things out.

2 Documented Design

2.1 Introduction to the Documented Design

In this section, I will outline the decisions, that have outlined the development of the ChemBox project and explain the programming techniques used to implement certain algorithms and structures.

The program is written in python, with the aim of using as little external frameworks and dependencies as possible, and therefore creating most of this project from scratch. For the GUI implementation, I chose to use the PyQt6 framework, which is a powerful tool for creating GUI applications in python.

2.2 Project Hierarchy

As mentioned in the analysis part of this document, the project is separated into three stand alone modules which are merged in the main class of the program, ChemBox.

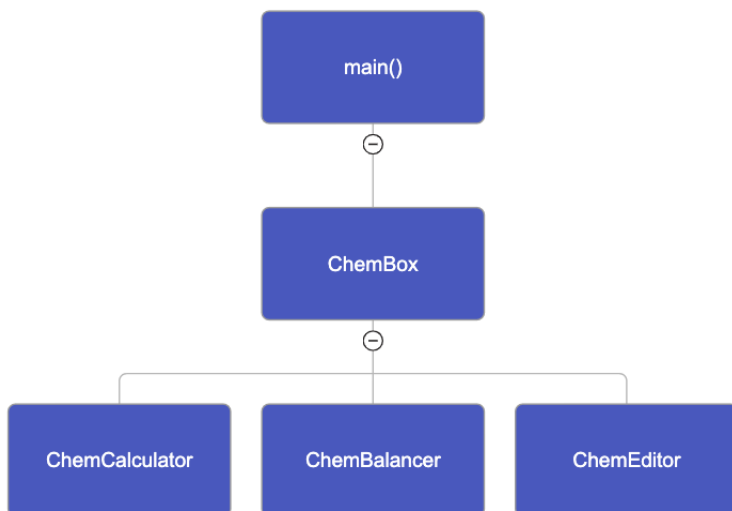


Figure 1: Program Hierarchy chart

I will explain the hierarchy and program flow of the system, beginning with the ChemCalculator module.

2.3 Structure of the GUI

The ChemBox class is the heart of the program, the point where all the different components are merged together to create one complete application. To give a better understanding of the system, I will explain how I composed the GUI and what each major component does. The ChemBox class contains only the constructor method, in which the layout of the graphical user interface is specified. The constructor defines the dimensions, the title and geometry of the window and then creates an instance of the TabBar class, which will act as the central widget of the program. Next, an instance of each of the three big components of the project, ChemCalculator, ChemBalancer and ChemEditor is created, and allocated to a separate tab of the tab bar (Listing 1: lines 18 - 25).

```
1 class ChemBox(QMainWindow):
2     def __init__(self):
3         super().__init__()
4
5         # set window properties
6         self.__left = 300
7         self.__top = 300
8         self.__width = 1280
9         self.__height = 720
10        self.__title = "ChemBox"
11        self.setWindowTitle(self.__title)
12        self.setGeometry(self.__left, self.__top, self.__width, self.
13        __height)
14        self.setFixedSize(self.__width, self.__height)
15
16        self.tab_bar = TabBar()
17        self.setCentralWidget(self.tab_bar)
18
19        self.chem_calculator = ChemCalculator()
20        self.tab_bar.tab1.setLayout(self.chem_calculator.main_layout)
21
22        self.chem_balancer = ChemBalancer()
23        self.tab_bar.tab2.setLayout(self.chem_balancer.balancer_layout)
24
25        self.chem_editor = ChemEditor()
26        self.tab_bar.tab3.setLayout(self.chem_editor.editor_layout)
```

Listing 1: Code snippet of ChemBox class No.1

The tab bar is used as the main widget of the program at all times, as it controls the navigation between modules. The goal was to achieve a design like the one illustrated in Figure 2, where 1, 2 and 3 in the small boxes at the top of the screen represent buttons in the tab bar for the three major modules of the ChemBox.

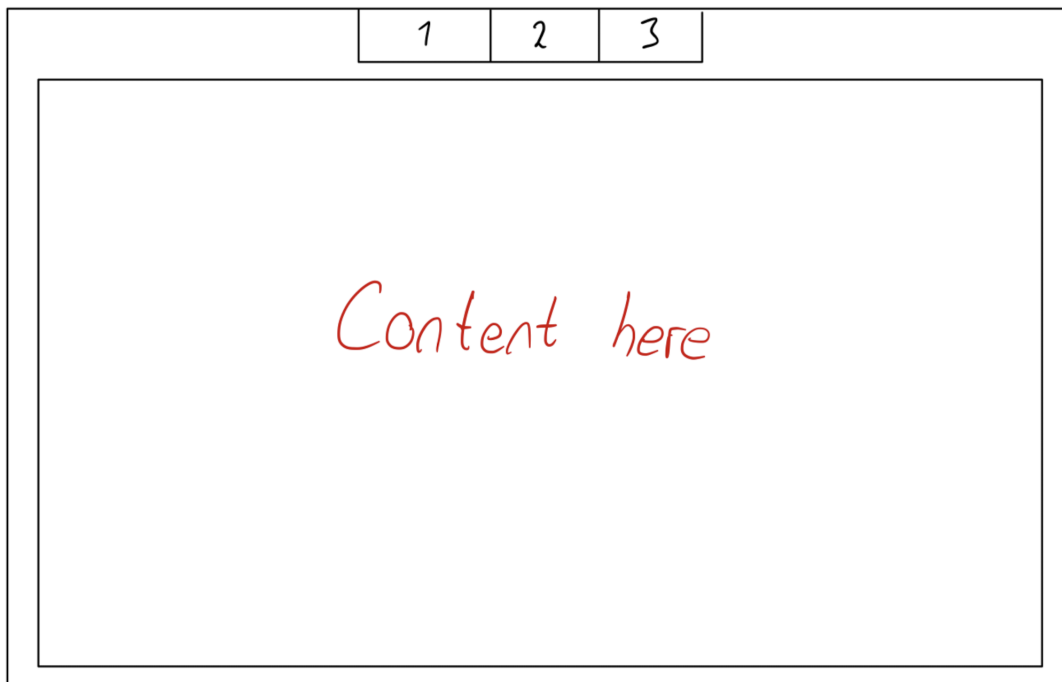


Figure 2: Mockup drawing of tab bar

The ChemCalculator module is the only part of the project with a special implementation of the user interface, which needs explanation.

When developing the ChemCalculator module, the first issue I encountered was how I would create a layout that would work for multiple separate sub-calculators. I had three different possible systems for displaying the module. The first option was creating a sidebar, where the user can choose the exact calculator they were looking for, which is then displayed on the screen. The second option was very similar, but with an additional tab bar at the top or bottom of the screen. This solution is not very aesthetically pleasing, and could cause confusion with the actual tab bar that allows the user to switch between the three main modules. The third option was putting every calculator on the same page, each in its own area, clearly separated from the others, and make the window scrollable. I decided against this option,

as this design could have gotten very messy. Therefore, I decided to create a sidebar for the ChemCalculator.

As there is no built-in sidebar widget in PyQt6, I have designed my own way of creating one. I did so by using the QTabWidget, which I also used for the tab bar, but not display it. I was able to benefit from the already existing widget, for the switching between calculators, and use buttons placed on the left side of the screen as a replacement for an actual sidebar. The buttons connect to a method, which changes the the current index of the tab widget to the according number of the calculator, which updates the page that is displayed. This is illustrated in the following code snippet (Listing 2), which only includes the example on a single button, for better readability of this document.

```
1 class ChemCalculator(QWidget):
2     def __init__(self):
3         super(QWidget, self).__init__()
4
5         self.side_bar_layout = QVBoxLayout()
6
7         self.gibbs_calc = GibbsFreeEnergyCalculator()
8
9         # Create buttons
10        self.gibbs_free_energy_tab_button = QPushButton("Gibbs Free Energy
11        Calculator")
12
13        self.gibbs_free_energy_tab_button.clicked.connect(self.
14        gibbs_free_energy_action)
15
16        # Create tabs
17        self.gibbs_free_energy_tab = QWidget()
18
19        # Initialise gibbs free energy calculator
20        self.gibbs_free_energy_tab.setLayout(self.gibbs_calc.layout)
21
22        # Add buttons to sidebar layout
23        self.side_bar_layout.addWidget(self.gibbs_free_energy_tab_button)
24
25        self.side_bar_widget = QWidget()
26        self.side_bar_widget.setLayout(self.side_bar_layout)
27
28        self.page_widget = QTabWidget()
29
30        self.page_widget.addTab(self.gibbs_free_energy_tab, "")
31
32        self.page_widget.setCurrentIndex(0)
33        self.page_widget.setStyleSheet('''QTabBar::tab{
34        width: 0;
35        height: 0;
36        margin: 0;
37        padding: 0;
38        border: none;
39        }''')
```

```

40     self.main_layout.addWidget(self.side_bar_widget)
41     self.main_layout.addWidget(self.page_widget)
42
43     self.main_widget = QWidget()
44     self.main_widget.setLayout(self.main_layout)
45
46     # Define actions for each button
47     def gibbs_free_energy_action(self):
48         self.page_widget.setCurrentIndex(5)

```

Listing 2: Example of ChemCalculator implementation

2.4 Algorithm Design for ChemCalculator

Essential for most individual calculators in the ChemCalculator module will be an algorithm to determine which of the input options the user left blank. In general, the user interface will always consist of a number of inputs, implemented as QLineEdit's, where the blank ones are the ones that our program will calculate.

The algorithm will have to take a list of inputs as a parameter, and use iteration to determine the blank one.

Algorithm 1 Algorithm to find empty input

```

 ← []
count ← 0
empty_input ← NONE
FOR  $i \leftarrow 0$  to  $Len(input\_list)$  DO
    IF  $input\_list[i]$  is empty THEN
        count ← count + 1
        empty_input ←  $input\_list[i]$ 
    END IF
END FOR
IF count = 1 THEN
    RETURN count
END IF

```

2.5 Algorithm Design for ChemBalancer

Figure 3 includes my initial attempt at visualising how an equation balancer could work. My first version of this module worked on a very similar system, although I struggled with finding a method of consistently balancing the equations after splitting them into their smallest possible components.

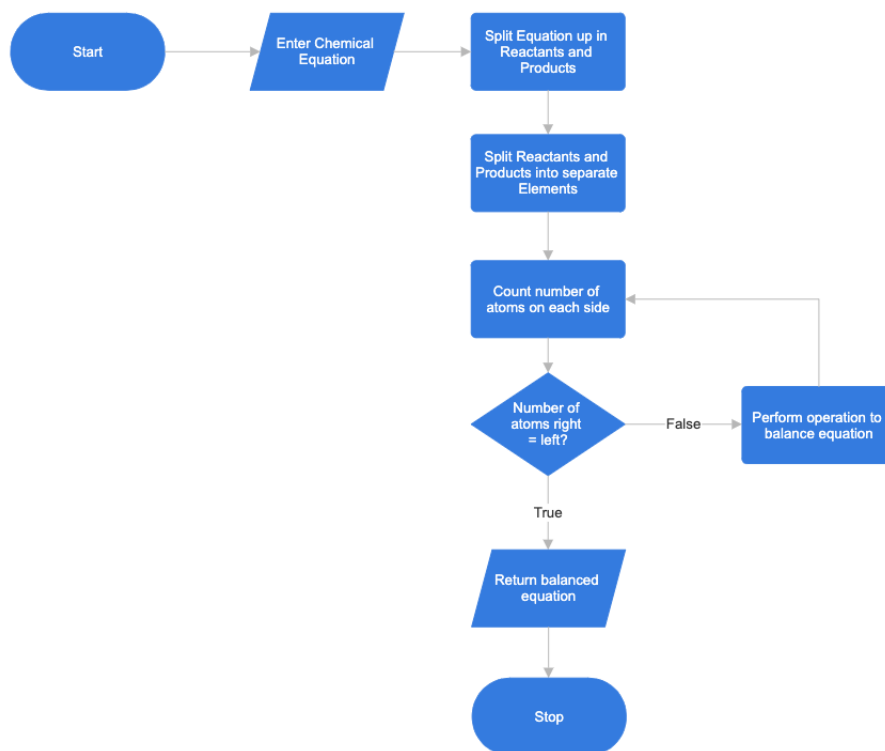


Figure 3: Initial flowchart for balancer

The following code blocks contains the code of my initial, failed, attempt at the balancer.

```
1 def splitEquation(self):
2     self.equationSplit = self.equationInput.text().split(" = ")
3     self.reactants = self.equationSplit[0]
4     try:
5         self.products = self.equationSplit[1]
6     except IndexError:
7         print("Wrong user input")
8     self.reactantComponents = self.reactants.split(" + ")
9     self.productComponents = self.products.split(" + ")
```

Listing 3: Method for splitting equation into components

The splitEquation method uses string manipulation to extract the individual molecules of the equation and store them in separate lists for reactants

and products.

After finding the individual molecules in the equation, the `parseComponent` method is called to get the elements used in the molecules, and the amount of them. This is done through if statements, as there is only a limited amount of different possibilities. The longest element symbol consists of three letters, where the first letter of every atom is always capitalised, and the following ones are not. The subscript number (number of atoms or element in a given molecule) can be found easily, as numbers within molecules always belong to the preceding element (H₂O - the 2 belongs to the hydrogen). Individual elements are found by checking for capital letters. An element starts with a capital letter and ends at the next capital letter, which is where the next element starts. The same techniques were used to find brackets and complex ions in the equation.

```
1 def parseComponent(self, component, countsDict, totalDict):
2     # Check for coefficient
3     try:
4         if component[0] in self.integers:
5             try:
6                 if component[0 + 1] in self.integers:
7                     try:
8                         if component[0 + 2] in self.integers:
9                             coefficient = int(component[0: 0 + 3])
10                        else:
11                            coefficient = int(component[0: 0 + 2])
12                        except IndexError:
13                            coefficient = int(component[0: 0 + 2])
14                    else:
15                        coefficient = int(component[0])
16                except IndexError:
17                    coefficient = int(component[0])
18            else:
19                coefficient = 1
20        for i in range(len(component)):
21            try:
22                openBracket = component.find("(")
23                closedBracket = component.find(")")
24            except IndexError:
25                continue
26            try:
27                if component[i].isupper() and component[i - 1] != "(":
28                    try:
29                        if component[i + 1].islower():
30                            try:
31                                if component[i + 2].islower():
32                                    element = component[i:(i + 3)]
33                                    if openBracket < i < closedBracket:
34                                        subCoefficient = self.
35                                else:
36                                    subCoefficient = 1
37                            # Check for subscript
38                        try:
```

```

40         if component[i + 3] in self.
integers:
41             try:
42                 if component[i + 4] in
self.integers:
43                     try:
44                         if component[i +
5] in self.integers:
45                             subscript =
46                             int(component[(i + 3): (i + 6)])
47                             else:
48                                 subscript =
49                                 int(component[(i + 3): (i + 5)])
50                                 except IndexError:
51                                     subscript = int(
component[(i + 3): (i + 5)])
52                                     else:
53                                         subscript = int(
component[i + 3])
54                                         except IndexError:
55                                             subscript = int(
component[i + 3])
56                                             else:
57                                                 subscript = 1
58                                                 except IndexError:
59                                                     subscript = 1
60                                                     else:
61                                                         element = component[i:(i + 2)]
62                                                         if openBracket < i < closedBracket:
63                                                             subCoefficient = self.
getSubCoefficient(component)
64                     else:
65                         subCoefficient = 1
66                     # Check for subscript
67                     try:
68                         if component[i + 2] in self.
integers:
69                             try:
70                                 if component[i + 3] in
self.integers:
71                                     try:
72                                         if component[i +
4] in self.integers:
73                                             subscript =
74                                             int(component[(i + 2): (i + 5)])
75                                             except IndexError:
76                                                 subscript = int(
component[(i + 2): (i + 4)])
77                                                 except IndexError:
78                                                     subscript = int(
component[i + 2])
79                                                     else:
80                                                         subscript = 1
81                                                         except IndexError:
82                                                             subscript = 1
83                     except IndexError:
84                         element = component[i:(i + 2)]

```



```

84         if openBracket < i < closedBracket:
85             subCoefficient = self.
getSubCoefficient(component)
86         else:
87             subCoefficient = 1
88     else:
89         element = component[i]
90         if openBracket < i < closedBracket:
91             subCoefficient = self.getSubCoefficient(
component)
92         else:
93             subCoefficient = 1
94         try:
95             # Check for subscript
96             if component[i + 1] in self.integers:
97                 try:
98                     if component[i + 2] in self.
integers:
99                     try:
100                         if component[i + 3] in
self.integers:
101                             subscript = int(
component[i + 1: i + 4])
102                             except IndexError:
103                                 subscript = int(
component[i + 1: i + 3])
104                             else:
105                                 subscript = int(component[i
+ 1])
106                             except IndexError:
107                                 subscript = int(component[i +
1])
108                             else:
109                                 subscript = 1
110                             except IndexError:
111                                 subscript = 1
112             except IndexError:
113                 element = component[i]
114
115         if openBracket < i < closedBracket:
116             subCoefficient = self.getSubCoefficient(
component)
117         else:
118             subCoefficient = 1
119         try:
120             # Check for subscript
121             if component[i + 1] in self.integers:
122                 try:
123                     if component[i + 2] in self.integers
:
124                     try:
125                         if component[i + 3] in self.
integers:
126                             subscript = int(
component[i + 1: i + 4])
127                             except IndexError:
128                                 subscript = int(component[i
+ 1: i + 3])

```

```

130         else:
131             subscript = int(component[i +
132
133         except IndexError:
134             subscript = int(component[i + 1])
135         else:
136             subscript = 1
137         except IndexError:
138             subscript = 1
139         try:
140             if element in countsDict:
141                 countsDict[element] = int(countsDict[element
142 ]) + subscript * coefficient * subCoefficient
143             else:
144                 countsDict[element] = subscript *
145 coefficient * subCoefficient
146
147             if element in totalDict:
148                 totalDict[element] = int(totalDict[element])
149 + subscript * coefficient * subCoefficient
150             else:
151                 totalDict[element] = subscript * coefficient
152 * subCoefficient
153         except UnboundLocalError:
154             continue
155
156         # Check for brackets / complex ion in equation
157         elif component[i] == "(":
158             try:
159                 if component[i + 2] == ")":
160                     element = component[i + 1]
161                     try:
162                         if component[i + 3] in self.integers:
163                             try:
164                                 if component[i + 4] in self.
165 integers:
166                                     try:
167                                         if component[i + 5] in
168 self.integers:
169                                             subscript = int(
170 component[(i + 3): (i + 6)])
171                                         else:
172                                             subscript = int(
173 component[(i + 3): (i + 5)])
174                                         except IndexError:
175                                             subscript = int(
176 component[(i + 3): (i + 5)])
177                                         else:
178                                             subscript = int(component[(i
179 + 3): (i + 4)])
180                                         except IndexError:
181                                             subscript = int(component[i +
182
183         else:
184             subscript = 1
185         except IndexError:
186             subscript = 1
187         elif component[i + 1].isupper():
188             if component[i + 2].islower():

```

```

177         try:
178             if component[i + 3].islower():
179                 element = component[(i + 1): (i
+ 4)]
180
181             # Check for subscript within
brackets
182
183             try:
184                 if component[i + 4] in self.
integers:
185
186                 try:
187                     if component[i + 5]
in self.integers:
188
189                     try:
190                         if component
[i + 6] in self.integers:
191
192                             subscript = int(component[(i + 4): (i + 7)])
193
194                             else:
195
196                                 subscript = int(component[(i + 4): (i + 6)])
197
198                                 except
IndexError:
199                                     subscript =
int(component[(i + 4): (i + 6)])
200
201                             else:
202                                 subscript = int(
component[i + 4])
203
204                             except IndexError:
205                                 subscript = int(
component[i + 4])
206
207                             else:
208                                 subscript = 1
209                             except IndexError:
210                                 subscript = 1
211
212             # Find subscript coefficient of
complex ion
213
214             subCoefficient = self.
getSubCoefficient(component)
215
216             else:
217                 element = component[(i + 1): (i
+ 3)]
218
219                 # Check for subscript within
brackets
220
221                 try:
222                     if component[i + 3] in self.
integers:
223
224                     try:
225                         if component[i + 4]
in self.integers:
226
227                             try:
228                                 if component
[i + 5] in self.integers:
229
230                                     subscript = int(component[(i + 3): (i + 6)])
231
232                                     else:

```

```

217 subscript = int(component[(i + 3): (i + 5)])
218
219 except
220
221 subscript =
222 int(component[(i + 3): (i + 5)])
223
224 except IndexError:
225 subscript = int(
226 component[i + 3])
227
228 else:
229 subscript = 1
230 except IndexError:
231 subscript = 1
232
233 # Find subscript coefficient of
234 complex ion
235 subCoefficient = self.
236 getSubCoefficient(component)
237
238 except IndexError:
239 element = component[(i + 1): (i + 3)]
240
241 # Check for subscript within
242 brackets
243 try:
244 if component[i + 3] in self.
245 integers:
246 try:
247 if component[i + 4] in
248 self.integers:
249 try:
250 if component[i +
251 5] in self.integers:
252 subscript =
253 int(component[(i + 3): (i + 6)])
254
255 else:
256 subscript =
257 int(component[(i + 3): (i + 5)])
258
259 except IndexError:
260 subscript = int(
261 component[(i + 3): (i + 5)])
262
263 except IndexError:
264 subscript = int(
265 component[i + 3])
266
267 else:
268 subscript = 1
269 except IndexError:
270 subscript = 1
271
272 # Find subscript coefficient of
273 complex ion
274 subCoefficient = self.
275 getSubCoefficient(component)
276
277 else:
278 element = component[i + 1]
279
280 # Check for subscript within brackets
281 try:
282 if component[i + 2] in self.integers
283 :

```

```

258         try:
259             if component[i + 3] in self.
integers:
260                 try:
261                     if component[i + 4]
subscript = int(
262                     component[(i + 2): (i + 5)])
263                     else:
264                         subscript = int(
265                             component[(i + 2): (i + 4)])
266                     except IndexError:
267                         subscript = int(
268                             component[(i + 2): (i + 4)]
+ 2])
269                     except IndexError:
270                         subscript = 1
271                     except IndexError:
272                         subscript = 1
273
274                     # Find subscript coefficient of complex
ion
275                     subCoefficient = self.getSubCoefficient(
component)
276                     except IndexError:
277                         print("wrong user input ")
278                     try:
279                         if element in countsDict:
280                             countsDict[element] += subscript *
subCoefficient * coefficient
281                         else:
282                             countsDict[element] = subscript *
subCoefficient * coefficient
283                         if element in totalDict:
284                             totalDict[element] += subscript *
subCoefficient * coefficient
285                         else:
286                             totalDict[element] = subscript *
subCoefficient * coefficient
287                     except UnboundLocalError:
288                         continue
289                     else:
290                         continue
291                     except IndexError:
292                         continue
293                     except IndexError:
294                         None

```

Listing 4: parseComponent method for finding elements

Although this code is overly complicated, it worked most of the times for basic or intermediate level equations.

The main problem this version of the balancer ran into, was the actual balancing of the equations. The version uses a loop to attempt balancing, where copies of the left and right hand side components are created together with

dictionaries to track the total count of elements. Each component then gets a randomly generated coefficient between 1 and 10, and the loop continues until the count on the left side equals the count on the right side. Once a balanced equation is found, the coefficients are normalised to their smallest integer values using the greatest common divisor (GCD) The balanced equation is then put back together and returned to the user.

```

1  def balance(self):
2      if self.balanced:
3          equation = str()
4          for dictionary in self.left:
5              compound = str()
6              for element in dictionary:
7                  compound += element
8                  if dictionary[element] > 1:
9                      compound += str(dictionary[element])
10             equation += compound
11             equation += " + "
12         equation = equation[:len(equation) - 3] + " = "
13         for dictionary in self.right:
14             compound = str()
15             for element in dictionary:
16                 compound += element
17                 if dictionary[element] > 1:
18                     compound += str(dictionary[element])
19             else:
20                 pass
21             equation += compound
22             equation += " + "
23         equation = equation[:len(equation) - 2]
24     else:
25         while not self.balanced:
26             tempLeft = list()
27             tempRight = list()
28             totalLeft = dict()
29             totalRight = dict()
30
31             for item in self.left:
32                 newDict = dict()
33                 for key in item:
34                     newDict[key] = item[key]
35                 tempLeft.append(newDict)
36
37             for item in self.right:
38                 newDict = dict()
39                 for key in item:
40                     newDict[key] = item[key]
41                 tempRight.append(newDict)
42
43             leftCoefficients = [randint(1, 10) for _ in range(len(
tempLeft))]
44             rightCoefficients = [randint(1, 10) for _ in range(len(
tempRight))]
45
46             for index in range(0, len(leftCoefficients)):

```

```

47         for key in tempLeft[index]:
48             tempLeft[index][key] *= leftCoefficients[index]
49             if key not in totalLeft:
50                 totalLeft[key] = tempLeft[index][key]
51             else:
52                 totalLeft[key] += tempLeft[index][key]
53
54         for index in range(0, len(rightCoefficients)):
55             for key in tempRight[index]:
56                 tempRight[index][key] *= rightCoefficients[index]
57                 if key not in totalRight:
58                     totalRight[key] = tempRight[index][key]
59                 else:
60                     totalRight[key] += tempRight[index][key]
61
62         self.balanced = True
63         for key in totalLeft:
64             if totalLeft[key] != totalRight[key]:
65                 self.balanced = False
66             else:
67                 continue
68
69         bigTup = tuple(leftCoefficients + rightCoefficients)
70         leftCoefficients = list(map(lambda x: int(x / reduce(gcd, bigTup
71         )), leftCoefficients))
72         rightCoefficients = list(map(lambda x: int(x / reduce(gcd,
73         bigTup)), rightCoefficients))
74
75         balancedEquation = str()
76         for index in range(0, len(self.left)):
77             if leftCoefficients[index] != 1:
78                 compound = str(leftCoefficients[index])
79             else:
80                 compound = str()
81             for key in self.left[index]:
82                 compound += key
83                 if self.left[index][key] != 1:
84                     compound += str(self.left[index][key])
85             balancedEquation += compound
86             balancedEquation += " + "
87         balancedEquation = balancedEquation[:len(balancedEquation) - 3]
88         + " = "
89         for index in range(0, len(self.right)):
90             if rightCoefficients[index] != 1:
91                 compound = str(rightCoefficients[index])
92             else:
93                 compound = str()
94             for key in self.right[index]:
95                 compound += key
96                 if self.right[index][key] != 1:
97                     compound += str(self.right[index][key])
98             balancedEquation += compound
99             balancedEquation += " + "
100         balancedEquation = balancedEquation[:len(balancedEquation) - 2]
101         return self.balancedLabel.setText(f"{balancedEquation}")

```

Listing 5: balance method

This version had a few obvious limitations, it relied on randomly generated coefficients, which is not a very reliable system of finding the correct coefficients. Although the random generation usually provides coefficients, it may not always produce the most optimal or smallest possible coefficients. The normalisation step using the greatest common divisor ensures that the coefficients are integer multiples of each other, but it does not guarantee the smallest possible coefficients. If a balanced equation cannot be found with the randomly generated coefficients, there is also a possibility of infinite looping.

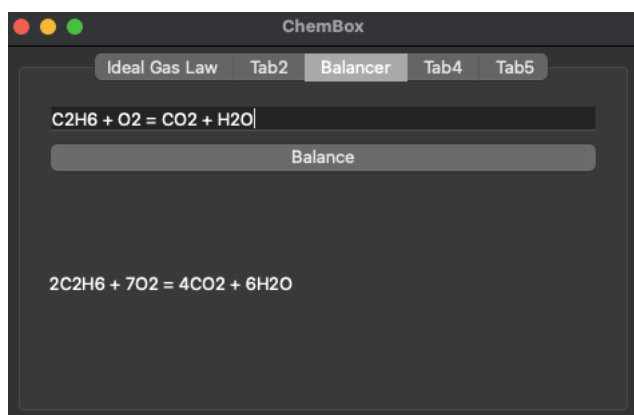


Figure 4: Initial balancer test

The test in Figure 4 shows, that the system works (at least for simple equations). Just the simple addition of two carbon atoms and four hydrogen atoms to get the equation $C_2H_6 + O_2 = CO_2 + H_2O$ is already too much for the system, as this will end in an infinite loop. This loop can be prevented by increasing the range of possibilities for the randomly generated integers (for example from between 1 and 10 to between 1 and 100) but this means there is always a limit to what is possible for the balancer. Figure 5 shows the state of the program after trying to balance the equation with the original range of one to 10. The program in the figure is in the state of an infinite loop.

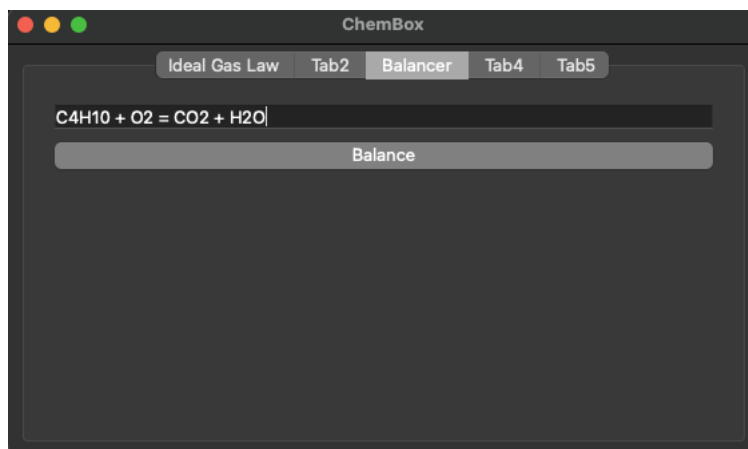


Figure 5: Initial balancer test No. 2

Whilst researching for better ways to implement a chemical equation balancer in python, I came across regular expressions, which around that same time, we also covered in our A-Level Computer Science lessons. Knowing about regular expressions helped me significantly improve the effectiveness of my balancer, when it comes to extracting individual elements and their amount from the equation. In my search for the best possible way to do so, I wrote a number of algorithms, which extend each other perfectly, to finish the job together.

This algorithm (Algorithm 2) takes the full equation as a parameter from the user input and removes any white spaces from it. It then splits it up into separate reactants and products, and lastly, it extracts the individual reagents within the reactants and products.

Algorithm 2 Algorithm to split equation

```

SUBROUTINE split_equation()
    stripped_equation  $\leftarrow$  USERINPUT
    stripped_equation  $\leftarrow$  stripped_equation.STRIP(" ")
    equation_split  $\leftarrow$  stripped_equation.SPLIT(" = ")
    reactants  $\leftarrow$  equation_split[0].SPLIT(" + ")
    products  $\leftarrow$  equation_split[1].SPLIT(" + ")
END SUBROUTINE

```

This second algorithm (Algorithm 3), which I designed for the task mentioned above, uses regular expressions, to identify separate reagents within a chemical compound by identifying brackets. The goal is to split up compounds like $Cu(NO_3)_2$ into $[Cu, (NO_3)_2]$, where Cu and $(NO_3)_2$ are

two separate entries in the list. The method then iterates over each reagent, and checks if it starts with an opening bracket. If it does, this indicates, that there is a chemical compound enclosed in the brackets. It extracts the inner compound and the subscript ((*OH*)₂ indicates that the (*OH*) group exists twice), and then passes these two variables as parameters to the "find_elements" method, along with the "index" and "side" parameters.

Algorithm 3 Algorithm to find reagents

```

SUBROUTINE find_reagents(compound, index, side)
  reagents ← SPLIT compound INTO reagents USING REGEX PATTERN
  FOR reagent IN reagents DO
    IF reagent BEGINS WITH "(" THEN
      inner_compound ← SUBSTRING(1, LEN(reagent))
      bracket_subscript ← SPLIT reagent BY ")" AND GET SECOND PART
      IF bracket_subscript EXISTS THEN
        bracket_subscript ← INT(bracket_subscript)
      ELSE
        bracket_subscript ← 1
      END IF
      find_elements(inner_compound, index, bracket_subscript, side)
    ELSE
      bracket_subscript ← 1
      find_elements(reagent, index, bracket_subscript, side)
    END IF
  END FOR
END SUBROUTINE

```

The next algorithm (Algorithm 4) uses a regular expression to obtain the elements and associated subscript values. Each extracted element is then stored together with the correlated subscript value as a tuple, inside a list (For example: $Cr_2O_7 \rightarrow [("Cr", "2"), ("O", "7")]$). The algorithm then iterates through each tuple (elements, subscript) in the element_counts list. With every iteration, a different subroutine named add_to_matrix is called, passing the current element, index, the product of the bracket_subscript and subscript, and the side argument.

Algorithm 4 Algorithm to find elements

```
SUBROUTINE find_reagents(reagent, index, bracket_subscript, side)  
  element_counts  $\leftarrow$  SPLIT reagent INTO element_counts USING REGEX PATTERN  
  FOR element, subscript IN element_counts DO  
    IF subscript DOES NOT EXIST THEN  
      subscript  $\leftarrow$  1  
    ELSE  
      subscript  $\leftarrow$  INT(subscript)  
    END IF  
    add_to_matrix(element, index, bracket_subscript * subscript, side)  
  END FOR  
END SUBROUTINE
```

While I was conducting my research, I found a neat solution to balancing equations, matrix operations. As we had only covered the rudimentary principles of matrix operations, I chose to use the article I had found on the internet to get a better understanding of the topic. After getting a grasp of the concept, and after trying out the code in the article, I had found a different way to tackle my problem. I decided not to reinvent the wheel, but to take the already existing algorithm and improve it, so it could be used for practically any tractable equation there is.

To address the problem, I decided to record a number of issues and possible improvements for the existing algorithm.

1. The first issue I had with the algorithm was, that it asked for separate inputs of reactants and products. Luckily, I had already solved this problem with the algorithms illustrated earlier.
2. Another issue the original algorithm had was the lack of structure and modularity. By organising the code in a class based structure, I ensured good readability for the code, as well as making it easier to maintain.
3. The original version had poor input validation and error handling. One common error I got with the early implementation of the algorithm was what I call a nullspace error. The nullspace is a linear subspace that contains a set of vectors that transform to the zeroth vector under a given linear transformation: multiplication with the matrix A, represented as $Ax = 0$. In the context of chemical equations, the nullspace contains the coefficients, that balance the equation.

Entering the unbalanced equation $K4[Fe(SCN)6] + K2Cr2O7 + H2SO4 = Fe2(SO4)3 + Cr2(SO4)3 + CO2 + H2O + K2SO4 + KNO3$ resulted in a "nullspace error" (Figure 6).

```

Traceback (most recent call last):
  File "/Users/tom/Downloads/ChemBox-db54135db0ee918912ceb046cf7ef95920d8f1bf/main.py", line 345, in runBalancer
    num = self.elementMatrix.nullspace()[0]
          ~~~~~^~~~~~
IndexError: list index out of range

```

Figure 6: Nullspace error

I managed to neutralise this error by having a more accurate and efficient method of separating out molecules, elements, etc. as shown in Algorithm 3 and Algorithm 4, before passing them on to the method that adds the items to the matrix.

I also made sure to use python's built in exception handling to catch any expected and unexpected errors or exceptions.

Using the current working system, inputting the chemical equation mentioned before results in a perfectly balanced equation (Figure 7).

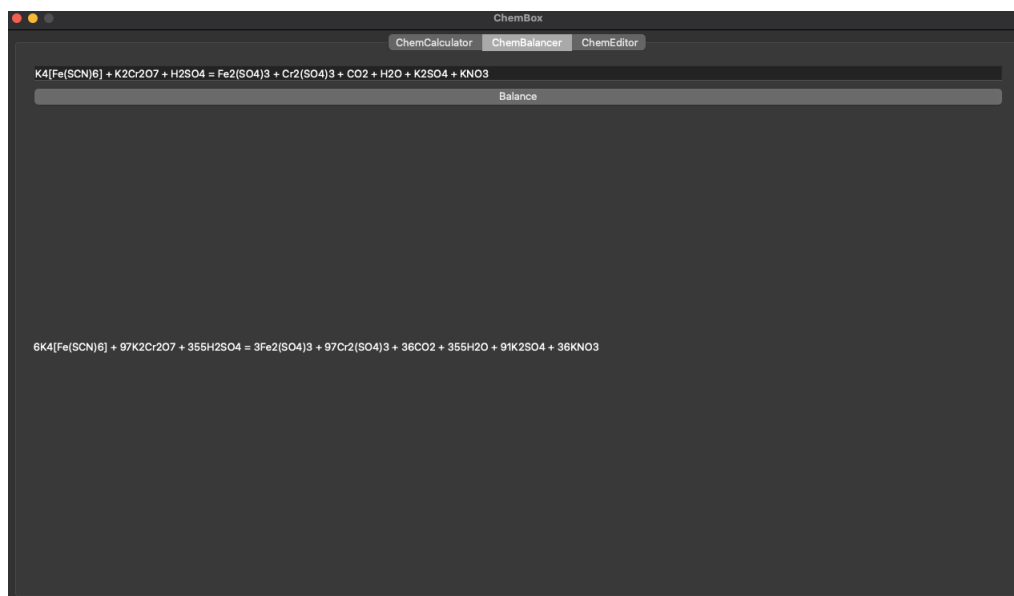


Figure 7: Balanced equation

3 Algorithm Design for ChemEditor

4 Testing

5 Evaluation