Unbalanced Chemical Equations Conversion to Mark-up Format and Representation to Vision Impaired Students

Azadeh Nazemi
Department of Electrical and Computer Engineering
Curtin University
Perth, Western Australia
azadeh.nazemi@postgrad.curtin.edu.au

David A. McMeekin
Department of Spatial Sciences
Curtin University
Perth, Western Australia
D.McMeekin@curtin.edu.au

Iain Murray
Department of Electrical and Computer Engineering
Curtin University
Perth, Western Australia
I.Murray@curtin.edu.au

ABSTRACT

Abstract--- This paper describes a method to represent unbalanced chemical equations to vision impaired students which allows them to navigate through classified data, such as species, elements, quantity numbers at the left and right hand sides of equations, reactants and products. Then they can find appropriate coefficients and balance chemical equations without involving to mathematical aspects of balancing and remembering a lot of information. The goal of this research was the development of an application which assists vision impaired students enrolled in chemistry course to be able to read chemistry literature containing formulae, chemistry representations of elements and other aspects of chemistry that has been difficult in the past to present in a way for vision impaired people to understand. Developed application by this research is an open source command line Bash Script application under Linux which accepts an unbalanced chemical equation as an input, processes, classifies information and represents it as Mark-up format or Alternative Audio Descriptive using Text to Speech.

General Terms

Open source command line application for vision impaired students

Keywords

Chemical Equation Balancing, Vision Impaired, Navigation, Mark-up Format, Bash Script, Assisitive Technology, Algebraic Balancing Method, Tactile Representation, Audio Representation, Text to Speech (TTS)

1. INTRODUCTION

Chemical representations refer to various types of formulae, structures, and symbols used to represent chemical processes and conceptual entities (e.g., molecules and atoms)[1]. Chemistry at the symbolic level is represented by symbols, numbers, formulae, equations, and structures [2]. Ben-Zvi et.al studies have shown that understanding symbolic representations is especially difficult for students because these representations are invisible and abstract while students' thinking relies heavily on sensory information. In addition, without substantial conceptual knowledge and visual-spatial ability, vision impaired students are unable to translate one given representation into another [3].

Mayo's [4] research results indicate that visually impaired students had difficulty interpreting common representations of the structures encountered in introductory chemistry and organic chemistry courses. These students have problems with balancing equations and are unable to produce accurate pictorial representations. They were also unable to interpret subscripts, coefficients, and implicit meanings of the equations [5].

For studying science, especially chemistry, the role of visual and symbolic modes of representations is extremely important. However, students with visual impairments cannot utilize these forms of visual and symbolic representations because of their impairment. Hence, the study of chemistry then becomes an area that is virtually unattainable for them.

According to the results of the research undertaken by Micklos Lewis & Bodner [6], current pedagogical practices need to be revised to enhance the conceptual understanding that all students develop of the symbolic representations used to describe chemical reactions.

2. CHEMICAL EQUATIONS AND BALANCING INTRODUCTION

Chemistry is the study of the properties of materials and the changes that those properties undergo. A chemical equation is the symbolic representation of a chemical reaction wherein the reactant entities are given on the left-hand side of the equation and the product entities are given on the right-hand side of the equation. According to the Law of Conservation of Mass and the Law of Conservation of Charge, the quantity of each element does not change in a chemical reaction. Thus, each side of the chemical equation must represent the same quantity of any particular element. Also, the same charge must be present on both sides of the balanced equation.

Chemical equations balancing plays a critical role in understanding the structure of chemical elements involved in chemical reactions. The ability to balance chemical equations in terms of mass and charge is a key skill that must be mastered by the serious chemistry student. Each reactant or product contains one, or more than one element and each element has a different quantity. In the result the balancing task must be performed so that equal quantities of each element are on the left and right sides of the equation. Balancing a chemical equation involves multiple steps:

1. Recognition of elements in each reactant on left side of the equation;
2. Recognition of elements in each product on the right side of the equation;
3. Finding quantity of each element on left side and right side;
4. Comparing the quantity of each elements on left and right sides; and
5. Appropriate coefficient calculation to insert before each reactant and products to make entire equation balanced.

These coefficients are the absolute values of the stoichiometric numbers and lead to chemical equation balancing. There are two methods for balancing chemical equations or finding coefficients:

- Inspection method is the usual method of balancing chemical equations. It is fast but confusing for complicated equations In addition the final result must be double checked to make sure that the chemical equation is indeed balanced;.
- The algebraic method uses algebraic equations and/or matrices. This method is systematic and can be applied to difficult reactions. It can be easily used with equation solvers but it is time-consuming to define corrected algebraic equations [7].

3. CURRENT METHODS TO REPRESENT CHEMICAL EQUATION TO VISION IMPAIRED STUDENTS

Traditionally there are two methods for representing symbolic level or chemical equations to vision impaired students. They are the:

1- Tactile method: The use of molecular models to teach balancing equations, which gives a concrete rather than abstract description of the problem, is an efficient approach. Using Braille or embossed paper is another tactile representation of chemical equations [8] and students can actively interact with the implicit information. Generally, the molecular formula shows the composition of a compound and the number of each type of atom in a molecule. For example molecular formula for benzene is: C_6H_6 . In terms of representation molecular formulae due to have subscript(s) are not entirely linear in nature. Since Braille is suitable for text representation, which is linear in nature, is not suitable for molecular chemical formulae representation.

2- Audio method or using assistive technologies such as Text to Speech (TTS). In audio representation, students passively hear all information from the left to right hand side without opportunity to rehear a specific part and/or interact with it. As it is perceived from previous section(s), balancing can be challenging even for sighted students who are able to use visual aids to communicate with and understand the equation's meaning.

For balancing, students need to have efficient knowledge of the chemical element structure as well as mathematical equation solving skills.

4. PROPOSED SOLUTION

Students with vision impairments will learn chemistry in the classroom and laboratory best and enjoy the most productive careers when they have access to the proper combinations of computer hardware and software and other assistive technology [10]. No vision impairment, including total blindness, should be a barrier to that process. Chemists who are blind use assorted assistive technology to work productively and safely in academia and industry [11]. Many adaptations are simple and readily available.

As stated previously, both current representation methods for chemical equation have disadvantages for vision impaired people hence, finding a new method suitable for vision impaired people is an important priority.

Chemical Mark-up Language - CML XML is a mainstream approach providing semantics for most chemistry, especially:

- Molecules reactions[12]
- Spectra and analytical data[13]
- Computational chemistry
- Chemical crystallography and material

CML has been developed by Peter Murray-Rust and Henry Rzepa since 1995. It is the de facto XML for chemistry, accepted by publishers and with more than 1 million lines of Open Source code supporting it. This research used CML idea to convert unbalanced chemical equation to a Mark-up format by embedding essential tags. Unbalanced chemical equation representation in Mark-up format provides an opportunity to navigate through the classified data. To achieve this purpose an application was developed in two main modules:

1. Basic module is responsible for
    - Classification of the implicit information in a chemical equation
    - Tagging classified information
    - Representing Mark-up format
2. Advanced module is responsible for generating algebraic equations related to chemical equation

To develop the application to convert the chemical equations to a Mark-up format, a Linux Bash Script was used. The Bash scripting language has several powerful text manipulation tools in command line mode which are suitable for vision impaired users. Vision impaired users can run the application in text mode as a command using Text to Speech without using any visual aspects.

5. METHODOLOGY

5.1 Species Classification and Reactants/Products Extraction

Species in a chemical reaction is a general term used to mean atoms, molecules or ions. A species can contain more than one chemical element. The general format for species is:

Species=$Symbol_{AtomicNumbr}^{Charge}$(PhysicalState)

The physical state of each reactant or product is represented by:

(l)=liquid

(g)=gas

(s)=solid.

Coefficients are useful for keeping the same number of atoms. Chemical equation in most cases comes as a text line, therefore to convert it to a Mark-up format text processing techniques are used. Suppose a chemical equation reads as a parameter such as this:

AgNO_3 *+Cu ->Cu(* NO_3 *)2 + Ag*

Equation=$1

To classify and equation the script needs to:

- Remove all space character from the chemical equation
- Convert yields sign to equal sign (−> to =)

```
Equation=$(echo $1 |sed 's/ //g;s/->/=/g;
s/-//g')
```

- Find "=" sign position

```
EqualSign=$(echo $Equation|grep -bo "="|
sed 's//:.*$//')
```

- Separate left and right side using "=" sign position:

```
Reactants=$Equation:0:$[$EqualSign+1])
```

```
Products=$Equation:$EqualSign)
```

- Extract reactants and product using "+ " sign position as the separator:

```
Reactant=($(echo Reactants|sed 's/+/ /g')
```

```
Product=($(echo Products|sed 's/+/ / g')
```

```
NoR=number of Reactants=${#Reactant[@]}
```

```
NoP=number of Products=${#Product[@]}
```

5.2 Chemical Elements Extraction and Symbol Replacement

This module is responsible to find the participating elements in a chemical reaction.

- It calculates the length of each species in terms of text processing, for example:

Table1. Two categories of symbols of elements

CATEGORY2					
SYMBOL	ELEMENT	SYMBOL	ELEMENT	SYMBOL	ELEMENT
He	Helium	Cs	Caesium	Es	Einsteinium
Li	Lithium	Ba	Barium	Fm	Fermium
Be	Beryllium	La	Lanthanum	Md	Mendelevium
Ne	Neon	Ce	Cerium	No	Nobelium
Na	Sodium	Pr	Praseodymium	Lr	Lawrencium
Mg	Magnesium	Nd	Neodymium	Rf	Rutherfordium
Al	Aluminium	Pm	Promethium	Db	Dubnium
Si	Silicon	Sm	Samarium	Sg	Seaborgium
Cl	Chlorine	Eu	Europium	Bh	Bohrium
Ar	Argon	Gd	Gadolinium	Hs	Hassium
Ca	Calcium	Tb	Terbium	Mt	Meitnerium
Sc	Scandium	Dy	Dysprosium	Ds	Darmstadtium
Ti	Titanium	Ho	Holmium	Rg	Roentgenium
Cr	Chromium	Er	Erbium	Cn	Copernicium
Mn	Manganese	Tm	Thulium		
Fe	Iron	Yb	Ytterbium		
Co	Cobalt	Lu	Lutetium		
Ni	Nickel	Hf	Hafnium		
Cu	Copper	Ta	Tantalum		
Zn	Zinc	Re	Rhenium		
Ga	Gallium	Os	Osmium		
Ge	Germanium	Ir	Iridium		
As	Arsenic	Pt	Platinum		
Se	Selenium	Au	Gold		
Br	Bromine	Hg	Mercury		
Kr	Krypton	Tl	Thallium		
Rb	Rubidium	Pb	Lead		
Sr	Strontium	Bi	Bismuth		
Zr	Zirconium	Po	Polonium		
Nb	Niobium	At	Astatine		
Mo	Molybdenum	Rn	Radon		
Tc	Technetium	Fr	Francium		
Ru	Ruthenium	Ra	Radium		
Rh	Rhodium	Ac	Actinium		
Pd	Palladium	Th	Thorium		
Ag	Silver	Pa	Protactinium		
Cd	Cadmium	Np	Neptunium		
In	Indium	Pu	Plutonium		
Sn	Tin	Am	Americium		
Sb	Antimony	Cm	Curium		
Te	Tellurium	Bk	Berkelium		
Xe	Xenon	Cf	Californium		

CATEGORY1	
SYMBOL	ELEMENT
H	Hydrogen
B	Boron
C	Carbon
N	Nitrogen
O	Oxygen
F	Fluorine
P	Phosphorus
S	Sulphur
K	Potassium
V	Vanadium
Y	Yttrium
I	Iodine
W	Tungsten
U	Uranium

$H_2 + O = H_2O$, the length of H_2O *is 3.*

```
Species=$reactant[$i]
SpeciesLength=$(echo $Speices| wc -c)
EquationLength=$(echo $Equation| wc -c)
```

- Allocate array[$SpeciesLength] to keep equation character by character;
- Investigate array members from the beginning of the array[$i]; and
- Parse equation from left to right to find elements.

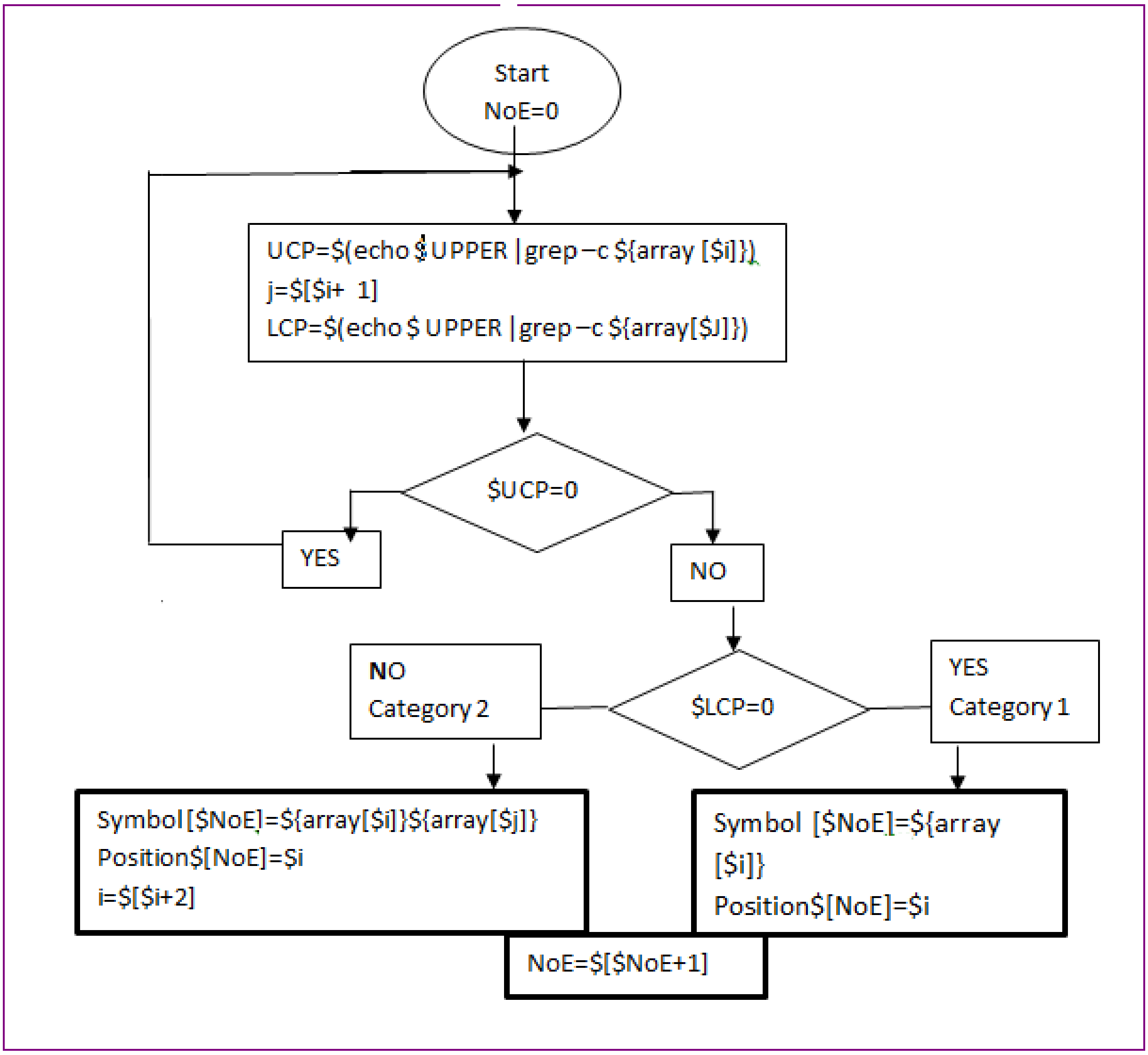
In terms of text processing chemical element symbols, they are divided into two categories:

1. Single uppercase character alphabet

2. Double character elements containing one uppercase character followed by a lower case character (as shown in Table 1)

If the array member is inside this string: upper="ABCDEFGHIJKLMNOPQRSTUVWXYZ" then read the next element. If the next element inside the string: lower="abcdefghijklmnopqrstuvwxyz" consider these two adjacent members of the array as an element in category 2. Otherwise the array member belongs to category 1. Figure 1 illustrates this procedure.

Fig 1: Flowchart for elements (symbols extraction)



AgNO$_3$ *+Cu ->Cu(NO*$_3$ *)2 + Ag*

N,O: category 1

Ag, Cu: category 2

For better understanding of the conceptual meaning of reaction and remembering the elements involved in the reaction which is shown in the equation, this module replaces the symbol of the extracted elements with a full name using the List of Periodic Table Elements. The following snippet shows the code for replacing the symbols by the full name from the elements.

```
for((k=0;k<$NumberElements;k++));do
sym=$symbol[$k]
ElementName[$k]= $(cat
chemicalelements.lib |awk '$1
/'$sym'/print $2'|head -1)
 done
```

AgNO$_3$ *+Cu ->Cu(NO*$_3$ *)2 + Ag*

Ag with Silver, N with Nitrogen, O with Oxygen and Cu with Copper are replaced

5.3 Calculations of Total Quantity of Element at Left and Right Sides and Comparison

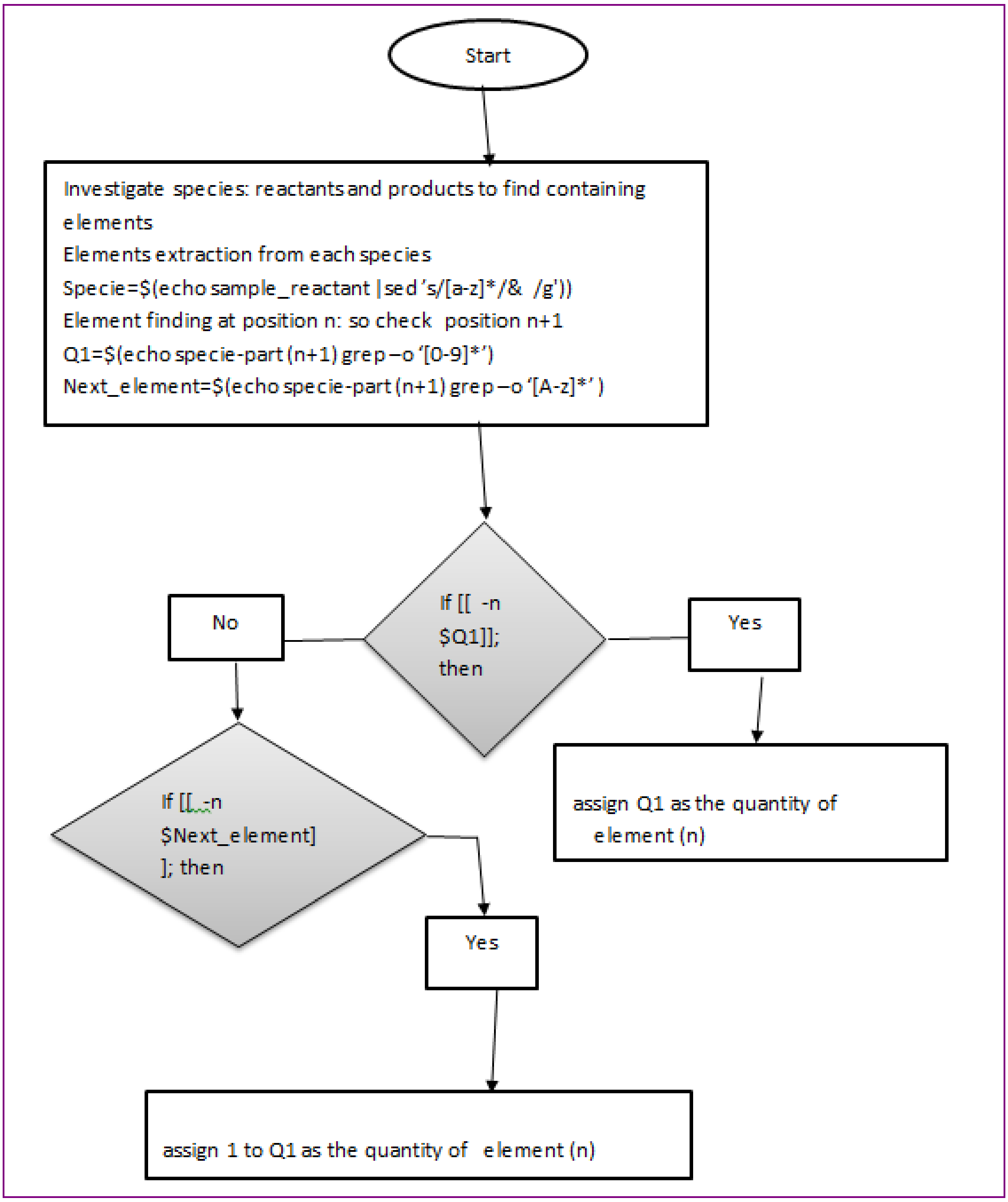
The quantity of each element appears in two forms:

- The first being: single quantity N: the number after an element indicates that there are "N" atoms of that particular element in each molecule. For example, the water molecule H 2 O has two hydrogen atoms. The quantity comes immediately after the symbol name. This kind of quantity is considered as Q1 in this research. Q1 only belongs to the element came before it. It is worth noting that if there is only one atom of an element in a molecule it is not written.

For example: $H_2 + O = H_2O$ the quantity of Oxygen is 1.

Figure 2 illustrates the procedure to find Q1

Fig 2 Single quantity finding



- The second being: parentheses or grouped quantity. Some formulae have parentheses and place a group of elements within parentheses '()' .The number comes immediately after the ')' parenthesis is a grouped quantity. This quantity is called Q2 by this research.

$(NH_4)\ 3PO_4 + Pb\ (NO_3)\ 4 = Pb_3\ \ (PO_4)\ 4 + NH_4NO_3$

This equation includes grouped quantity. Number 3 (in the first reactant) which comes after the closed bracket, belongs to all the elements that are enclosed within the '()'.It means the quantity of N is 3x1=3 and the quantity of H is 3x4=12.

In this case the total quantity of each element within the parentheses '()' is a product of Q1 x Q2 . This quantity after the parentheses '()' must be considered for all elements surrounded by the '()'. Thus to find the grouped quantity it must:

- Parse forward through the species from left to right until reaching) at position j;
- Parse back from) position toward start of equation until reaching (at i; and
- Use substring to extract part of species inside the bracket from position i and length j-i.

```
BracketSurrounded=$species:$i:$[$j-$i]
```

- Use previous part to find out elements and their quantity inside the parentheses.

```
Q2 or BracketRelatedQuantity= $(echo
$specie:$[$i+1]|grep -o '[0-9]* '|head -
1)
```

- Obtain Q2 and calculate Q1xQ2 for all elements inside the parentheses.

The following example shows the quantity of each element inside the parentheses after finding Q2.

$(E1_{Q11}E2\ \ E3_{Q13})Q2$

E1:Q11xQ2
E2:Q2
E3:Q13xQ2

5.4 Insert Unknown Coefficients before Species Including Reactants and Products

The coefficients are the numbers in front of each species. They have very important meaning and indicate the relative amount of atoms in reaction. This module is responsible for considering a unique unknown coefficient for each reactant or product as shown below:

a $(NH_4\ )\ 3PO_4$ +**b** $Pb\ (NO_3)\ 4$= **c** $Pb_3\ (PO_4)\ 4$+ **d** NH_4NO_3

The following snippet shows the conversion process of species to species with coefficient:

```
 alphabet="abcdefghiklmnopqrstuwxyz"
reactant=($(echo $1|sed 's/=.*//g'|sed
's/+/ /g'))
product=($(echo $1|sed 's/.*=//g'|sed
's/+/ /g'))
NoR=${#reactant[@]}
NoP=${#product[@]}
for ((r=0 ;r<$NoR;r++));do
CoeffReactant[$r]=${alphabet:$r:0}\${reac
atnt[$r]}
done
for ((p=0 ;p<$NoP;p++));do
CoeffProduct[$p]=${alphabet
[$p+NoR]:2}${product[$p]}
done
```

5.5 Defining Algebraic Equation

If the number of participating elements in the reaction is considered "n" that means there are "n" algebraic equations to be extracted by performing following steps for all participating elements in reaction one by one :

1. Consider an specific element
2. Remove all species which do not contain the specific element at the left and right
3. Keep all species containing the specific element at the left and right;
4. For remaining species obtain the mathematical production of the quantity of elements in species by the coefficient and replace species with obtained mathematical production

Table 2 indicates the algebraic equation for the example from the previous section.

Table 2 .Algebraic equations

Element	Related algebraic equation
N	$3a+4b=2d$
H	12=4d
P	a=4c
O	4a+12b=16c+3d
Pb	b=3c

5.6 Tagging Classified Information and Mark-up Format Generation

Reviewing previous sections shows which classified data have been extracted. Mark-up Format is the collection of this information with meaningful tags (Table 3).

Table 3. Mark-up format tags

Level	Information	Tags
2	Species	<species>,</species>
1	Reactants Products	<reactant>,</reactant> <product>,</product>
3	Symbol of Elements	<element>,</element>
3	Full name of Elements	<name>,</name>
3	Quantity in left and right	<Q_left>,</Q_left> <Q_right>,</Q_right>
2	Coefficients	<coefficient>,</ coefficient>

The following snippet indicates sample tags used by Mark-up Format:

```
<!DOCTYPE html><html>
<body> Unbalanced chemical equation:
AgNO3 + Cu = Cu(NO3)2 + Ag
<h1><reactant>reactants</reactant>
<h2><specie>AgNO3</specie>
<h3><element>Ag</element>
<name>Silver</name>
<left>1</left>
<element>N</element>
```

```
<name>Nitrogen</name>
<left>1</left>
<element>O< /element>
<name>Oxygen</name>
<left>3</left></h3></h2>
<h2><specie>Cu</specie>
<h3><element>Cu</element>
</name>Copper</name>
<left>1</left>
</h3></h2> </h1>
<h1><product>product </product>
<h2><specie>Cu(NO3)2</specie>
<h3> <element>Cu</element>
<name><copper</name>
<right>1</right>
<element>N</element>
<name>Nitrogen</name>
<right>2</right>
<element>O</element>
<name>Oxygen</name>
<right>6</right></h3></h2>
<h2><specie>Ag</specie>
<h3><element>Ag</element>
<name>Silver</name>
<right>1</right>
</h3> </h2> </h1>
</body>
</html>
```


Table 4.Results for three samples

Equation with Coefficients	Reactants	Products	Elements	L	R	Algebraic Equation
a Fe + b Cl2 = c FeCl3	Fe Cl2	FeCl3	Cl= Chlorine Fe= Iron	2 1	3 1	2b=3c a=c
a KMnO4 + b HCl = c KCl + d MnCl2 + e H2O + f Cl2	KMnO4	KCl MnCl2 H2O	Cl=Chlorine H=Hydrogen K=Potassium Mn=Manganese O=Oxygen	1 1 1 1 4	5 2 1 1 1	b=c+2f+2d b=2e a=c a=d 4a=e
a K4Fe(CN)6 + b H2SO4 + c H2O = d K2SO4 + e FeSO4 + f (NH4)2SO4 + g CO	K4Fe(CN)6 H2SO4 H2O	*K2SO4* *FeSO4* (NH4)2 SO4	C=Carbon Fe=Iron H=Hydrogen K=Potassium N=Nitrogen O=Oxygen S=Sulphur	*6* *1* *4* *4* *6* *5* *1*	*1* *1* *8* *2* *2* *17* *4*	*6a=g* *a=e* *2b+2c=8f* *4a=2d* *6a=2f* *4b+c=4d+4e+4f+g* *b=d+e+f*

6. TESTING

Over 100 simple, intermediate and complicated unbalanced chemical equation have been applied to the developed application in this research and equivalent description for these equations automatically provided as they are shown in table 4.

7. CONCLUSION

The obtained results from testing module clearly demonstrated that this application is reliable for chemical equation balancing and therefore it can be trusted by students. Hence, the research undertaken here to develop an application which provides opportunity to both sighted and vision impaired students to balance chemical equations, regardless of the involved mathematical aspects, achieved its desired goal. The embedded navigation ability feature in this application allows students to skip through different parts of the equation and to simply replay ambiguous parts.

By the time visually impaired students enrol in chemistry they should have acquired techniques for solving mathematical problems [14], thus the future work required, and already commenced, is the development of an evaluation module for the application. The evaluation module will assist users in inspecting the accuracies of the balancing of the equations which they have carried out.

8. REFERENCES

[1] Wu, Hsin-Kai; Krajcik, Joseph S.; Soloway, Elliot (2001)."Promoting understanding of chemical representations: Students' use of a visualization tool in the classroom." Journal of Research in Science Teaching 38(7): 821-842. <http://hdl.handle.net/2027.42/34515>

[2] Gabel, D. (1998). The complexity of chemistry and implications for teaching. In B.J. Fraser & K.G. Tobin (Eds.), International handbook of science education (pp. 233 ± 248). Boston, MA: Kluwer Academic Publishers.

[3] Ben-Zvi, R., Eylon, B., & Silberstein, J. (1987). Students' visualization of a chemical reaction. Education in Chemistry, July, 117 ± 12

[4] Mayo P., (2004), Assessment of the impact chemistry text and figures have on visually impaired students' learning, Doctoral Dissertation, Purdue University, West Lafayette.

[5] Yarroch W. L., (1985), Student understanding of chemical equation balancing, J. Res. Sci. Teach., 22(5), 449–459.

[6] Bodner, Amy L. Micklos Lewis and George M. 2013. "Chemical Reactions: What Understanding Do Students with Blindness Develop?" Chemistry education research and practice: 625-636.Mehmet Sahin, Nurettin Yorek 2009. "Teaching Science to Visually Impaired Students: A Small- Scale Qualitative Study∗." 6: 19-25.

[7] Jamin Santiago (2013) Balancing Chemical Equations – Easy Algebraic Method, Creative Commons Attribution 3.0 Unported (CC BY 3.0).Availabe at: http://jaminsantiago.files.wordpress.com/2013/04/balancing-chemical-equations-easy-algebraic-method.pdf

[8] Edman, Polly. Tactile Graphics; American Foundation for the Blind: New York, 1992.

[9] Dorothy L. Miner, Ron Nieman, Anne B. Swanson,and Michael Woods. 2001. "Teaching Chemistry to Students with Disabilities: A Manual for High Schools, Colleges, and Graduate Programs 4th Edition." 10-96'

[10] Scadden, L. A. An overview of technology and visual impairment. Technology and Disability, 1991, 1, 11. (62)

[11] Woods, M. Working Chemists with Disabilities, Expanding Opportunities in Science; Blumenkopf, T. A., et al., Eds.; American Chemical Society: Washington, DC, 1996; ISBN 0841235023

[12] Holliday, G. L.; Murray-Rust, P.; Rzepa, H. S. (2006), "Chemical Markup, XML and the World Wide Web. Part 6. CMLReact; An XML Vocabulary for Chemical Reactions", J. Chem. Inf. Mod. 46 (1): 145–157, doi:10.1021/ci0502698

[13] Kuhn, S.; Helmus, T.; Lancashire,, R. J.;Murray-Rust, P.; Rzepa, H. S.; Steinbeck, C.; Willighagen, E. L. (2007), "Chemical Markup, XML, and the World Wide Web. 7. CMLSpect, an XML Vocabulary for Spectral Data", J. Chem. Inf. Mod. 47 (6): 2015–2034, doi:10.1021/ci600531a, PMID 17887743

[14] Tombaugh, Dorothy. 1981. "Chemistry and the Visually Impaired " Journal of Chemical Education 58: 222-225.