A
Dissertation
On

"A Knowledge Based Expert System in Scientific Domain"

Submitted in Partial fulfillment of the requirement
for the award of the degree of
MASTER OF ENGINEERING
(Computer Technology & Application)

Submitted By:
Praveen Kumar
College Roll No. 10/CTA/09
University Roll No. 8549

Under the Guidance of:
Mr. R. K. Yadav
Assistant Professor,
Dept. of Computer Engineering
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)

DEPARTMENT OF COMPUTER ENGINEERING
DELHI COLLEGE OF ENGINEERING
UNIVERSITY OF DELHI

June 2011

CERTIFICATE

This is to certify that the work contained in this dissertation entitled "A Knowledge Based Expert System in Scientific Domain" submitted by Praveen Kumar**,** College Roll No: 10/CTA/09, Delhi University Roll No. 8549 is an account of his work carried out under my supervision in the academic year 2010-2011. This is submitted in partial fulfillment of the requirement for the Master of Engineering in Computer Technology & Application.

I wish his all the success in future.

Mr. R. K. Yadav

Assistant Professor
Department Of Computer Engineering
Delhi Technological University
(Formerly Delhi College of Engineering)
Delhi-110042

ACKNOWLEDGEMENT

First of all it is a great pleasure to express my gratitude to my supervisor Mr. R. K. Yadav, Assistant Professor, Department of Computer Engineering, Delhi Technological University *(Formerly Delhi College of Engineering)* for his invaluable guidance, motivation and support. With his continuous inspiration only, it becomes possible to complete this dissertation.

I would like to take this opportunity to express my sincere regards to Dr. Daya Gupta, HOD, Department of Computer Engineering, Delhi Technological University *(Formerly Delhi College of Engineering)* for her support and encouragement.

I am also thankful to all the faculty members and Staff of the Computer Engineering Department for their valuable support.

Last but not the least; I am grateful to my Family members for their unconditional support and motivation during this work.

Praveen Kumar
Roll No. 10/CTA/09, University Roll No. 8549
Department of Computer Engineering
Delhi Technological University
(FormerlyDelhi College of Engineering)
Delhi-110042

CONTENTS

List of Figures and Tables

Abstract

If we have some query pertaining to *Inorganic chemistry*, we either search through some pages of a text book, or we do *googling*. Even after going through so many pages of search results, we may not get what we want. So there a need for a system which can precisely and compactly, answer such queries.

Knowledge formulation is done on inorganic chemistry (at 1st year science level) and a standard set of questions were taken for testing.

We have successfully implemented a system which answers to the questions related to definitions, chemical reactions and their balancing, pH calculation of buffer solutions, redox reactions and many more concepts, with GUI interface for the query.

All the questions can be asked in normal English Language .

Chapter 1: INTRODUCTION

In this chapter the need of a learning module in inorganic chemistry is highlighted which leads to the motivation behind this project. The objective is clearly defined and the layout of the thesis is outlined.

1.1 Need

At present if we desire to get some information related to a specific subject in scientific domain, we either search through some pages of textbook or use some search engine for the purpose. Even after doing so, it is not always guaranteed that we will get the desired information. So, we felt that there is a need for a system that can precisely answer such queries.

1.2Motivation

Present search engines have the following problems associated with them.

- Search engines require that a specific text containing the answer to user's query must reside in the searched corpus.
- Document containing the correct answer must reside fairly high among the ranked list of documents it returns given the user's specified keywords.
- The user needs to scan the document for the appropriate passage.

In scientific domain, we always desire to have a system which can compactly and precisely answer our queries. For that, we need to represent our knowledge in some structure to retrieve information from that structure along with a system that can understand what the user query is. We wish to provide total flexibility to the user as far as the way of writing queries is considered.

1.3 Existing Work

1.3.1 AURA

AURA (Automated User-Centered Reasoning and Acquisition System) is a knowledge formulation tool for systematic knowledge like Physics, Chemistry and Biology. The goal of this project is to build a generic knowledge acquisition capability for the science streams.

Using AURA system, the scientists will be able to formulate their knowledge in the three science domains, and the high school students will be able to pose questions and get user appropriate explanations. A lot of work has already been done in the field of biology. It provides a graphical interface for knowledge capturing so that an expert can enhance the system by introducing the system

of the knowledge it doesn't have.We have increased the flexibility as the user can give the questions in natural English Languge.

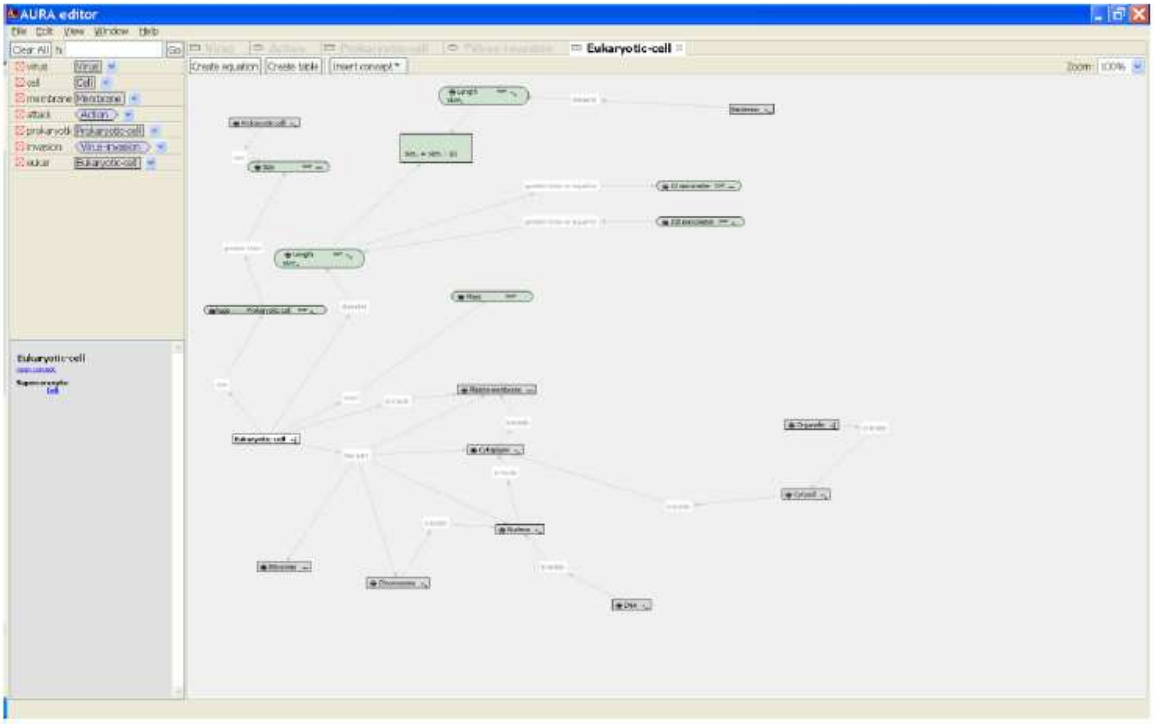


Figure 1.1: AURA Snapshot

This diagram shows knowledge formulation for eukaryotic cell in AURA editor

1.3.2 Wolfram Alpha

Wolfram Alpha, released by Stephen Wolfram on May 15, 2009 is a computational knowledge engine which aims to bring expert-level knowledge and capabilities to the broadest possible range of people spanning all professions and education levels. Their goal is to accept completely free-form input, and to serve as a knowledge engine that generates powerful results and presents them with maximum clarity. It provides specific answers rather than providing a list of documents or web pages that might contain the answer as search engines do.

Wolfram Alpha's long-term goal is to make all systematic knowledge immediately computable and accessible to everyone. But presently, it fins difficult to answer those queries pertaining to subjects that are to be represented by numbers.

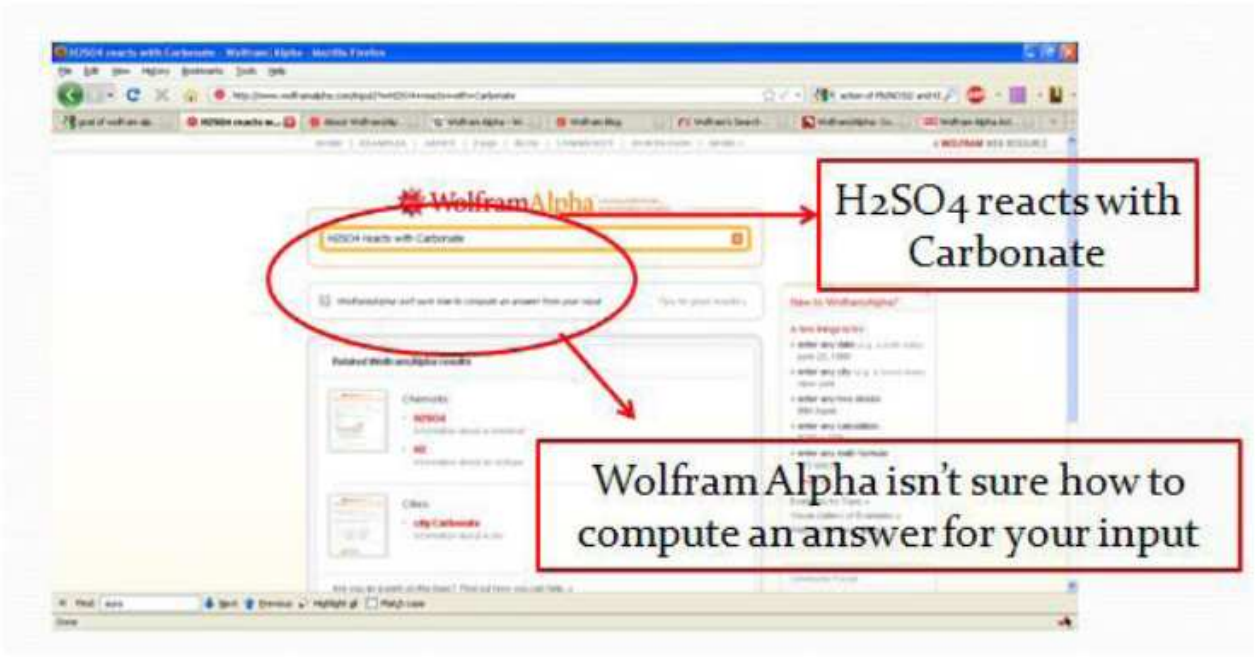


Figure1.2: Screenshot of Wolfram Alpha

1.4 OBJECTIVE

Our main objectives are as follows:

- To represent our knowledge in structured form so that we are able to get semantically correct results of our queries.
- Generate custom answers to every question, i.e. to make a system which is capable of answering questions for which the text does not exist in some document.
- Produce user and domain appropriate justifications for the answer.

For this purpose, we represent the knowledge of a particular domain in structured form and build a system which can capture domain knowledge and give semantically correct results of our queries. We have chosen 'inorganic chemistry' at the level of 1st year science as scientific domain for representation of knowledge. Inorganic chemistry being structured and hierarchical serves as a general platform for various domains. It has facts, rules and some exceptions to them which make it a perfect choice for implementation.

1.5 LAYOUT OF THE THESIS

Initial part explains about the knowledge representation schemes , how the knowledge is formulated , the later part explains how exactly the given question is parsed to form the CUL question and about various answer checking mechanisms.

- **Chapter 2 : Knowledge representation Schemes Used**
  This chapter explains about various knowledge representation schemes and used inorganic chemistry along with the overall design of the system.

- **Chapter 3: Description Of Grammar For various Types Of Questions**
  This part is used to explain the various parsing grammars along with the examples that are deployed in understanding the given natural English language question.

- **Chapter 4 : Answering Questions**
  This section deals with the various ways in which the options can be provided and also how to resolve those options that will assist in finding out the correct answer . So , explanation is provided about various answer checking techniques. Seven have been used in our system.

- **Chapter 5 : Implementation**
  This chapter main aim is to elaborate the software implementation of our system.

- **Chapter 6 : Results and Discussions**
  This chapter deals about the conclusions, the problems faced in understanding the question and future work that needs to be done.

CHAPTER 2: KNOWLEDGE REPRESENTATION SCHEMES USED

2.1. Predicate Logic

The most important knowledge representation language is arguably predicate logic. Predicate logic allows us to represent fairly complex facts about the world, and to derive new facts in a way that guarantees that, if the initial facts were true then so are the conclusions. It is a well understood formal language, with well-defined syntax, semantics and rules of inference. Predicate logic introduces two new quantification symbols to propositional logic:
Universal quantification: ∀ x. ("for all x, it is the case that ...");
Existential quantification: ∃ x. ("there exists an x, such that ...")
Predicate calculus is built up from *atomic sentences*. Atomic sentences consist of a predicate name followed by a number of arguments. These arguments may be any *term*. Terms may be:

- Constant symbols: such as “praveen”
- Variable symbols: such as “X”
- Function expressions: such as “friend(praveen, vipin)”

Sentences are formed by combining atomic sentences with logical connectives, eg.

- ∀ X ∃ Y (person(X) ^ person(Y) ^ friends(X,Y)) → likes(X,Y)

i.e. for every person X, there exists a person Y who is a friend of X and X likes Y.

- ∀ X (person(X)) → ∃ Y likes(X,Y)

i.e., every person has something that they love.

2.2 Frame Logic

Frame Logic is a knowledge representation and ontology language which combines the declarative semantics and expressiveness of deductive database languages with the rich data modeling capabilities supported by the object oriented data model. The basic idea behind F-logic is to consider complex data types as in object-oriented database, combine them with logic and use the result as a programming language.
F-Logic consists of

- Object Identifiers: Eatables, fruits, etc.
- Variables: X, Y, Z, etc.
- Predicates: isSweet, isSour, etc.

F-Logic allows specifying signature information and organizing the structures in hierarchical fashion. Each frame has a name and has slots - properties/attributes of the entity, having some name and value. In the higher levels of the frame hierarchy, typical knowledge about the class is

stored where as in the lower levels; the value in a slot may be a specific value, to overwrite the value which would otherwise be inherited from a higher frame.

1. We have used Predicate logic for implementing rules and for handling exceptions.

2. We have also used Frame logic for storing various data like Elements, Compounds, and Radicals etc.

2.3 DESCRIPTION OF CHEMISTRY USED

The main focus is on Inorganic chemistry which is the branch of chemistry concerned with the properties and behavior of inorganic compounds. This field covers all chemical compounds except the myriad organic compounds

Many inorganic compounds are salts, consisting of cations and anions joined by ionic bonding. Examples of salts are magnesium chloride MgCl2, which consists of magnesium cations Mg2+ and chloride anions Cl−; or sodium oxide Na2O, which consists of sodium cations Na+ and oxide anions O2−. In any salt, the proportions of the ions are such that the electric charges cancel out, so that the bulk compound is electrically neutral. The ions are described by their oxidation state and their ease of formation can be inferred from the ionization potential (for cations) or from the electron affinity (anions) of the parent elements.

Important classes of inorganic salts are the oxides, the carbonates, the sulfates and the halides. Many inorganic compounds are characterized by high melting points. Inorganic salts typically are poor conductors in the solid state. Another important feature is their solubility in e.g. water (see: solubility chart), and ease of crystallization. Where some salts (e.g. NaCl) are very soluble in water, others (e.g. SiO2) are not.
The simplest inorganic reaction is double displacement when in mixing of two salts the ions are swapped without a change in oxidation state.

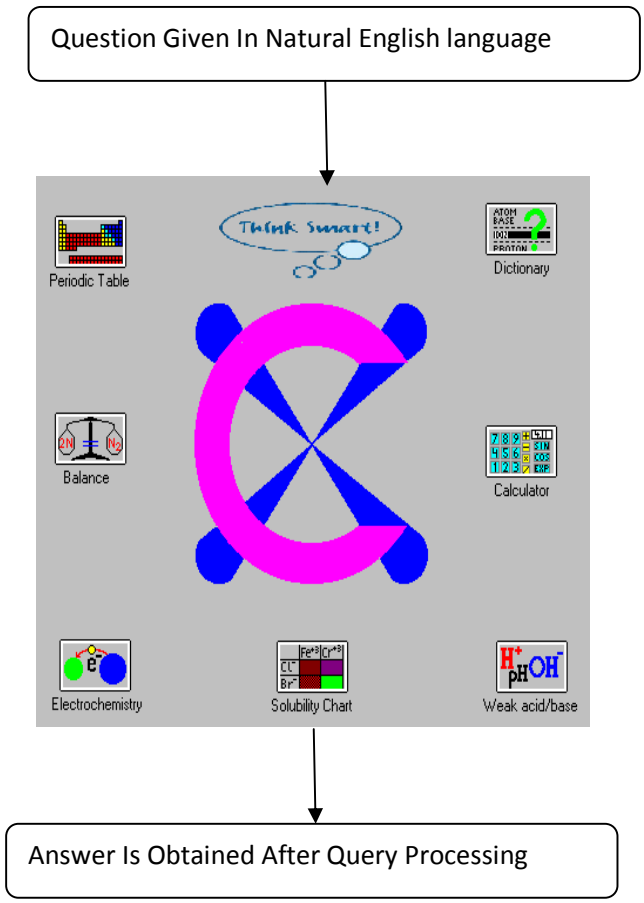
In redox reactions one reactant, the oxidant, lowers its oxidation state and another reactant, the reductant, has its oxidation state increased. The net result is an exchange of electrons. Electron exchange can occur indirectly as well, e.g. in batteries, a key concept in electrochemistry.

When one reactant contains hydrogen atoms, a reaction can take place by exchanging protons in acid-base chemistry. In a more general definition, an acid can be any chemical species capable of binding to electron pairs is called a Lewis acid; conversely any molecule that tends to donate an electron pair is referred to as a Lewis base. As a refinement of acid-base interactions, the HSAB theory takes into account polarizability and size of ions.

2.4 Overall Design of the System

Handling of a question involves three parts.

- The first part is to understand what is being asked in the Natural language Query (English) and to generate the corresponding query in Computer Understandable Language (CUL).
- The second part is to work out the correct answer from the knowledge base, for the query made available in CUL form.
- The third part is to represent the result in natural language form, supported by an explanation part, as to how the result is arrived at by the system.



2.4.1 Question Formulation

This is the very important module of the overall system, for this pre-processing is required. This module takes in the question given in natural English Language and thereby converts it to the CUL query .For this purpose the questions are divided into various categories like :

a. What

Example 1:
What is the pH of a 0.05 M solution of hypochlorous acid? (For HOCl, Ka = 0.000000038)
a. 8
b. 10
c. Between 7 and 10
d. Between 4 and 7

Example 2:
What is the empirical formula for a compound that contains 17.34 % hydrogen and 82.66 % carbon?
a. C5H
b. C2H5
c. CH3
d. CH2

b. Which

Example 1:
Which of the following is in correct order of increasing acidity?
a. HClO, HClO4, HClO2, HClO3
b. HClO4, HClO3, HClO2, HClO
c. HClO4, HClO, HClO2, HClO3
d. HClO, HClO2, HClO3, HClO4

Example2:
Which of the following correctly represents the balanced chemical reaction between aluminium and S8?
a. Al + S8 -> 8Al2S3
b. 12Al + S8 -> Al3S2 ,
c. 16Al + 3S8 -> 8Al2S3
d. 4Al + S8 -> 4AlS2

c. When

Example 1:
When lithium metal is reacted with nitrogen gas, under proper conditions, the product is
a. LiN
b. Li2N
c. Li3N
d. LiN3

Example 2.:
When calcium carbonate is heated it decomposes forming :
a. Solid Ca and CO2 gas
b. Gaseous CaCO and CO2 gas
c. Solid CaO and CO2 gas
d. Gaseous Ca and CO2

Article

Example 1
The concentration of H+ ions in 0.075 M solution acetic acid is ? (For acetic acid, Ka = 0.00018)
a. 1.16 x 10-3M
b. 1.35 x 10-6M
c. 2.4 x 10-4M
d. 0.25M

Example 2
The pOH of a solution containing 2.250 g of LiOH , in 250.0 mL of solution is :
a. 0.425
b. 13.58
c. 0.376
d. 13.62

a. Fact First

Example 1:
Sodium iodize is used in air bags to inflate it. the products of the decomposition reaction are
a. sodium and nitrogen
b. sodium and ammonia
c. ammonia and nitrogen
d. Na and O2

Example 2:
A solution nickel nitrate and sodium hydroxide mixed . Which of the following statements is true?
a. A precipitate will not form
b. A precipitate of sodium nitrate will be produced.
c. Nickel hydroxide and sodium nitrate will precipitate.
d. Nickel hydroxide will precipitate.

More questions can be found in the appendix

The different category of questions are being treated in a different way depending on their respective grammars The Think smart as shown in the diagram refers to this aspect of the module. To form a proper query, we need to separate out the relevant data from the question and

the entity which the question wants to know. These are obtained from the sequence of states through which the words in the question passes. The combination of these two results in the formation of the CUL query.
The various classes that help us in understanding the query are as follows:

CLASS	Observations
reaction	Fused , mixed , reaction, react, passed , added , heated ,heat ,decomposition, etc.
require	Definition , property, weight, products, liquid, gas, Metal, species, pH,solubility,normality,etc.
property	Colorless, soluble, precipitate, weak electrolyte, etc.
compare	Strongest, most, least ,highest, lowest, weakest, etc.
check	True, false, wrong, correct, etc.
Numerical kind	pH, Pka,k1,K2,Ka,volume,concentration,molar, normality, solubility, molar.

2.4.2 Query Evaluation

This refers to the processing of the CUL query as obtained from the above step. For this purpose, The whole database and various rules are looked through and checked for applicability to the given query.
If no such rule or information exists, it results in the exception else we gets the required value after processing.
At the present the queries dealing with these areas of chemistry can be answered:

Type Of Question	Corresponding Sample Query
Chemical Reactions	Reaction : HCl + NaOH
pH calculation	Calculate : pH , CH3COOH ,0.001
Equilibrium constant	Calculate : Ka, HCl , 0.3, 2.63
Oxidation Number	Calculate : oxno, Pb, Pb(NO3)2
Empirical Formula	empirical:H2 ,17.34 ,C ,82.66
Finding net ionic equation	Ionic : Pb(NO3)2 + 2 NaCl -> PbCl2 + 2NaNO3
Balancing of reactions	Balance : NaCl + H2O
Solubility	Soluble : BaSO4
Formula of chemical	Formula : sodium oxalate
Spectator ions in a reaction	Act_spectator : Ba(NO3)2 + Na2SO4 -> NaNO3 + BaSO4
Ions in a compound	Ions : CH3COONa
Electrolyte identification	Electrolyte : NaOH
Lewis Acids	Act_lewis : FeBr3 + Br - -> FeBr4
Amphoteric oxide	Amphoteric : CaO

2.4.3 Query Answering

The value obtained from above module (Query evaluation) is checked against each of the options , the option that matches the obtained requirement is resulted as correct . There can be more than one correct answer.

CHAPTER 3: DESCRIPTION OF GRAMMAR FOR VARIOUS TYPES OF QUESTIONS

As we all know that whenever we try to understand a question, human mind works in a way that it tries to figure out meaning of it from some of the relevant words (keywords) , the similar phenomena is being implemented in the system for various questions , thus , many words like is, are , the , etc. are generally ignored because they hardly contributes in understanding .

For our system , the grammar that we have used is Context Free Grammar(CFG) which is a type 2 Grammar . Description of CFG is given in the appendix.

The parsing technique used is *adaptive Shift Reduce Parsing.* We have called our technique as adaptive because based on the category of question, the level until which the Shift Reduce parsing is deployed for the purpose of states exploration varies.

3.1 WHAT

These questions are divided mainly in two parts:

a. The entity about which question wants to know about.

b. The data given that will assist figuring out that entity.

.

3.1.1 GRAMMAR

What -> Require Separator Data
Require -> product/pH / ions /oxidation No. / formula , etc .
Separator -> of/for/when/ between /to/in
Data ->D1/D2/D3/D4/_____/ Dn
D1-> chemical operation chemical Data
D2 -> Nv Nkind chemical Data
D3 -> Nkind chemical Nv Data
D4 -> chemical operation Data
D5 -> chemical Data
D6->operation chemical chemical Data
D7-> chemical operation Data
D8-> chemical chemical operation Data
D9 -> chemical chemical Data
D10 -> Type_operation chemical Data
D11 -> NKind NValue Data

…..
Dn -> null

Here Nv denotes numerical value , Nkind denotes numerical kind

There are two techniques of finding required data. First is the technique that the information extracted in require directly gives what is the required entity , but this is not exactly the case , for example , in require we may have something like *most likely products* **,** from this we have to separate out the most important word i.e. products and then correspondingly form the query .

Example

Question is
What are the products formed when lithium is reacted with nitrogen

a. LiN
b. Li3N
c. LiN3
d. Li2N

Formation of query:

Relevant keywords along with their order are *products formed* and lithium *reacted nitrogen* (words like of, is, when , between ,etc. serve only as state classifier) .

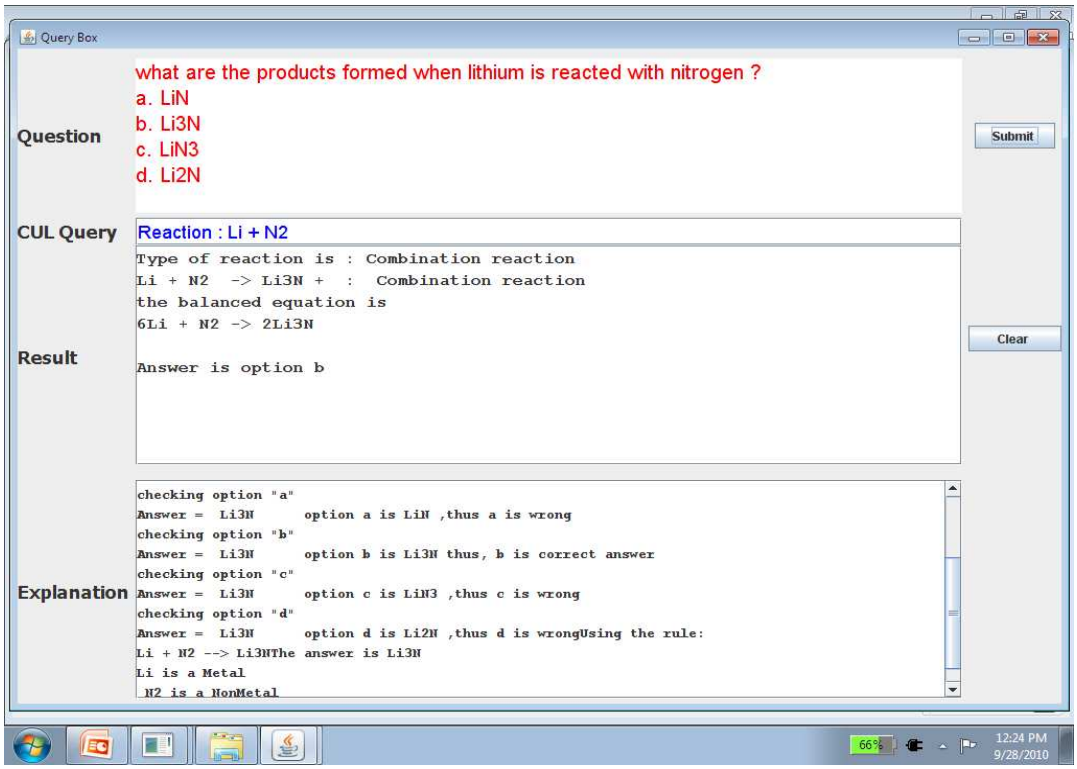
Require = products formed, Chemical1 = lithium, chemical2 = nitrogen, operation = reacted.

Thus states are -> Products formed chemical operation chemical
-> Products formed D1
-> Products D1
-> require D1

Thus, query formed = Products: Li + N2

Answer checking is quite simple in *what* kind of questions, because the query is formed from the question only and no help of answer is required for query formation. Thus, we get the correct answer after query processing is terminated, and that answer is used to check which option is correct.

SCREENSHOT OF THE ABOVE EXAMPLE



3.2 WHICH

The formation of questions of *which* type as compared to above case . It is because the number and kind of questions that can be asked are numerous. To handle these kind of questions, the following steps are taken .

We divide the question in two parts , the required entity and the property it needs to satisfy(generally in this case).The way of dividing question in these two is very similar as above .

a. After which, we need to extract out the given information from the property. This infect is quite complex due to its great variety and for each variety ,different kind of parsing structure will be formed.

3.2.1 GRAMMAR

Which -> following / following require property/ following property/require property/property
Require -> gas , solution , oxides , statements ions , compounds , etc .
Property -> P1 / P2 / P3 / P4 /____/Pn/prop1 .
P1-> compare prop1
Compare -> strongest, weakest, smallest ,lowest etc .
Prop1 -> conductivity, solution, Bronsted acid , base , atomic radius ,etc.
P2-> not prop1 /non electrolyte
P3 -> check prop2
Check -> true /false/correct/incorrect/not true/cannot
Prop2 -> order of increasing acidity / Data/ null /
Prop1 -> it is found based on occurrence position of words
P4 -> amphoteric .

Data is described later .

Here it is very difficult to restrict prop1 because it can be of unlimited varieties , some examples are

a .Amphoteric in character.
b.Amphoteric oxide.
c. Produce a precipitate.
d.Correctly represents balanced chemical reaction between chemical1 and chemical2.
e.Contain unpaired electron.
f.Produce hard water.
g.Oxidizing behavior of chemical.
h.React with chemical.
i.Disproprotionate in cold alkali.
j.produce H2.
k.Insoluble in water.

These are only some of the few examples that can be formed .Thus , the understanding of questions in *which* becomes very very complicated . Now , in most of the cases , we again need to extract information from the given property .

Property -> X1/X2/X3/X4/X5_ _ _ _ _/Xn / Data

X1 -> Balanced chemical equation Data
X2 -> amphoteric
X3 -> produce gas Data
X4-> not liberate chemical Data

Data is same as used in case of which

Data ->D1/D2/D3/D4/_____/ Dn
D1-> chemical operation chemical Data
D2 -> Nv Nkind chemical Data
D3 -> Nkind chemical Nv Data
D4 -> chemical operation Data
D5 -> chemical Data
D6->operation chemical chemical Data
D7-> chemical operation Data
D8-> chemical chemical operation Data
D9 -> chemical chemical Data
D10 -> Type_operation chemical Data
.
.
.
Dn -> **null**

Now after carrying out the parsing of the question using the above grammar , we have all the required information to form the CUL query .

The question can be divided into various types:

a .**Comparison->** In this the comparison over some property between the given option is carried out.

b. **Validation->** In these type the correctness of the various options is being examined.

c. **Identification->** In this the given options are checked against whether they satisfy the given property or not.

As it is clear, that broadly speaking, as it is necessary to use the options for CUL formation, the difficulty increases manifolds.

Example:

Question is

Which of the following compounds is insoluble in water?

a. Pb(NO3)2

b. Li2CO3

c. Ba(OH)2

d. BaSO4

Formation of query:

Relevant keywords along with their order are following Compounds insoluble in water

Thus states are -> following compounds insoluble in water

-> following compounds prop1

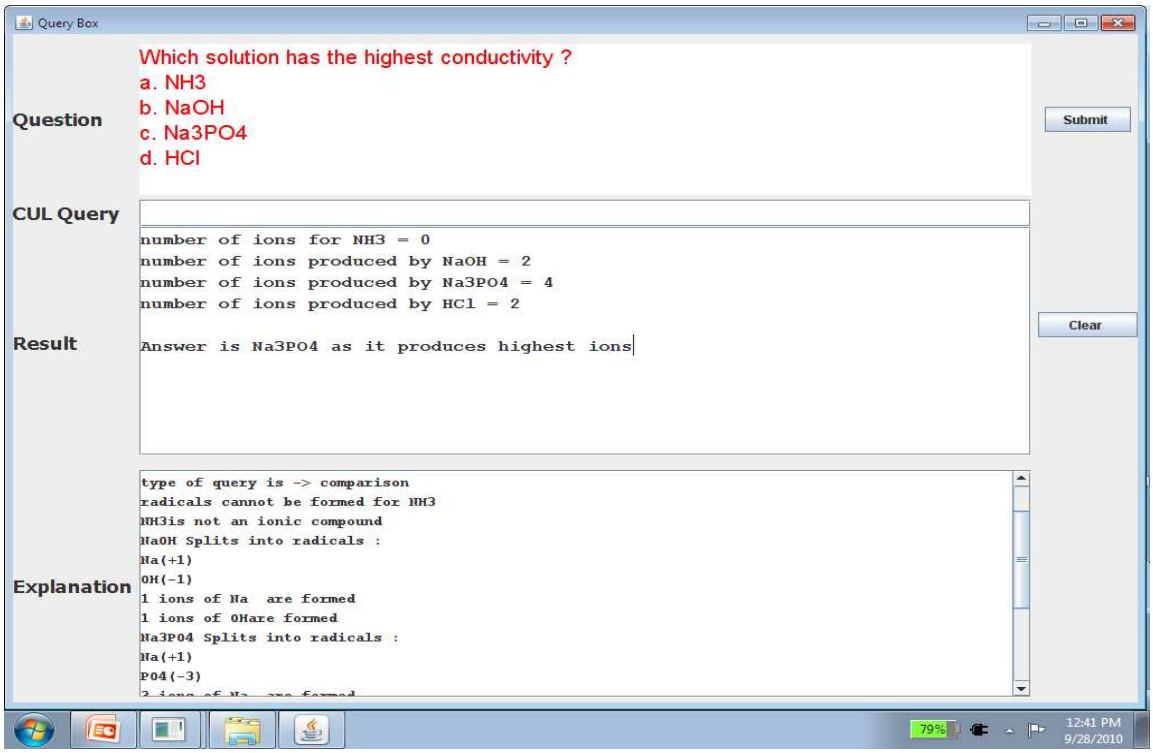
-> following require prop1

Type of query -> IDENTIFICATION

➔ Query Formed : soluble : BaSO4 ; soluble : Pb(NO3)2 ;

soluble: Ba(OH)2 ;soluble : Li2CO3

SCREENSHOT OF THE ABOVE EXAMPLE



3.3 ARTICLE

These questions can begin in one of the two ways, either from *The* or from *A or An*. I am using different structures for these.

3.3.1 GRAMMAR

When beginning with The

The -> Require separator Data / Data require / Data

Separators -> of / for / in / is

The Data grammar is same as used in the previous cases. Sometimes it may happen that requirement is not defined explicitly and we have to infer from the Data given only, that what may be the requirement, so it leads to lot of ambiguity. So in these cases, a look up of the answers becomes necessary for the understanding purpose

The requirement in these cases is generally found on the basis of positions of occurrence of words. In most of the cases, separators words are being utilized for classification of query into require and data parts.

When beginning with A/An

A/An -> Data Require / Require Separator Data

For first case, we are treating it differently from fact first types (as explained later) mainly for simplicity purposes . Division of these two types into different structures reduces the complexity because treatment of these as same increases complexity a lot, so it's better to treat them as having different structures. Also to understand what the exact requirement is, we need to consider the (, or.) position, the position of the keyword **The and** then proceed in a similar manner as from the grammar of **The.** For second case, the requirement is mainly found based on the position of separator.

Example:

Question is

A sample of vinegar having acetic acid (Ka = 0.000018) has pH 2.90 , the molar concentration of acetic acid in vinegar is

a. 0.088

b. 0.890

c. 0.126

d. 0.014

Formation of Query

Relevant keywords along with their order of occurrence are : *vinegar , acetic acid , Ka , 0.000018 , pH 2.90 , the molar concentration ,* and *acetic acid* .

Thus, states are -> vinegar acetic acid Ka 0.000018 pH 2.90 the molar concentration acetic acid

->chemical1 chemical2 NKind1 Nvalue1 NKind1 NValue1 Nkind chemical2

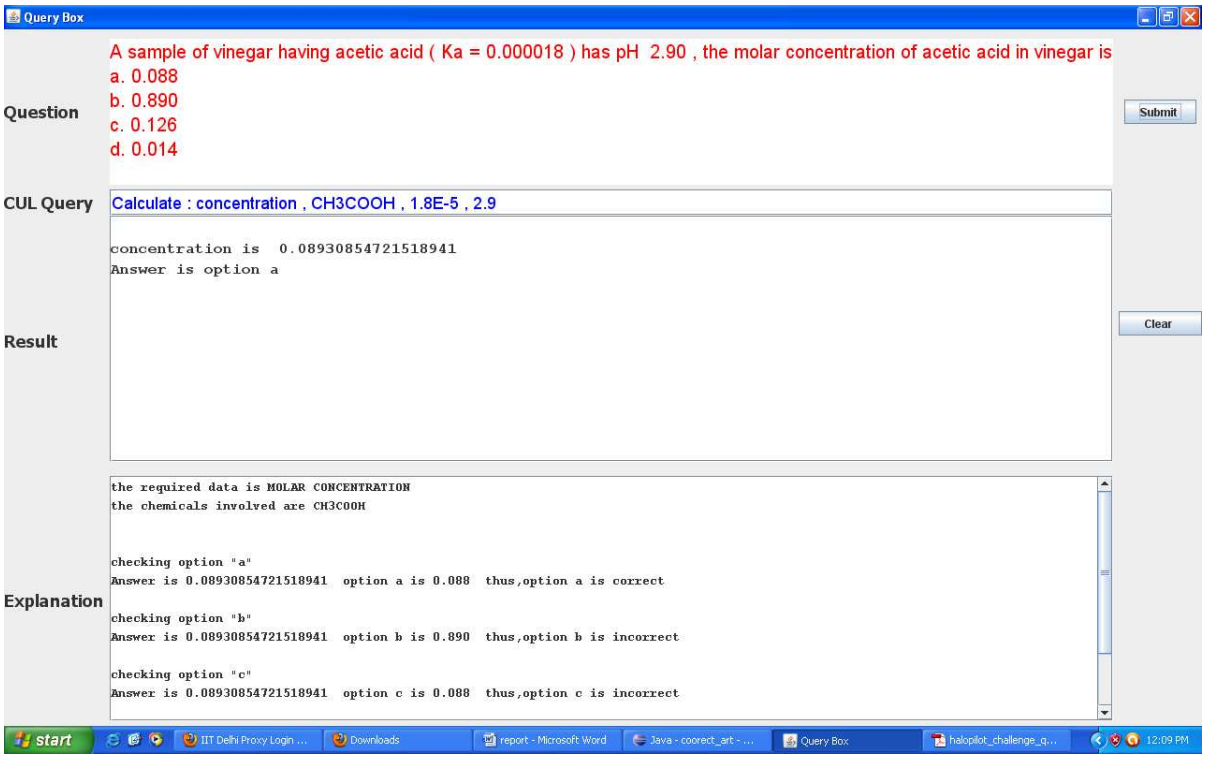
-> Data9 Data11 Data11 require

-> Data require

The intermediate states Data9, Data 11 & Data 11 are very important because they contain the crux of the information contained in the question.

Query Formed : Calculate : concentration , CH3COOH , 1.8E-5 , 2.9

SCREENSHOT OF THE ABOVE EXAMPLE



3.4. WHEN

This is the other class in which the type of questions can be divided. Here the questions are divided mainly in these two parts:

Action->The action that is being performed on some chemical entities.

Require->The required data.

In this mainly the **(,)** acts as separator between the action and require. The action basically shows the operations that are being performed, whose results we need to find out the obtained requirement. Finding the exact requirement is a tough job because mainly there is a whole sentence instead of direct requirement. So for this purpose, we have our set of requirements like number of moles, amount of chemical, balanced equation, etc . Along with the well defined requirements. These requirements are searched within the obtained require, and if any of these are present, we assume it to be the main requirement.

3.4.1 GRAMMAR

When -> Action require / Action

Action may be of various kinds and these are inferred on basis of the position of occurrence, for example some of actions are :

a. ChemicalA is reacted with ChemicalB.

b. reaction between ChemicalA and chemical.

c. Some value of ChemicalA is mixed with some value of chemical.

d. Operation on chemical.

Sometimes, as in the previous case , we may not be given the requirement explicitly , but the action in itself is so well defined , that its not a tough task to infer the correct requirement for that question.

Example

Question is

When calcium carbonate is heated it decomposes forming:
a. Solid Ca and CO2 gas
b. Gaseous CaCO and CO2 gas
c. Solid CaO and CO2 gas
d. Gaseous Ca and CO2

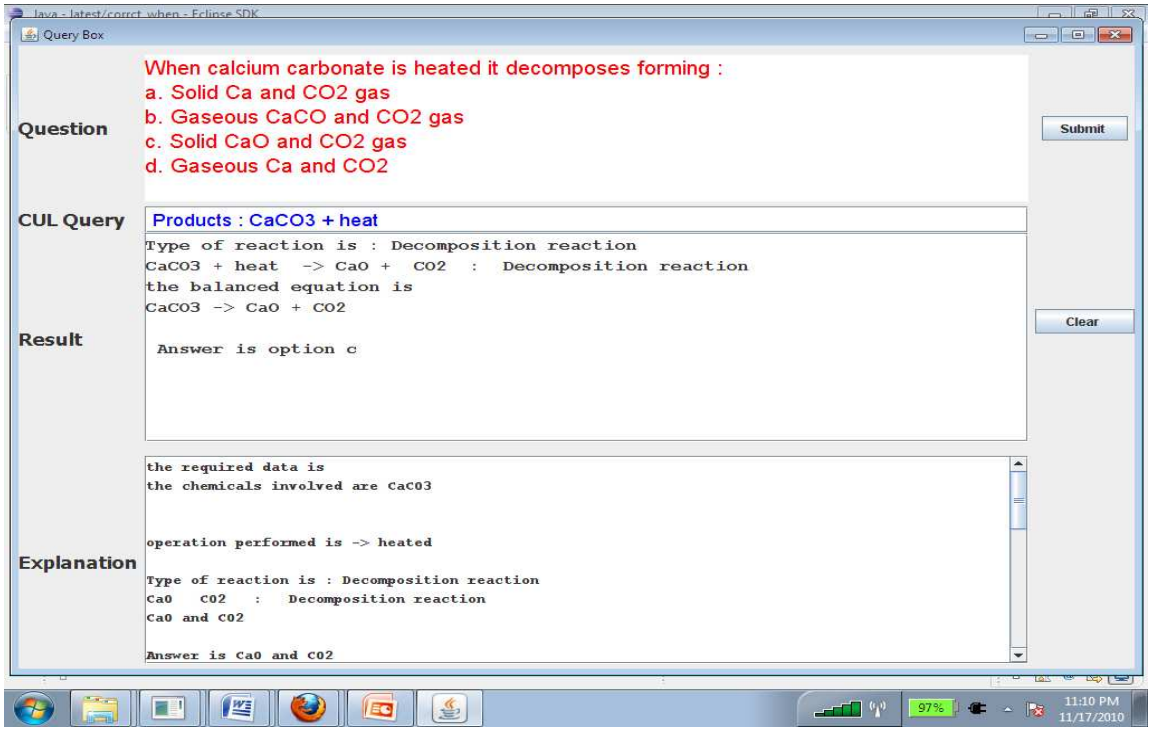
The relevant keywords with their order of occurrence are : calcium carbonate , heated , decomposes .

Thus , states are : Calcium carbonate heated decomposes

: Chemical operation

As we can clearly see that there is no explicit mention of what exactly we want , but the sentence is well framed to reach to conclusion that products are required .

Thus, query Formed: **Products: CaCO3 + heat**



3.5 FACT FIRST

This is perhaps the most difficult portion of query understanding. Because, it is very difficult to know from first few words , what the question exactly wants to know , as in contrast to the previous cases , where the identification of questions is much more simpler . In these cases , we are assuming that the question will consist of two sentences .The first sentence will be the fact first sentence i.e. it will contain the majority of the information required for the CUL query formation .The second part will actually describe the kind of question i.e. , which , what , how or when , etc . Based on the word occurring immediately after the full stop

3.5.1 GRAMMAR.

S1 -> Whole Reaction Data / Data

Here the data will be the same as in the previous defined cases.

S2 -> which/when/what/article…..

For each of *which* , when or what kind of sentence occurring after the full stop , the way to proceed is the same as explained in their respective grammar .

The main difficulty lies in combining the information obtained in a well defined way to form the query. The main reason for this is that the information from the two sentences are explored independent of each other and thus, combining the relevant information is an headache. Also in this case , even with a slight modification in the question forms , the way of treatment varies by a great extent thereby increasing the overhead manifolds and posing a great problem to the implementation .

EXAMPLE

Question is

The pKa of HNO2 is 3.37 . The pH of a 0.01 M aqueous solution of HNO2 is:
a. 5.37
b. 2.00
c. 1.69
d. 0.69

Sentence 1 = The pKa of HNO2 is 3.37.

Sentence 2 = pH of a 0.01 M aqueous solution of HNO2 is:

Sentence 1 analysis :

The -> pKa HNO2 3.37

-> NKind HNO2 NValue

->Nkind chemical1 Nvalue

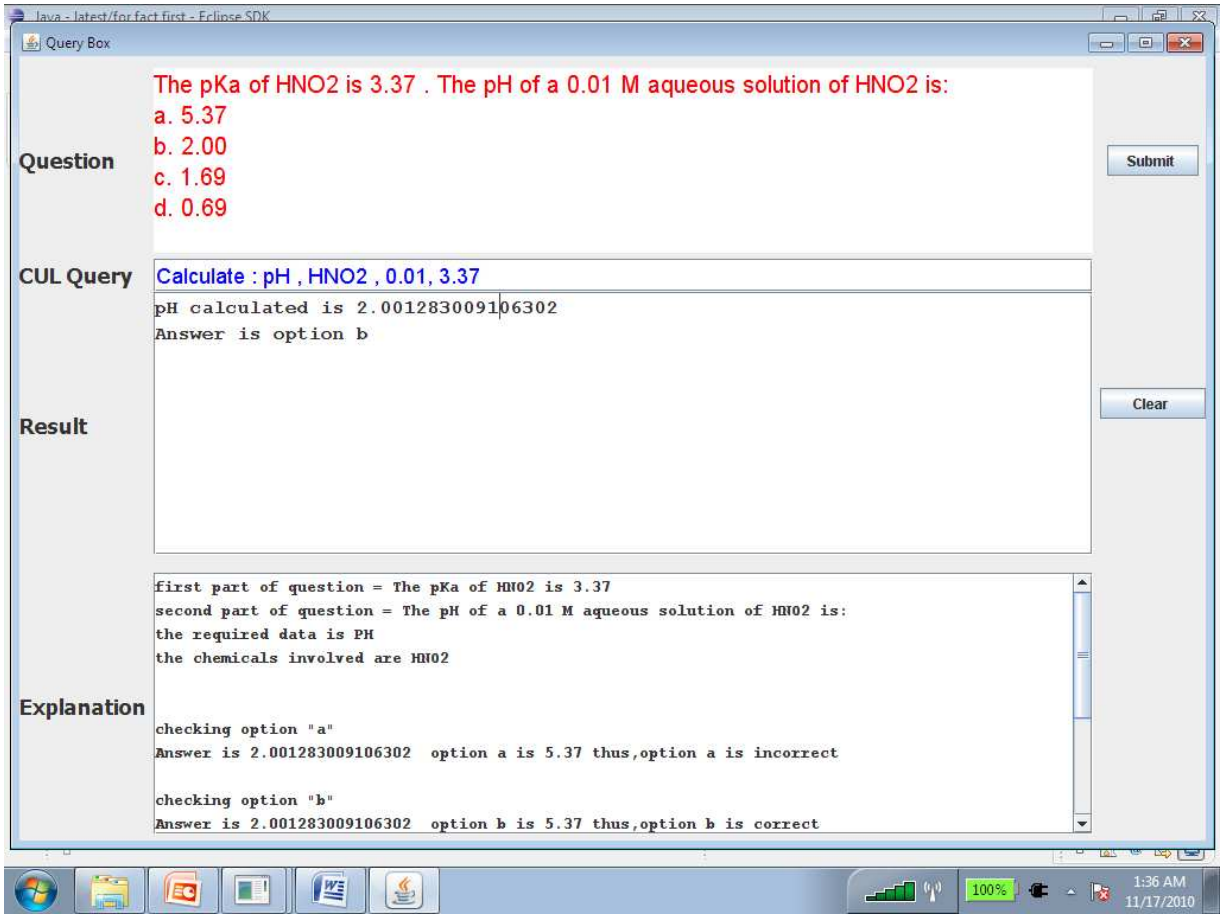
Sentence 2 analysis :

The -> pH 0.001 M solution HNO2

-> NKind Nvalue NKind chemical1

Here as the word of is used after pH , and no immediate value is being defined for it , we assume that this is the entity whose value we need to determine .

Thus , we form a query from the information present in this sentence and later on, before evaluating the other required information is provided from the first sentence parsed structure .



CHAPTER 4: ANSWER QUESTIONS

The objective of this phase is to check the correctness of the various options and then give the correct answer along with the explanation. As the option can be in various types, thus simple String Matching is not going to work in all the cases . So the techniques of answer checking are as follows:

4.1 Simple String Matching :

This refers to the direct checking of the obtained answer with the options one by one for equality. For this purpose the function *equals* as defined in the String class of java is being utilized.

For example :

the pH of a 0.1 M solution of HCl is
a. 1.3
b. 2.4
c. 5.1
d. 1.0

Here the answer obtained is 1.0 which is directly checked giving the result that Option d is correct.

4.2 Checking in a range of Values:

In this case, direct string matching cannot work because the correct answer has to be semantically inferred from the options. We need to find the correct value and identify that it falls in which of the given range.

Example:

The pH of a 0.001 M solution of HCl is

a. less than 2
b. between 2 and 4
c. between 5 and 7
d. more than 8

Here the correct answer is 3, thus the correct option is b, but the checking of options in this case is very complex. If between appears, then the two numbers are identified and then the range is checked. If less is found, the n the answer should be less than the number in the option and vice versa for the more option.

4.3 Checking with the closest possible:

For numerical questions, it is not always necessary that the correct option is same as the obtained answer because the options may be displayed after rounding off the correct answer. In such cases, we need to check out which option is closest , this is done by finding out the difference between the option and given answer, and correct option is min(||option – answer||) where i=a, b, c, d.

4.4 Chemical Names Given:

In these, the answer that we obtain is the chemical formula but the options contain the chemical names of compounds and not the formula. In such cases, we extract the formula of the compound in the option from our database and then simple string matching as previously explained is being done.

Example:

What is the result of the reaction between water and SO3 ?

a. Sulphorous acid.

b. Sulphuric acid.

c. No reaction.

d. Hydrogen sulphide and oxygen.

In this case , the answer obtained is H2SO4 which is the chemical formula for sulphuric acid . In these, each option is checked , if it is chemical formula then directly string matching is done else chemical formula is obtained for the purpose . Thus the final answer is *option b.*

4.5 More than 1 chemical in the products obtained: In these cases, first we need to find out that what are the chemicals present in each of the option, every compound's chemical formula is explored for this objective. The new check that whether the chemicals in the option and the answer are same and the number of matching's should be same as the number of chemicals in the answer.

Example

What are the products formed when ammonium nitrate is heated ?
a. nitrous oxide and water
b. nitric oxide and hydrogen
c. nitrogen dioxide and hydrogen
d. nitrogen dioxide and water

In this case the answer obtained is N2O and H2O(i.e. nitrous oxide and water). Thus we have to find that what the chemicals in the option are. If the number of matching's reaches 2, the nthat option will be the correct answer. In this case, option a is right. In other cases, number of matching's is either 1 or 0 .

4.6 Comparison between options needs to be performed: In such case, we have to make comparisons between the options for the required property like conductivity, strength, etc. Thus, the option which results maximum or minimum based on what is asked in the question is the correct answer.

Example:
Which solution has the highest conductivity?
a. NH3
b. NaOH
c. Na3PO4
d. HCl

In this case, we have to find conductivity for each based on the number of ions formed after complete dissociation. Thus as Na3PO4 results in maximum number of *ions (4)* in the given options thus , correct answer will be option c.

4.7 Option satisfying some property: This type of checking generally arises when the question belongs to *which* category. In these, the options are also utilized in forming query and then the result obtained is used to check whether the given option satisfies the required property or not.

Example:

Which of the following compounds will produce a gas when HCl is added to the solid compound?
a. Ba(OH)2
b. CaCO3
c. CuSO4
d. Na3PO4

In this question, first the reaction between each option and HCl is performed, and then the products are checked whether any of them is a gas. Here as CaCO3 on reaction produces CO2 which is gas, hence the correct option is b.

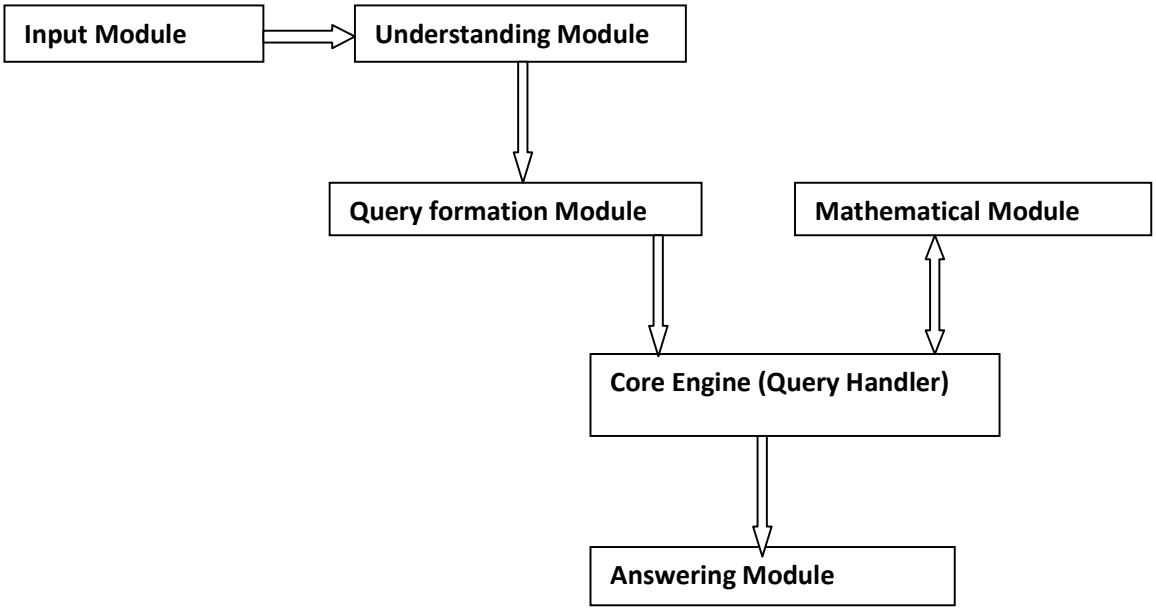
CHAPTER 5: IMPLEMENTATION AND RESULTS

The rule based system has been implemented on Eclipse platform. The system uses JForm Designer for designing the user-interface. For fast access to the database, the system uses java Hash Map. The advantage of using Hash Map is that it can handle Chemistry database of any size. The system can access data in O(1) time, thus making it very fast for handling large number of elements and compounds. To answer queries pertaining to chemical reactions, the system contains number of rules.
We are using F-logic to specify classes like element class, molecule class, compound class, chemical reactions class etc. We have implemented our project in Java and JForm Designer for graphical user interface. We have used **HashMap <key, Value>** - inbuilt structure for hashing, for fast access to elements.

The overall scheme consists of following modules:

- Input module
- Understanding Module
- Query Formation module
- Core-engine interface
- Mathematical module
- Answering module

These modules work together to get the final output of the query. Here is a pictorial representation of the various modules used in this system and their inter-dependence



5.1 Input Module

The Objective of this module is to get the English language question from the user of the system .The question must be given along with four options a, b, c, and d else exception will occur. It breaks the questions into tokens so as to separate out various options from the question, store them in their respective variables. This module also performs the function of deciding the type of question based on its structure and accordingly the respective class is called.

5.2 Understanding Module

The main objective of this module, as the name suggests, is to understand the question. Understanding means finding out what are the various entities as given in the question and what entity we need to find out. Based on the type of question, the understanding module is different for each of them .Thus, this section utilizes the various grammars that are defined previously.

5.3 Query Formation Module

The objective of this module is to form the CUL query based on the information extracted above. The structure of the final CUL query formed is as follows:

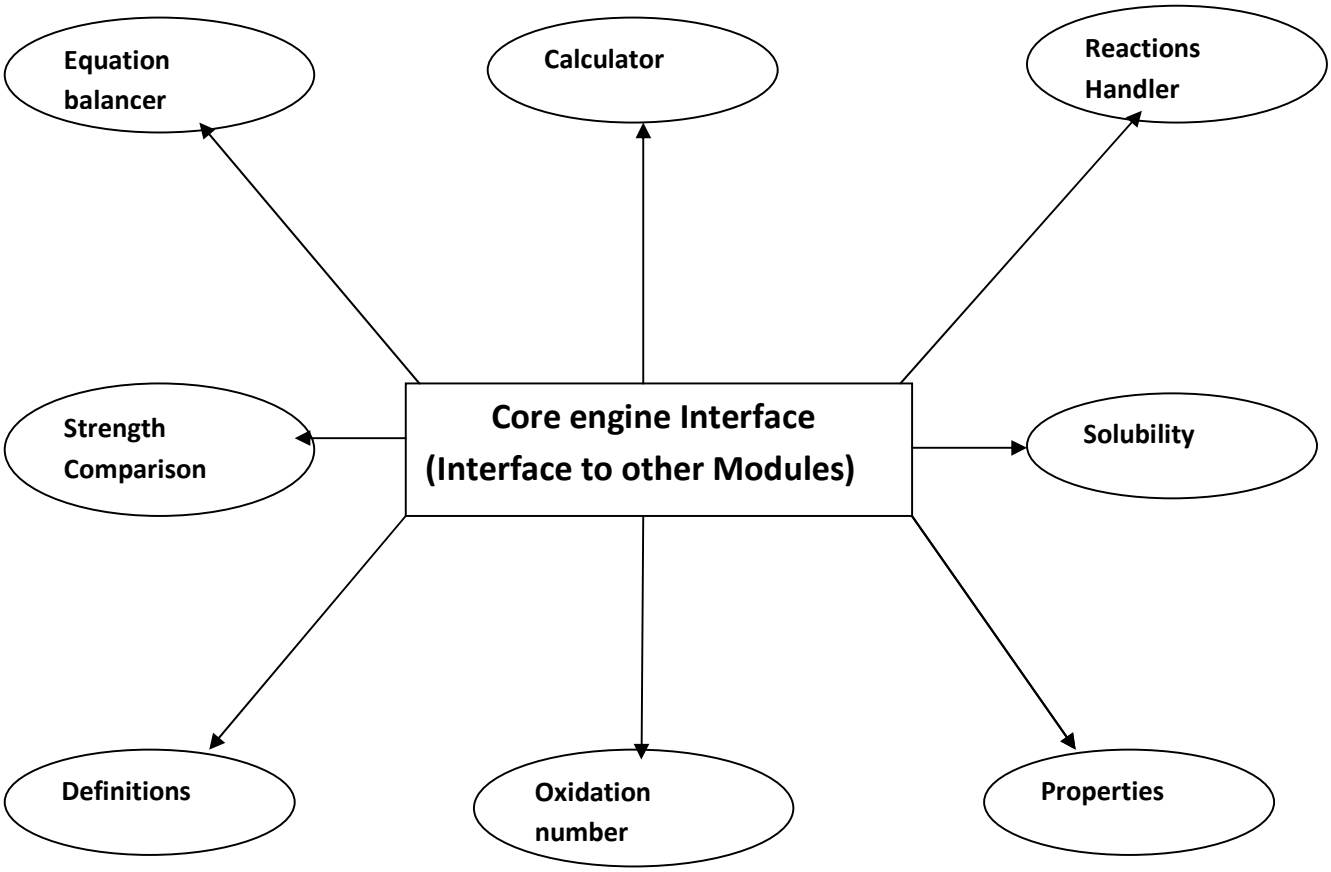
a. The first part consists of the entity we need to find about.

b. The second part consists of the requires data that will assist in finding the a part.

Based on the first part, the corresponding class to solve is being utilized by passing the parameters of the second class.

5.4 Core Engine Interface

This is the most important module of the system which interacts with every other module to get the answer to any query. It is a **query handler** which processes the query according to the *tag* and gets the result accordingly. Various tags include Definition, Property, Reaction, calculate pH, balance, Ions, oxidation number, etc. Each tag has a specific action associated with it. Here is a pictorial representation of the core engine interface.

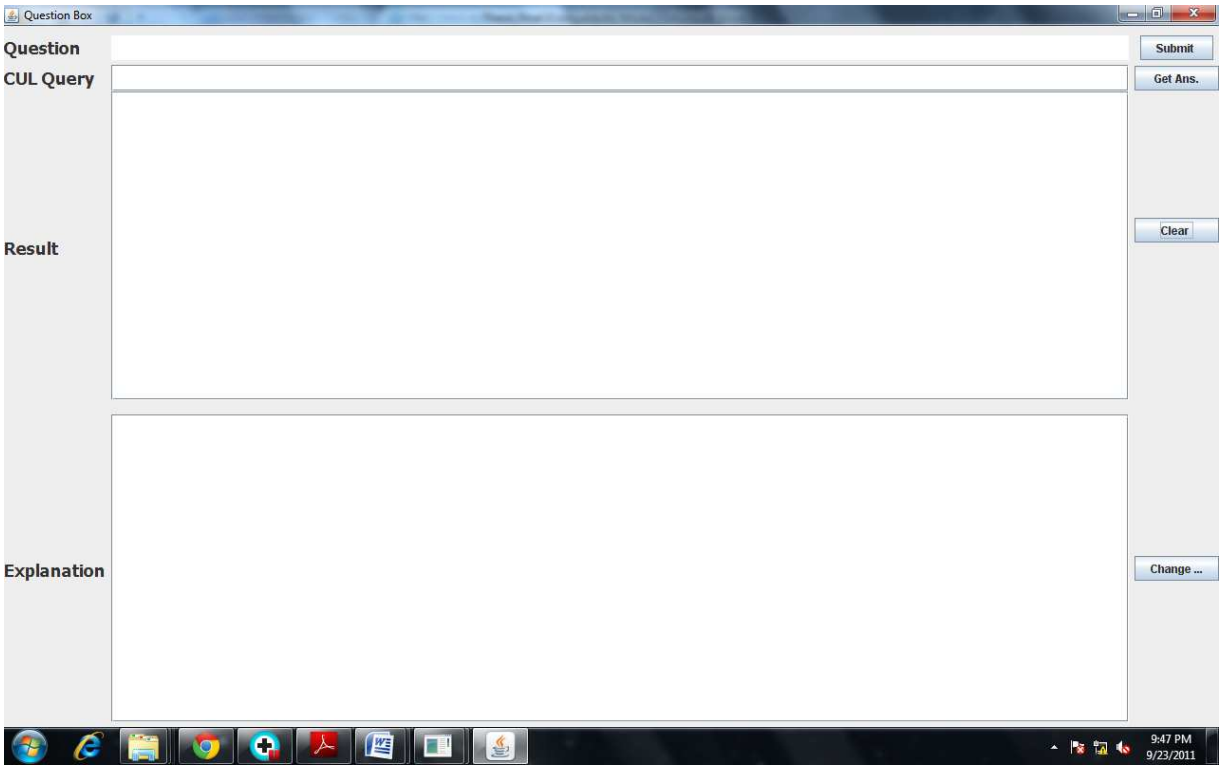


5.5 Mathematical Module

A large number of mathematical functions are needed to handle various types of queries. *For example*, we need to solve a polynomial of degree n, for calculating the concentration of [H+] ions, which is required for calculating the pH of acidic or basic solution. Similarly, we need matrix libraries for solving *system of dependent linear equations* which helps in balancing chemical equations. This module is called up by the core engine interface while solving for queries which include mathematical computations like ph calculation, balancing chemical equations, etc. It has other small functions like gcd, lcm of two numbers, etc. which assist in **solving other queries also.**

5.6 Output Module

This module is responsible for display of the query and its results. We have implemented Graphical User Interface (GUI) using JformDesigner. It helps us to insert Buttons, Text Area, ScrollPanel into JFrame. This module helps in displaying query as well as its answers and explanation in the Query Box.



5.7 RESULT

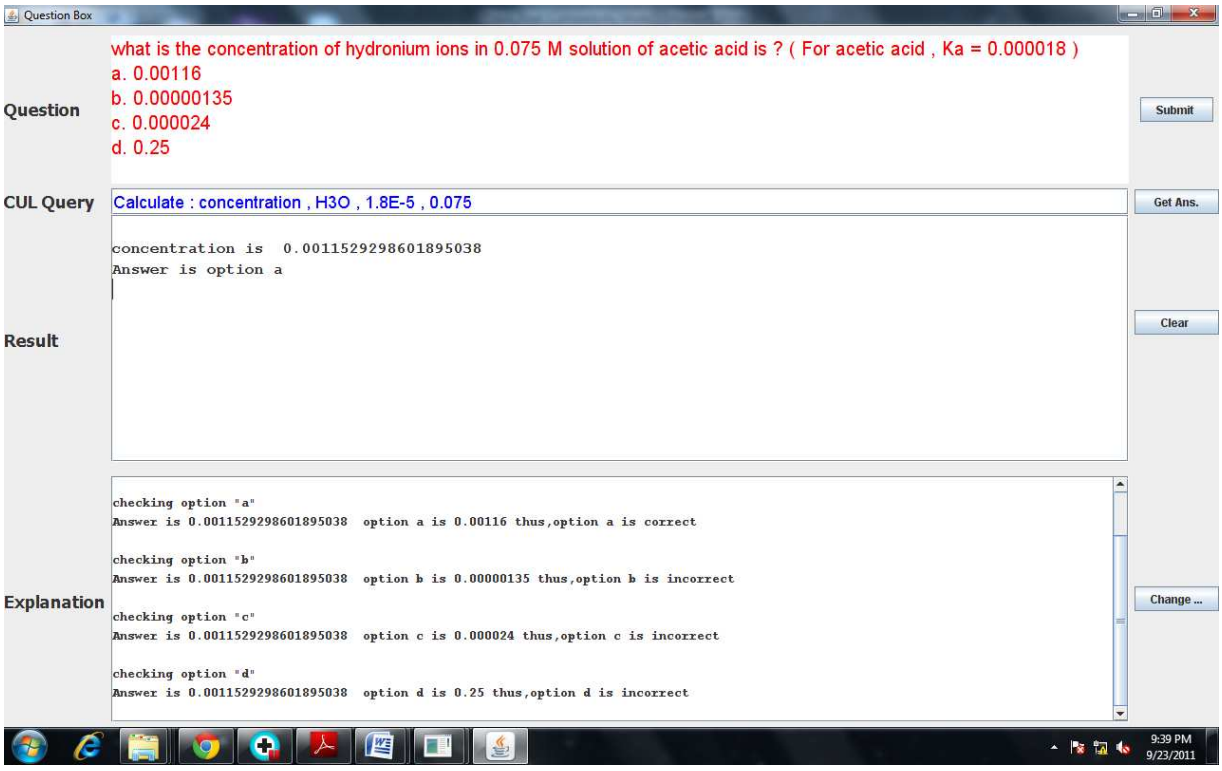
1.Number of questions used for training	80
2.Number of questions used for testing	210
3.Number of questions correctly answered by the system	174

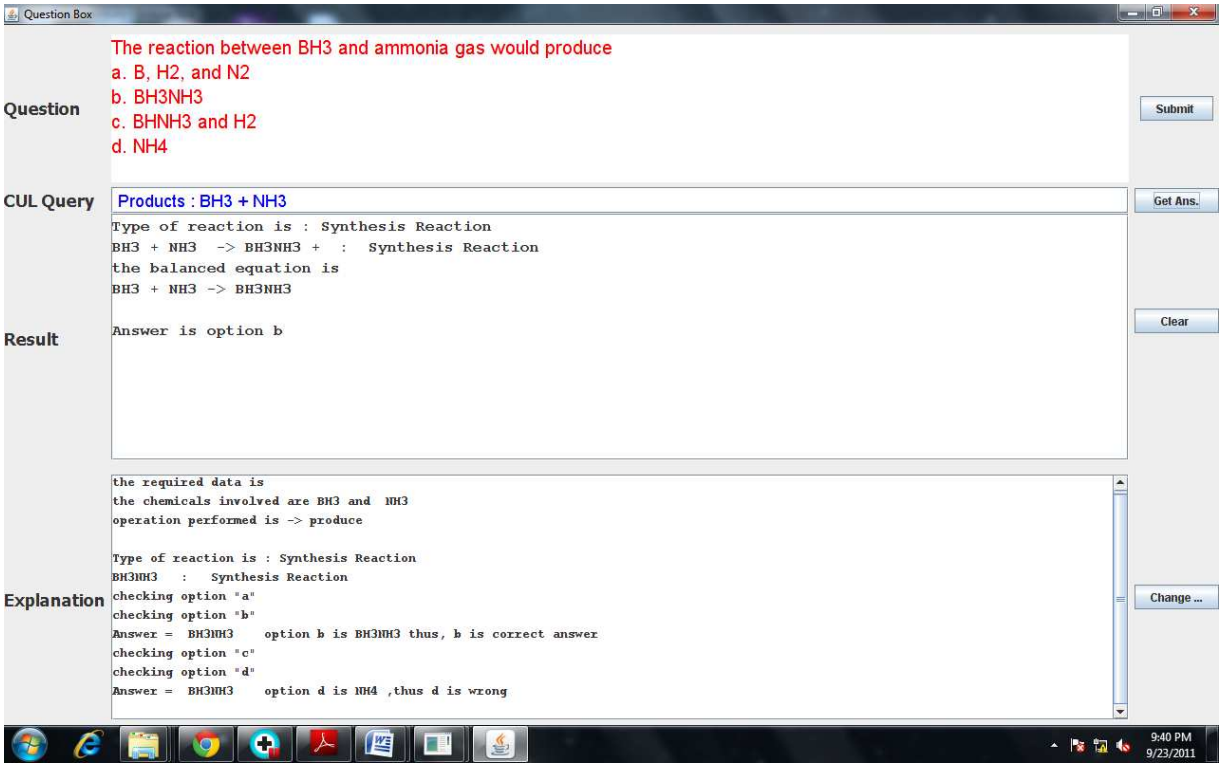
% Accuracy of the system to answer correct queries= (174/210)

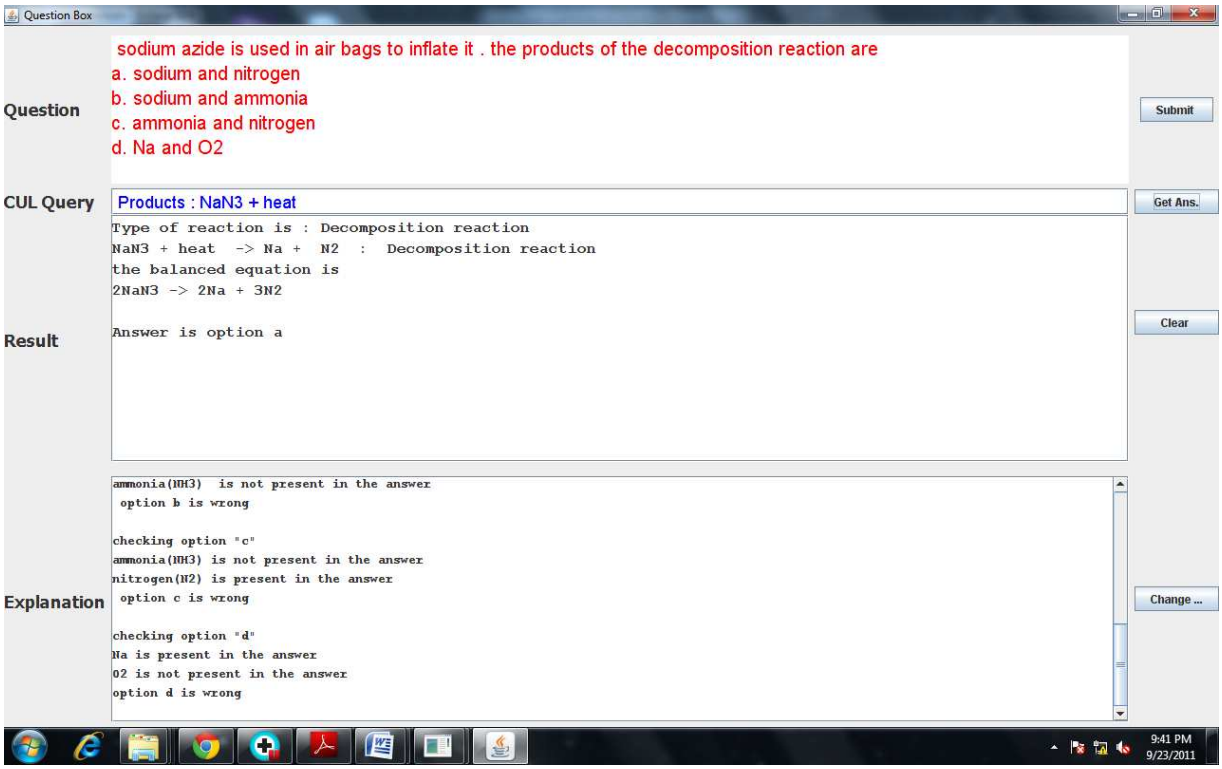
=.82857

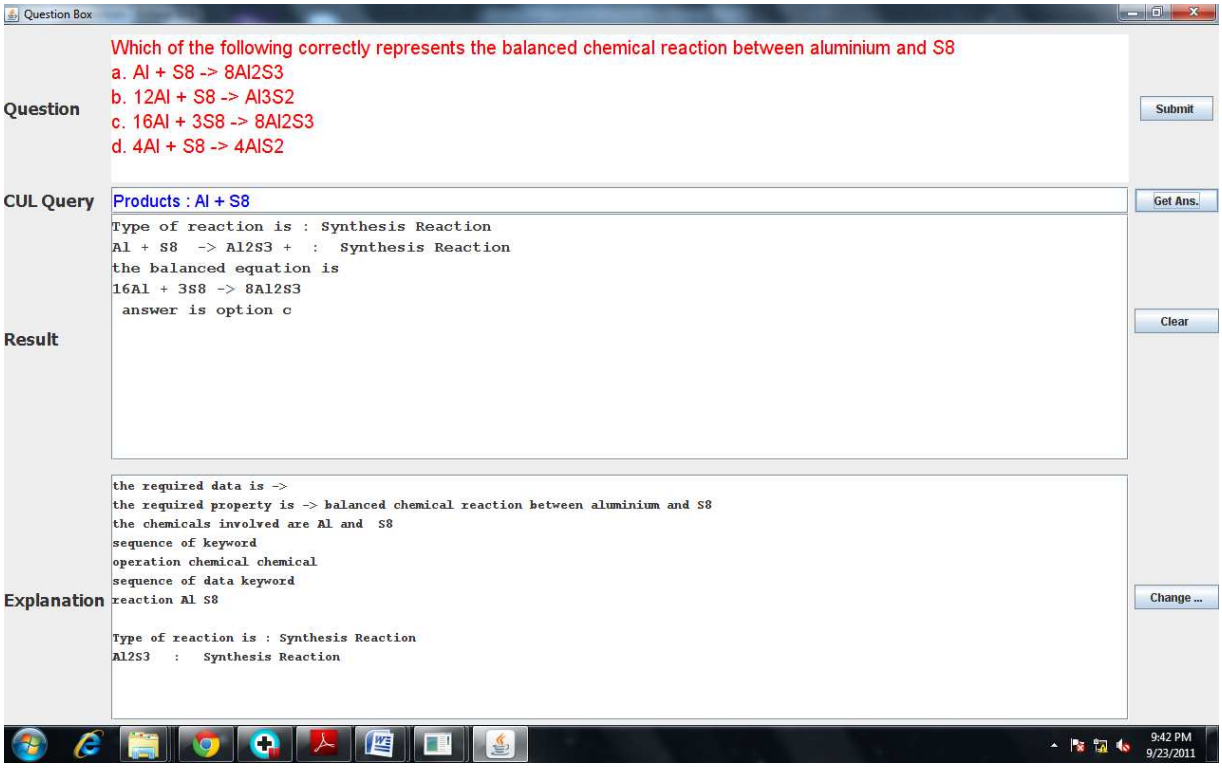
So system operates with the **82.857 %** accuracy**.**

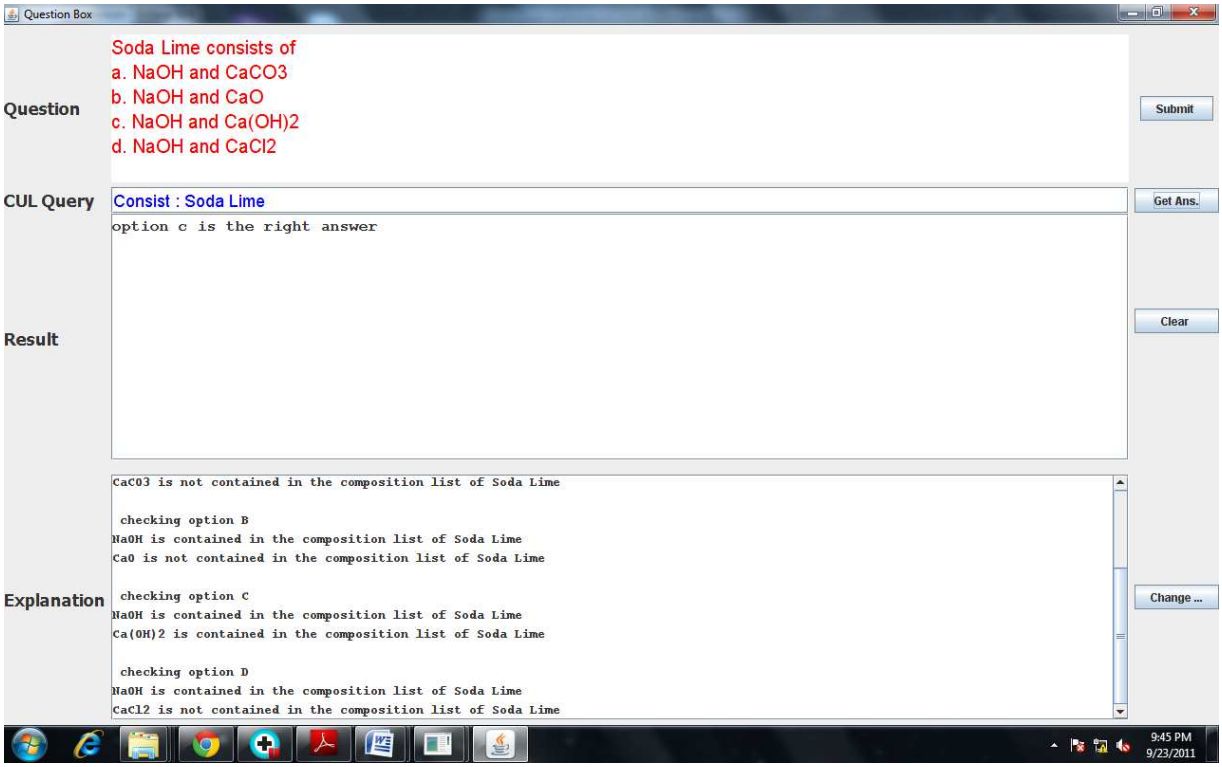
5.8 SCREENSHOTS




Question Box
Question
The reaction between BH3 and ammonia gas would produce
a. B, H2, and N2
b. BH3NH3
c. BHNH3 and H2
d. NH4
Submit
CUL Query
Products : BH3 + NH3
Get Ans.
Result
Type of reaction is : Synthesis Reaction
BH3 + NH3 -> BH3NH3 + : Synthesis Reaction
the balanced equation is
BH3 + NH3 -> BH3NH3
Answer is option b
Clear
Explanation
the required data is
the chemicals involved are BH3 and NH3
operation performed is -> produce
Type of reaction is : Synthesis Reaction
BH3NH3 : Synthesis Reaction
checking option "a"
checking option "b"
Answer = BH3NH3 option b is BH3NH3 thus, b is correct answer
checking option "c"
checking option "d"
Answer = BH3NH3 option d is NH4 ,thus d is wrong
Change ...
9:40 PM
9/23/2011


Question Box
Question
sodium azide is used in air bags to inflate it . the products of the decomposition reaction are
a. sodium and nitrogen
b. sodium and ammonia
c. ammonia and nitrogen
d. Na and O2
Submit
CUL Query
Products : NaN3 + heat
Get Ans.
Result
Type of reaction is : Decomposition reaction
NaN3 + heat -> Na + N2 : Decomposition reaction
the balanced equation is
2NaN3 -> 2Na + 3N2
Answer is option a
Clear
Explanation
ammonia(NH3) is not present in the answer
option b is wrong
checking option "c"
ammonia(NH3) is not present in the answer
nitrogen(N2) is present in the answer
option c is wrong
checking option "d"
Na is present in the answer
O2 is not present in the answer
option d is wrong
Change ...
9:41 PM
9/23/2011


Question Box
Question
Which of the following correctly represents the balanced chemical reaction between aluminium and S8
a. Al + S8 -> 8Al2S3
b. 12Al + S8 -> Al3S2
c. 16Al + 3S8 -> 8Al2S3
d. 4Al + S8 -> 4AlS2
Submit
CUL Query
Products : Al + S8
Get Ans.
Result
Type of reaction is : Synthesis Reaction
Al + S8 -> Al2S3 + : Synthesis Reaction
the balanced equation is
16Al + 3S8 -> 8Al2S3
answer is option c
Clear
Explanation
the required data is ->
the required property is -> balanced chemical reaction between aluminium and S8
the chemicals involved are Al and S8
sequence of keyword
operation chemical chemical
sequence of data keyword
reaction Al S8
Type of reaction is : Synthesis Reaction
Al2S3 : Synthesis Reaction
Change ...
9:42 PM
9/23/2011



5.9 DISCUSSION

Current system based upon the knowledge base and the queries presented to it works fine with acceptable accuracy but some improvements are there which could be done.

Following issues need to be solved for the enhancement of efficiency.

1. There are are many questions which are semantically exactly similar but syntactically there is a lot of difference between the two. So, in such cases the resultant queries must exactly be the same.

For example, these questions can be considered .

1 .What are the products formed when sodium hydroxide reacts with hydrochloric acid ?

2. Which of the following products are formed when sodium hydroxide is made to react with hydrochloric acid?
3. The reaction between NaOH and HCL produces.
4. When NaOH is reacted with HCL under proper conditions, the product formed is?

2. There are many words or sequence of words which in terms of chemistry semantically mean the same thing. To solve this we have something known as synonym set. In this we are keeping an account of all such words. For example:

Synonyms of reaction -> mixed, dissolves, react, fused, and added.

Another set -> balanced chemical equation, net chemical equation, balanced chemical reaction, correctly represents equation.

CHPATER 6: CONCLUSION AND FUTURE WORK

6.1 CONCLUSITION

There are two parts to handle queries in the scientific domain. The first part is to understand the query posed in the natural language. The second part is to search for the relevant information from the database to process the query. In this project we have concentrated on the first part. As such, we have simplified the nature of queries posed to our system in CUL (Computer Understandable Language).Second part is also very significant and some work has been done for this to build structured database, so that it can support large number of queries by searching the database itself provided it has correctly maped in HashMap for fast retrieval.

.6.2 FUTURE WORK

1. Till Now the focus has been mainly on the Natural Language Understanding, and so work also needs to be done on Natural Language Generation so that the given system can work well as a teaching module.

2. Regression testing will be done for the analysis of correctness of the system. In this, the comparison between the correct answer and obtained answer is being performed for the purpose.

3. Reading the questions from the input file will be made possible

4. Knowledge Formulation of other areas of chemistry.

5. Graphical Knowledge Updating Module for capturing knowledge.

6. Attempt to handle more variety of questions. For this purpose, the grammar needs to be expanded.

7. Utilizing the options also in formation of the query.

REFERENCES

1. Brown, T. L., H. E. LeMay, et al. (2003). Chemistry: *The Central Science*. New Jersey, Prentice Hall

2. Cosley,D.D. Frankowski,L. Terveen and J. Riedl,(2007).SuggestBot:*Using Intelligent Task Routing to Find Work in Wikipedia*.In T. Lau and A. R. Puerta (Eds.), IUI 2007: International Conference on Intelligent User Interfaces(pp. 32-41). New York : ACM.

3. Krotzsch, M. D. Vrandecic, M. Volkal ,et al.(2006). *Semantic Wikipedia*. Proceedings of the 15th International Conference on the World Wide Web, Edinburg, Scotland, pp. 585-594

4. Clark, P., J. Chaw, et al. (2007). *Capturing and Answering Questions Posed to A Knowledge-Based System*. International Conference on Knowledge Capture Systems (KCAP), Whistler, Canada

5. Chaudhari, V. K., B. John, et al. (2007). *Enabling Experts to Build Knowledge Bases from Science Textbooks*. International Conference on Knowledge Capture Systems.(KCAP). Whistler, Canada.

6. Chaudhri, V. K., K., B. Porter, et al. (2004). A *Question-Answering System for AP Chemistry*. The Ninth International Conference on the Principles of Knowledge Representation and Reasoning , Whistler, Canada

7. Gauss- Jordan Elimination “http://mathcorner.hoatwq.net/pages/gaussjordan/gaussjordan.

8. GaussJordanElimination“http//www.mpihd.mpg.de/astrophysik/HEA/internal/Numerical-Recipes/f2-1.pdf”

A. *Natural Language Understanding* (2nd Edition): James Allen , Benjamin/Cummings Pub. Co., 1995.
B. *Foundation of Statistical Natural language processing*: Chris Manning , MIT Press ,2003.

APPENDIX A1:

Questions of *what* kind:

1. What are the products formed when lithium is reacted with nitrogen?
a. LiN
b. Li3N
c. LiN3
d. Li2N

2. What is the pH of a 0.01 molar solution of HCN? (Ka for HCN is 0.0000000004)
1. 2.5
2. 5.7
3. 1.4
4. 6.5

3. What is the pH of a 0.041 molar solution of H2SO4?
a. 1.4
b. 2.4
c. 1.7
d. 4.3

4. What are the products formed when ammonium nitrate is heated?
a. nitrous oxide and water
b. nitric oxide and hydrogen
c. nitrogen dioxide and hydrogen
d. nitrogen dioxide and water

5. What are the ions in the chemical sodium nitrate?
a. K and NO3
b. K2 and NO2
c. K and NO2
d. Na and NO3

6. What are the most likely products for the reaction of NH3 with oxygen are:
a. NO and water
b. N2 and H2
c. N2 and water
d. H2 and NO

7. What products are formed when NaOH and CaCO3 are reacted?
a. sodium carbonate and calcium carbonate
b. calcium hydroxide and sodium carbonate
c. no reaction occurs

d. Na2CO3 and Ca

8. What is the result of the reaction between hydrochloric acid and sodium carbonate?
a. no reaction
b. carbonic acid and sodium chloride is produced
c. carbonic acid and sodium are produced
d. sodium chloride and hydrogen is produced

9. What are the products when sodium sulphate and calcium carbonate is reacted?
a. sodium carbonate and calcium sulphate
b. Na2CO3 and CaSO3
c. no reaction
d. Na2CO3 and CaSO4

10. What products are formed when Na2CO3 and CaCO3 are reacted?
a. sodium carbonate and calcium sulphate
b. Na2CO3 and CaSO4
c. no reaction occurs
d. Na2CO3 and CaSO4

APPENDIX A2:

questions of *which* category

1. Which gas is evolved when PbO2 is treated with conc. HNO3?
a. NO2
b. O2
c. N2
d. N2O

2. Which of the following compounds will produce a gas when HCl is added to the solid compound?
a. Ba(OH)2
b. CaCO3
c. CuSO4
d. Na3PO4

3. Which solution has the highest conductivity?
a. NH3
b. NaOH
c. Na3PO4
d. HCl

4. Which solution has the lowest conductivity?
a. NH3
b. NaOH
c. Na3PO4
d. HCl

5. Which of the following oxides is amphoteric in character?
a. CaO
b. CO2
c. CaO3
d. ZnO

6. Which one of the following is not an amphoteric oxide?
a. PbO2
b. SnO
c. B2O3
d. ZnO

7. Which of the following has the lowest conductivity?

a. CuS04
b. KOH
c. BaCl2
d. HF

8. Which metal does not liberate H2 from dilute aqueous hydrochloric acid at 298K?
a. Mg
b. Zn
c. Cu
d. Al

9. Which of the following is a non electrolyte?
a. NaCl
b. CH3COOH
c. NH3
d. CH3CH2OH

10. Which of the following combinations produce a precipitate ?
a. KCl + Ca(NO3)2
b. AgNO3 and NaCl
c. Al(CH3COO)3 + KNO3
d. HCl + KOH

APPENDIX A3:

Questions of *article* category

1. The most likely products for the reaction of NH3 with oxygen are:
a. NO and water
b. N2 , H2 , and H2O
c. N2 and water
d. H2 and NO

2. The pH of a 1.0 M solution of HCl is :
a. 1.0
b. 0.1
c. 0.0
d. less than zero

3. The pH of a 0.1 M solution of HCl is
a. 1.3
b. 2.4
c. 5.1
d. 1.0

4. The concentration of hydronium ions in 0.075 M solution of acetic acid is? (For acetic acid , Ka = 0.000018
a. 0.00116
b. 0.00000135
c. 0.000024
d. 0.25

5. The spectator ions in the reaction of barium nitrate with sodium sulphate are:
a. Na ions and barium ions.
b. sodium ions and sulfate ions .
c. Na ions and sulfate ions.
d. sodium ions and nitrate ions .

6. A vinegar having acetic acid (Ka = 0.000018) has pH 2.90 , the molar concentration of acetic acid is
a. 0.088
b. 0.890
c. 0.126
d. 0.014

7. A 0.3 M solution of acetic acid has a pH of 2.63. The Ka of this acid is
a. 0.000018
b. 0.0000000018
c. 0.0000013
d. 0.000019

8. A 0.3 M solution of acetic acid has a pH of 2.63. The ionization constant of this acid is
a. 0.000018
b. 0.0000000018
c. 0.0000013
d. 0.000019

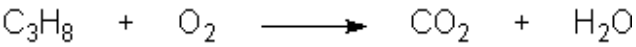
9. The reaction between BH3 and ammonia gas would produce
a. B, H2, and N2.
b. BH3NH3
c. BHNH3 and H2.
d. NH4

10. The conjugate base of [H2PO4]- is :
a. [PO4]3-
b. [HPO4]2-
c. H3PO4
d. [H3P2O7]

APPENDIX A4

Questions of *when* category:

1. When the following reaction equation is properly balanced, the number of moles of O_2 will be...



(a) 1.5 moles
(b) 3.5 moles
(c) 3 moles
(d) 5 moles

2.. When two ionic compounds are dissolved in water, a double replacement reaction can...

(a) Never occur since all ions in water are "spectator ions".
(b) Occur if two of the ions form an insoluble ionic compound, which precipitates out of solution.
(c) Occur if the ions react to form a gas, which bubbles out of the solution.
(d) Occur only if the ions form covalent bonds with each other.

C. When a small amount of iodine is shaken with trichloromethane, and water containing dissolved potassium iodide, a purple trichloromethane and a brown aqueous layer are obtained. Suppose more solid potassium iodide were dissolved in the water and the system again allow to come to equilibrium. What change or changes, if any, would you expect to see in the colours of the two layers?

a. The brown colour of the aqueous layer becomes paler, but the purple colour of the trichloromethane layer does not change.

b. The brown colour of the aqueous layer becomes more intense, byt the purple colour of the trichloromethane layer becomes paler.

c. The The brown colour of the aqueous layer becomes paler in colour, but the purple colour of the trichloromethane layer does not change.

d. Neither the purple colour of the trichloromethane layer, nor the brown colour of the aqueous layer changes in any way.

4. When 157.0 grams of $CaSO_4$ are dissolved in enough water to yield a volume of 7.25x102 milliliters of solution. The molarity of this solution will be...

(a) 0.0016 M
(b) 0.837 M
(c) 1.15 M
(d) 1.59

5. When lithium metal is reacted with nitrogen gas, under proper conditions, the product is :
a. LiN
b. Li2N
c. Li3N
d. LiN3

6. When methane, CH4 , gas reacts with oxygen , the following changes occur
a. Carbon dioxide is formed and the oxidation number of oxygen remain unchanged.
b. Carbon dioxide and water are formed and the oxidation number of oxygen remains unchanged.
c. Carbon dioxide and water are formed and the oxidation number of oxygen change zero.
d. Carbon monoxide and water are formed and thm -4 to +4.

7. When the equation is properly balanced , the number of moles of O2 will C6H14 + O2 -> CO2 + H2O
a. 1.5
b. 13
c. 19
d. 38

APPENDIX A5

Questions of *fact first* category:

1. Sodium azide is used in air bags to inflate it. The products of the decomposition reaction are
a. sodium and nitrogen
b. sodium and ammonia
c. ammonia and nitrogen
d. Na and O2

2. A solution nickel nitrate and sodium hydroxide mixed. Which of the following statements is true?
a. A precipitate will not form .
b. A precipitate of sodium nitrate will be produced.
c. Nickel hydroxide and sodium nitrate will precipitate.
d. Nickel hydroxide will precipitate.

3. Pb(NO3)2 + 2 NaCl -> NaNO3 + PbCl2 . The net ionic equation would include which of following?
a. all of them
b. Only Pb(NO3)2 and PbCl2
c. Pb2+ , Cl-, and PbCl2
d. Na+, NO3-, and NaNO3

4. H2Te + O2F2 -> TeF6 + HF + O2. Which of the following is true regarding above reaction ?
a. The oxidation number of H changes.
b. The oxidation number of F changes from +1 to -1.
c. The oxidation number of Te changes from +6 to -2.
d. There are no changes in oxidation states or the above answers are not correct.

5. If 2.68 g of hydrated sodium sulfate, $Na_2SO_4 \cdot nH_2O$, on heating produces 1.26 g of water, what is the empirical formula of this compound?

$Na_2SO_4 \cdot H_2O$

$2Na_2SO_4 \cdot H_2O$

$Na_2SO_4 \cdot 7H_2O$

6. If additional calcium phosphate is added to the above reaction mixture, what will happen to the overall reaction?

a. There will be no change in the overall reaction.
b. The reaction will occur at a faster rate.
c. Less of the reactants will react in order to compensate for the increase in the amount of one of the products of the reaction.
d. More of the reactants will have to react in order to compensate for the increase in the amount of one of the products of the reaction.

7. One of the functions of the catalytic converter in your car is to oxidize carbon monoxide to carbon dioxide. If 15.0 g of carbon monoxide reacts with 9.0 g of oxygen, how many grams of which compound remains unreacted? The balanced chemical equation is...

a. 0.4 g of oxygen remains unreacted
b. 0.8 g of carbon monoxide remains unreacted
c. 7.1 g of carbon monoxide remains unreacted
d. 8.1 g of oxygen remains unreacted

8. On oxidation, a compound $C_4H_{10}O$ can be converted into a compound C_4H_8O. The *original* compound could be a

I. primary alcohol
II. secondary alcohol
III. Tertiary alcohol

a. I, II and III are correct.
b. I and II are correct.
c. II and III are corrects.
d. I is the only correct response.

9. Some of the reactions are given. Which does not occur?

a. $B_2H_6 + 6H_2O \rightarrow 2H_3BO_3 + 6H_2$
b. $B_2H_6 + 2CO \rightarrow 2OC{\cdot}BH_3$
c. $B_2H_6 + 3RCH{=}CH_2 \rightarrow (RCH_2CH_2)_3B$
d. $B_2H_6 + H^+ \rightarrow [B_2H_7]^+$

APPENDIX B

CFG Grammar : A context-free grammar (CFG), sometimes also called a phrase structure grammar, is a grammar that naturally generates a formal language in which clauses can be nested inside clauses arbitrarily deeply, but where grammatical structures are not allowed to overlap. *CFG* are expressed by Backus–Naur Form, or *BNF*. In terms of production rules, every production of a context free grammar is of the form

$$V \rightarrow w$$

Where V is a single nonterminal symbol, and w is a string of terminals and/or nonterminals (w can be empty). These rewriting rules applied successively produce a parse tree, where the nonterminal symbols are nodes, the leaves are the terminal symbols, and each node expands by the production into the next level of the tree. The tree describes the nesting structure of the expression. V can therefore change while w is fixed (can't change).

In a context free grammar the left hand side of a production rule is always a single nonterminal symbol. In a general grammar, it could be a string of terminal and/or nonterminal symbols. The grammars are called *context free* because – since all rules only have a nonterminal on the left hand side – one can always replace that nonterminal symbol with what is on the right hand side of the rule. The *context* in which the symbol occurs is therefore not important.

Context-free languages are exactly those which can be understood by a finite state computer with a single infinite stack. In order to keep track of nested units, one pushes the current parsing state at the start of the unit, and one recovers it at the end.

Context-free grammars play a central role in the description and design of programming languages and compilers. They are also used for analyzing the syntax of natural languages

APPENDIX C

Shift Reduce Parsing:

A shift-reduce parser uses a *parse stack* which (conceptually) contains grammar symbols. During the operation of the parser, symbols from the input are *shifted* onto the stack. If a prefix of the symbols on top of the stack matches the RHS of a grammar rule **which is the correct rule to use within the current context**, then the parser *reduces* the RHS of the rule to its LHS, replacing the RHS symbols on top of the stack with the nonterminal occurring on the LHS of the rule. This shift-reduce process continues until the parser terminates, reporting either success or failure. It terminates with success when the input is legal and is *accepted* by the parser. It terminates with failure if an error is detected in the input.

The parser is nothing but a stack automaton which may be in one of several discrete *states*. A state is usually represented simply as an integer. In reality, the parse stack contains states, rather than grammar symbols. However, since each state corresponds to a unique grammar symbol, the state stack can be mapped onto the grammar symbol stack mentioned earlier.

The operation of the parser is controlled by a couple of tables:

Action Table

The *action table* is a table with rows indexed by states and columns indexed by terminal symbols. When the parser is in some state *s* and the current look ahead terminal is *t*, the action taken by the parser depends on the contents of *action*[*s*][*t*], which can contain four different kinds of entries:

Shift *s′*

Shift state *s′* onto the parse stack.

Reduce *r*

Reduce by rule *r*. This is explained in more detail below.

Accept

Terminate the parse with success, accepting the input.

Error

Signal a parse error.

Goto Table

The *goto table* is a table with rows indexed by states and columns indexed by nonterminal symbols. When the parser is in state *s* immediately **after** reducing by rule *N*, then the next state to enter is given by *goto*[*s*][*N*].

The current state of a shift-reduce parser is the state on top of the state stack. The detailed operation of such a parser is as follows:

1. Initialize the parse stack to contain a single state s_0, where s_0 is the distinguished *initial state* of the parser.

2. Use the state *s* on top of the parse stack and the current lookahead *t* to consult the action table entry *action*[*s*][*t*]:
    - If the action table entry is *shift s'* then push state *s'* onto the stack and advance the input so that the look ahead is set to the next token.
    - If the action table entry is *reduce r* and rule *r* has *m* symbols in its RHS, then pop *m* symbols off the parse stack. Let *s'* be the state now revealed on top of the parse stack and *N* be the LHS nonterminal for rule *r*. Then consult the goto table and push the state given by *goto*[*s'*][*N*] onto the stack. The lookahead token is not changed by this step.
    - If the action table entry is *accept*, then terminate the parse with success.
    - If the action table entry is *error*, then signal an error.
3. Repeat step (2) until the parser terminates.