Knowledgebase Approach for Organic Conversions

P. G. T. H. Kashmira

Department of Physical Sciences
Rajarata University
Mihintale, Sri Lanka

Abstract – **This paper explains about the approaches which are used to automate the organic conversions. This solution provides all the possible paths which can be used to perform organic conversions as well as optimum path among all the possible paths.**

Keywords – **Alkyl Groups, Functional Groups**

I. INTRODUCTION

Compounds which mainly consist of "carbon" are called as Organic compounds. Organic conversion is a process of converting a given initial compound into an expected compound by applying different reactants and conditions. There may be more than one possible path to perform organic conversions. Therefore sometimes it may need to find an optimum path among them.

Automated tools have become all important because of the importance of accessing right information at the right time. Organic conversions are using students and chemists for various studies and industries in their day today activities. It creates need for an automated tool for the organic conversions. I implemented this tool as a knowledge base system.

The objectives of this research are to find suitable logical structures for different organic compounds as well as reactions and to apply searching strategies to find all the possible paths for conversions.

In the rest of this article I will explain how I represented the different compounds and reactions in the knowledge base, how can find conversion paths and how can optimize the conversion paths.

II. REPRESENTATION MECHANISM

First I observed the structure of the compounds and the nature of their reactions. I found that the organic compounds are consisting of different alkyl groups and functional groups. Therefore first I designed the representation mechanism for the alkyl groups. Then I designed the representation for the different compounds by considering the alkyl groups. Finally I designed the logical rules to represent organic reactions.

A. Representation for the Alkyl Groups

Atoms of the alkyl groups and, the number of connecting bonds between the atoms of the alkyl groups are considered as the parameters. And also 'Hydrogen' atom is considered as the atomic alkyl group which can be connected into the carbon atom/atoms of the functional group. Therefore the first parameter of this atomic alkyl group is the 'hydrogen' atom and second parameter is one.

Other than that different non atomic alkyl groups can be created by foaming different branches at the carbon atoms of the main carbon chain of the alkyl group. These non atomic alkyl groups again can be considered as the collection of sub alkyl groups.

All the alkyl groups connect to carbon atom/atoms of the functional group by the single bond. When non atomic alkyl groups are connecting to carbon atoms of the functional group at left, right, up or down directions; different representations can be arouse. It happens due to the considering sequence of the branches at the major carbon atom in the alkyl group. To avoid it I used the proper sequence and the format. It is shown in Table 1. In that table; R1, R2, R3 are used to represent the sub alkyl groups within the non atomic alkyl groups. These sub alkyl groups can be atomic or non atomic.

TABLE 1
FORMAT OF THE ALKYL GROUP REPRESENTATION

Alkyl group	Join to carbon atom of the functional group at	Representation
R1 | —— C — R3 | R	Right	[C,[R1,1],[R2,1],[R3,1]]
R1 | R3 —— C—— | R2	Left	[C,[R1,1],[R2,1],[R3,1]]
R3 | R1 — C— R2 |	Up	[C,[R1,1],[R2,1],[R3,1]]
| R1 — C— R2 | R3	Down	[C,[R1,1],[R2,1],[R3,1]]

B. Representation for compounds

Alkyl groups, their number of connecting bonds to functional group and, atoms of the functional group with their number of connecting bonds are considered as the parameters.

Then I separated the carbon atoms of the functional group in the selected compound. Then I used the following parameters to represent each carbon atom in detail.

For each carbon atom; if they have alkyl groups/ other atoms of the functional group, take them according to the sequence of the left, up, down and right directions to carbon with their connecting bonds.

Then take the number of connecting bonds for the other adjacent carbon in the functional group. If there is no other adjacent carbon in the functional group; take the number of connecting bonds as zero.

Finally combine all the carbon atoms of the functional group with their parameters to represent an organic compound. Table 2 shows few examples about the representation mechanism of the compounds. In that table; R1, R2, R3, R4, R5, R6 are used to represent the R groups and the other letters are having the same meaning of the atoms.

TABLE 2
EXAMPLES FOR THE COMPOUND REPRESENTATION

Compound Name	Compound Structure	Representation
Alkane	R2 | R1 — C — R4 | R3	[c,[R1,1],[R2,1],[R3,1],[R4,1],0]
Alcohol	R2 | R1 — C— OH | R3	[C,[R1,1],[R2,1],[R3,1],[[o,[h,1]],1,0]
Acid	O ‖ R1 — C— OH	[C,[R1,1],[o,2],[[o,[h,1]],1],0]

C. Representation for reactions

Organic reactions can be categorized into two groups. They are; reactions which take organic compounds as a reactant and the reactions which take inorganic compounds as a reactant. I used following parameters to represent both types of the reaction rules in the knowledgebase.

They are; initial compound which participate into the reaction, reactants which should be applied to the reaction, compound which resulted from the reaction, cost of the reactant.

At this representation phase, I identified the problem of take infinite number of carbons and hydrogen atoms for the reactants by the reactions which use organic compound as a reactant. I have limited the number of hydrogen atoms of the alkyl groups into fixed number.

III.   ALL THE PATHS FOR THE CONVERSION

First, I considered the state space tree [8] of the conversion problem. Then, I applied the Depth First Search strategy [13] into the state space tree. By that found all the paths for the conversions. Search strategy is implemented as follows.

If there is a reaction from compound 'A' to compound 'B';

Then there is the conversion from 'A' to 'B'.

Otherwise, if there is a reaction from compound 'A' to some other compound 'D', and there is a conversion from compound 'D' to compound 'B';

Then there is a conversion from compound 'A' to compound 'B' too.

IV. OPTIMIZING CONVERSION PATHS

I considered two optimization factors for the optimization part of the conversions. They are;

Number of steps for the conversion

Cost of the reactants for the reactions of the conversion.

When optimization factor equals to the number of steps, used the Depth Limited Search Strategy to find the answers for the conversions. It avoids the drawback of Depth First Search which is getting into stuck by going down the wrong path of very deep search trees. Its implementation can be expressed as follows.

If there is a reaction from compound 'A' to compound 'B';

Then there is the conversion from 'A' to 'B' with N or fewer amounts of steps.

Otherwise, if there is a reaction from compound 'A' to some other compound 'D', and there is a conversion from compound 'D' to compound 'B' with M or fewer amount of steps and M equals to N-1;

Then there is a conversion from compound 'A' to compound 'B' with N or fewer amount of steps too.

When optimization factor equals to the number of steps, I had to find the total cost for each conversion paths. It was achieved as follows.

If; There is a reaction to create compound B from compound A with cost C,
Then;There is a conversion from compound A to compound B with cost C.
If; There is a reaction to create some other compound D from compound A with cost C1, and there is a conversion from D to B with cost C2,
Then; There is a conversion from A to B with cost C1 + C2.

Then I stored the cost of each conversion in a list. Then I sorted the list which contains the costs of conversion paths. Then select the conversion which has similar cost as the first element of above sorted list. Then obtain the reactants and the intermediary compounds of the selected conversion.

V. RESULTS

I considered two scenarios when checking the performance of the system. First I tested the performance of the system by considering only the reactions which use inorganic compounds as the reactant. (266 rules in the KB) Obtained results are shown in table3.

TABLE 3
PERFORMANCEs OF THE CONVERSIONS FOR THE ONE CATEGORY OF REACTIONS

Number of steps for the conversion	Average time taken to give the answers
2	30ms
3	30ms
4	36ms
5	38ms
6	1s
7	1s 50ms
8	2s 16ms
9	3s 58ms
10	7s 49ms

Then I tested the performance of the system by considering all the reactions which use inorganic and organic compounds as the reactant. (328 rules in the KB) Obtained results are shown in Table4. Here I have limited the number of hydrogen atoms of the alkyl groups in the reactant into three.

TABLE 4
PERFORMANCEs OF THE CONVERSIONS FOR BOTH CATEGORIES OF REACTIONS

Number of steps for the conversion	Average time taken to give the answers
2	34ms
3	40ms
4	72ms
5	1s 58ms
6	7s 22ms
7	23s 50ms
8	1min 7s
9	6min 35s
10	16min 20s

VI. CONCLUTIONS

This research was mainly focused on finding the suitable approaches for conversion paths as the knowledgebase system. Throughout my research I found that suitable general logical representations for different organic compounds and reactions.

Other than that I have identified the problem of take infinite number of carbons and hydrogen atoms for the reactants by the reaction rules which represent organic compound as a reactant at the unification phase. To solve that problem, I limited the number of hydrogen atoms into the fix number.

According to the results obtained from the system, it can be said that the performance of the system depends on the number of reaction rules in the knowledgebase and the number of steps for the conversion.

ACKNOWLEDGMENT

REFERENCES

[1] Education publication department, "Advanced Level Organic Chemistry".
[2] Professor G. P. Wannigama, "Organic Chemistry".
[3] Lionello Pogliani, "Graph Theoretical Concepts and Physiochemical Data".
[4] Ramon Garcia-Domenech, Jorge Galvez, Jesus V. de Julian-Ortiz, and Lionello Pogliani ,"Some New Trends in Chemical Graph Theory".
[5] V. K. Balakrishnan, Graph Theory.
[6] *SMILES*;
    http://www.daylight.com/dayhtml/doc/theory/theory.smiles.html.
[7] *SMARTS*;
    http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html.
[8] Stuart Russell, Peter Norvig, "Artificial Intelligence - A Modern Approach", 2nd Edition.
[9] IUPAC publications on Nomenclature and Symbolism;
    http://www.chem.qmul.ac.uk/iupac/
    http://www.acdlabs.com/iupac/nomenclature/
[10] Kuruvilla Mathew, Mujahid Tabassum, "Experimental Comparison of Uninformed and Heuristic AI Algorithms for N Puzzle and 8 Queen Puzzle Solutions", Swinburne University of Technology, Malaysia.
[11] Korf, R. E. "Depth-first iterative-deepening: An optimal admissible tree search", Department of Computer Science, Columbia University, New York.
[12] Ernst Ahlberg Helgee, "Improving Drug Discovery Decision Making using Machine Learning and Graph Theory in QSAR Modeling" , University of Gothenburg, Sweden.
[13] lvan Bratko, "PROLOG PROGRAMMING FOR ARTIFICIAL INTELLIGENCE " – 3rd Edition