

発表題目	示性式に近い線形表記による化学構造入力プログラムの開発	
発表者 (所属)	速水健一・今泉喬・本郷雄史 (仙台電波工業高等専門学校)	
連絡先	989-3124 仙台市青葉区上愛子北原 1 番地 Phone : 022-391-5508(代表) E-mail : hayami@cc.sendai-ct.ac.jp	
キーワード	示性式、化学式、化学構造入力、線形表記法 Rational formula, Chemical formula, Chemical structure input, Line notation	
開発意図 適用分野 期待効果 特徴など	化学構造をキーボードから線形(一行)入力できるもので、入力するための操作の方法や入力式の規則を習得することなく、たとえば示性式に近いものを入力するだけで平面構造をコンピュータに取り込むことのできるアプリケーションを開発した。	
環境	適応機種名	
	OS名	MS-DOS・Windows・Solaris・Linux
	ソース言語	C
	周辺機器	
流通形態 (右のいずれかに をつけて ください)	<ul style="list-style-type: none"> ・ 化学ソフトウェア学会の無償利用ソフトとする ・ 独自に配布する ・ ソフトハウス、出版社等から市販 ・ ソフトの配布は行わない その他 ・ 未定 	具体的方法

1. 背景

通常、化学構造を入力する(コンピュータに取り込む)には、GUI環境下でマウスなどを用いて、平面または立体的に構造を描画しながら行う。しかし、それらとは別に、手軽にキーボードから線形表記することで(文字列として一行で)入力できて、特に入力するための操作や、線形表記の方法を習得することなく、たとえば示性式や化学構造名に近い形で入力するだけで平面構造をコンピュータに取り込むことのできるアプリケーションの開発を行った。

GUI環境下で入力するものでは、一般にそれに必要なハードウェアに気を配らなければならない。機種依存性が生じたり、移植性が難しくなる。また、入力構造の形(ノードのつながり)を理解していないと入力が難しいことがある。

線形表記法による化学構造の入力は古くから考えられてきた。また、その表記法については、さまざまなもの(WLN、SMILES など)が提案されてきた[1]。これらは本来データベースを効率よく検索するためのものである。そこで、示性式とは違って表記法に関する習得が必要となる。教科書などにある化学式をそのまま入力することはできない。

2. 目的

表1の3つを目標に開発した。

表1. 目的

1). 化学構造を文字列として一行入力できるものにする
2). 示性式に近いものをコンピュータに理解させる
3). 式を認識させた最終形式は、結合行列とする

3. 方法

入力された式は、表 2 の項目順に処理される。そして、結合行列が出力される。

現在、稼動可能な環境を表 3 に示す。

1). MS-DOS, i86(16bit 可), PC-98x1 と PC/AT 互換機
2). Windows3.1/95/98 の DOS プロンプト, i86, PC-98x1 と PC/AT 互換機
3). WindowsNT/2000 の Command プロンプト, i86, PC/AT 互換機
4). Solaris/Linux, telnet 可
5). Solaris/Linux+Apache1.3, CGI, Internet Explorer または Netscape Navigator

表 2. 処理項目

[処理 1]: 式の置換と展開をする
[処理 1.1]: 式に含まれる慣用的な記号を置換する
[処理 1.2]: 式に含まれる数字やカッコを展開する
[処理 2]: 式に含まれる結合種表記や、価数表記、架橋表記をもとに結合行列を生成する
[処理 2.1]: ノード(式を構成する要素)を抽出する
[処理 2.2]: 架橋表記に従ってエッジ(結合)を生成する
[処理 2.3]: カッコの内側から、非末端要素から末端要素に対して、最も近い要素間から順にエッジを生成する
[処理 3]: 生成した結合行列に対して、必要に応じてオプションを適用する
[処理 3.1]: 結合行列のソート(ノード番号について、エッジ数について)を行う
[処理 3.2]: 分離構造の検査をする
[処理 3.3]: 余った結合に水素を付加する
[処理 3.4]: 不完全(部分)構造の検査をする

4. 結果と考察

アセチルサリチル酸を例として o-Ph(COOH)OAc と入力した場合には図 1 の処理過程を経て入力式の認識結果を結合行列として出力することができた。

現在までに可能なことを表 4 にまとめた。

1). 数字の使用
2). カッコの使用
3). 結合種表記
4). 逆順表記
5). 辞書機能による慣用的な記号や一部の単語の認識
6). 価数の指定
7). 架橋の指定
8). 環構造の記述

参考文献

[1]. 湯田浩太郎, 第 18 回情報化学討論会・第 23 回構造活性相関シンポジウム要旨集, P73 (1995)

入力	→ o-Ph(COOH)OAc	
処理過程	<p>[処理 1.1]: 慣用的な記号の置換 → (HC\$*1\$CH\$C\$CH\$C\$C\$*1)((C=O(OH)))O(C=O(CH3))</p> <p>[処理 1.1]: 数字の展開 → (HC\$*1\$CH\$C\$CH\$C\$C\$*1)((C=O(OH)))O(C=O(CHHH))</p> <p>[処理 2.1]: ノードの抽出 → HCCHCCHCCHCCOOHOCOCCHHH</p> <p>[処理 2.2]: 架橋(C\$*1 間)の生成 → (HC*CHCCHC\$*)((CO(OH)))O(CO(CHHH))</p> <p>[処理 2.3]: カッコと結合種表記を考慮してエッジの生成 → (HC*CHCCHC\$*)((CO(OH)))O(CO(CHHH)) → (HC*CHCCHC\$*)((CO(O*))O(CO(CHHH)) → (HC*CHCCHC\$*)((C**))O(CO(C**)) → (HC*CHCCHC\$*)((C**))O(C**)) → (C*C*CC*C\$*)((**))O(**)) → (*****))O(**)) → (*****))O(**)) → (*****))O(**)) → (*****))O(**))</p> <p>[処理 3.1]: ソート HCCHCCHCCHCCOOHOCOCCHHH → 0000CCCCCCCCCHHHHHHHH</p>	
出力	<pre>[Hex] Total 1 node(s) AS:NN: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, a, b, c, d, e, f, g, h, i, j, k, O: 0, 8, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, O: 1, 0, 8, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, O: 2, 0, 0, 8, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, O: 3, 0, 0, 0, 8, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, C: 4, 2, 1, 0, 0, 6, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, C: 5, 0, 1, 0, 0, 0, 6, 0, 0, 4, 4, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, C: 6, 0, 0, 2, 1, 0, 0, 6, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, C: 7, 0, 0, 0, 0, 1, 0, 0, 6, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, C: 8, 0, 0, 0, 0, 0, 4, 1, 0, 6, 0, 4, 0, 0, 0, 0, 0, 0, 0, 0, 0, C: 9, 0, 0, 0, 0, 0, 4, 0, 0, 0, 6, 0, 4, 0, 0, 0, 0, 0, 0, 1, 0, 0, C: a: 0, 0, 0, 0, 0, 0, 0, 4, 0, 6, 0, 4, 0, 0, 0, 0, 0, 0, 1, 0, 0, C: b: 0, 0, 0, 0, 0, 0, 0, 0, 4, 0, 6, 4, 0, 0, 0, 0, 0, 0, 0, 1, 0, C: c: 0, 0, 0, 0, 0, 0, 0, 0, 4, 4, 6, 0, 0, 0, 0, 0, 0, 0, 1, H: d: 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, H: e: 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, H: f: 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, H: g: 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, H: h: 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, H: i: 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, H: j: 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, H: k: 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1,</pre>	<pre> OO H10 = C4--C7--He O1 Hf C5 O2 \$ \$ = H11-C9 C8--C6 \$ \$ H13-Cb H12-Ca O3 \$ \$ Cc Hd H14</pre>

図 1. 処理過程 (例: アセチルサリチル酸)