

语法分析1-化学分子式解析

实验内容

熟悉PLY基本编写规范

熟悉yacc_example中分析四则运算的程序

编写程序，计算化学分子式中元素的数目，并完成以下测试：

```
atom_count("He") == 1
atom_count("H2") == 2
atom_count("H2SO4") == 7
atom_count("CH3COOH") == 8
atom_count("NaCl") == 2
atom_count("C60H60") == 120
```

参考语法

```
species_list : species_list species
species_list : species
species : SYMBOL
species : SYMBOL COUNT
```

lex.py是用来将输入字符串标记化

yacc.py用来对语言进行语法分析

实验步骤

1. 使用lex进行标记

我们将化学式分的字符串分成两大类：**字母**和**数字**，分别用**symbol**和**count**来表示，当他们各自符合我们规定的规则时，即符合**t_SYMBOL**和**t_COUNT**时，我们将这些字符分别标记，用于下面的yacc语法分析

```
tokens = (
    "SYMBOL",
    "COUNT"
)

t_SYMBOL = (
    r"[lroudsemf]?|Os?|N[eaibdpos]?|S[icernbmg]?|P[drmtboau]?|"
    r"H[eofgas]?|A[lrsgutcm]|B[eraik]?|Dy|E[urs]|F[erm]?|G[aed]|"
    r"I[nr]?|Kr?|L[iaur]|M[gnodt]|R[buhenaf]|T[icebmalh]|"
    r"U|V|W|Xe|Yb?|Z[nr]")

def t_COUNT(t):
    r"\d+"
    t.value = int(t.value)
    return t

def t_error(t):
```

```
print("Illegal character")
```

```
lex.lex()
```

2.使用yacc进行语法分析

yacc可以分析lex解析出的序列标记，化学解析式有以下四种语法：

```
species_list : species_list species
species_list : species
species :      SYMBOL
species :      SYMBOL COUNT
```

我们创建**原子**类 (Atom) 后，分别对他们进行解析：

```
class Atom(object):
    def __init__(self, symbol, count):
        self.symbol = symbol
        self.count = count

    def __repr__(self):
        return "Atom(%r, %r)" % (self.symbol, self.count)

def p_chemical_equation(p):
    """
    chemical_equation :
    chemical_equation : species_list
    """
    if len(p) == 1:
        p[0] = []
    else:
        p[0] = p[1]

def p_species_list(p):
    "species_list : species_list species"
    p[0] = p[1] + [p[2]]

def p_species(p):
    "species_list : species"
    p[0] = [p[1]]

def p_single_species(p):
    """
    species : SYMBOL
    species : SYMBOL COUNT
    """
    if len(p) == 2:
        p[0] = Atom(p[1], 1)
    elif len(p) == 3:
        p[0] = Atom(p[1], p[2])

def p_error(p):
    print("Syntax error in input!")

yacc.yacc()
```

运行结果

```
def atom_count(s):  
    count = 0  
    for atom in yacc.parse(s):  
        count += atom.count  
    return count  
  
if __name__ == "__main__":  
    print(atom_count("He") == 1)  
    print(atom_count("H2") == 2)  
    print(atom_count("H2SO4") == 7)  
    print(atom_count("CH3COOH") == 8)  
    print(atom_count("NaCl") == 2)  
    print(atom_count("C60H60") == 120)
```

True

True

True

True

True

True