

Fórmulas Químicas

1 Objectivo do Projeto

Um grupo de historiadores está a estudar uma série de experiências antigas de química e precisa de ajuda a organizar os dados. Têm os apontamentos de um grupo de químicos que determinou as fórmulas empíricas de vários compostos orgânicos, alguns deles também caracterizados quanto aos seus atributos químicos (alcanos, ácidos carboxílicos, cetonas, etc.).

O objectivo deste projeto é criar um programa capaz de interpretar um ficheiro de texto onde os historiadores, que não sabem programar, possam facilmente especificar que operações querem fazer com os seus dados. Executando as instruções nesse ficheiro, o programa deverá organizar os dados numa base de dados e usá-la para obter as estatísticas e gráficos que os historiadores especificarem.

2 Descrição do Problema

Pretende-se um programa cuja função principal se chame `processar` e receba o nome do ficheiro de texto com as instruções de processamento, fornecido pelos investigadores. O programa deve ler e executar os comandos descritos nesse ficheiro, que deve conter os seguintes comandos:

- **BASE DADOS ficheiro**

Este comando deve preceder qualquer comando que interaja com a base de dados e especifica o nome do ficheiro da base de dados que vai ser usado. Esse ficheiro pode já existir ou pode ser criado pelo sistema de gestão de bases de dados. Note que a base de dados pode mudar, havendo mais que um comando deste tipo. Se o programa encontrar outro comando destes deve garantir que a ligação à base de dados anterior é fechada e que as operações subsequentes sejam feitas na base de dados corrente.

- **CRIAR.TABELAS**

Ao encontrar este comando o programa deve criar duas tabelas na base de dados cujo nome foi especificado no comando `BASE_DADOS`:

- Tabela Compostos com campos, pela ordem indicada, para o identificador do composto (um número inteiro único para cada composto), o nome do composto (uma string), a fórmula química do composto (uma string) e o ponto de ebulição Celsius (um número decimal).
- Tabela Atributos com campos, pela ordem indicada, para o identificador do composto (um número inteiro) e o nome do atributo associado a esse composto (uma string). Como um composto pode ter vários atributos e vários compostos podem partilhar atributos, ambos os campos admitem valores repetidos. No entanto, não faz sentido que a tabela Atributos tenha duas vezes o mesmo atributo para o mesmo composto, portanto o par formado pelo identificador do composto e o atributo não pode ser repetido na tabela.

- CARREGAR ficheiro

Ler um ficheiro com o nome indicado e carregar a informação para as tabelas da base de dados nomeBD. O ficheiro tem, em cada linha, separados por ponto e vírgula (;), os valores do identificador do composto, nome do composto, fórmula do composto, ponto de ebulição e, finalmente, os atributos químicos associados ao composto. Os atributos podem ser um ou mais e, no caso de serem vários, estarão separados por vírgula (,).

Exemplo de um ficheiro a carregar:

```
1;Acetic Acid;C2H4O2;117.9;Carboxylic acid
2;Acetone;C3H6O;56.2;Ketone
3;Alanine;C3H7NO2;213;Amine,Carboxylic acid
4;Aniline;C6H7N;184.1;Amine,Aromatic
5;Benzene;C6H6;80.1;Aromatic
```

O identificador do composto, o seu nome, fórmula e ponto de ebulição devem ser inseridos na tabela Compostos. Os atributos químicos do composto devem ser inseridos, em conjunto com o identificador do composto, na tabela Atributos.

- REPORT nome ficheiro

Especifica o ficheiro para o onde as listas de compostos correspondentes ao critério de selecção serão gravadas. Todos os comandos do tipo LIST (ver abaixo) que sigam um comando REPORT devem acrescentar estas listas ao ficheiro especificado no comando REPORT. Se um novo ficheiro for especificado em REPORT então os comandos subsequentes do tipo LIST deverão ser gravados nesse ficheiro. Por exemplo, neste caso:

```
REPORT relatorio.txt
LIST CH *
LIST CH3 Alkane
LIST CH2O *
LIST CH2O Ester
REPORT relatorio2.txt
LIST C3H7 *
LIST C4H9 *
```

os primeiros quatro comandos LIST deverão resultar na escrita dos resultados no ficheiro relatorio.txt e os outros dois, a seguir ao segundo comando REPORT, deverão resultar na escrita do ficheiro relatorio2.txt.

- LIST formula_empirica;atributo

Escrever no ficheiro seleccionado no comando REPORT os nomes e fórmulas químicas dos compostos que tenham o atributo especificado e que sejam compatíveis com a fórmula empírica especificada.

A fórmula empírica de um composto indica a proporção de átomos de cada elemento desse composto. Por exemplo, a fórmula química do benzeno é C_6H_6 . Empiricamente, pode-se determinar que o benzeno tem o mesmo número de átomos de hidrogénio que tem de carbono e, assim, a sua fórmula empírica será CH .

A forma de determinar se uma fórmula química corresponde a uma fórmula empírica é verificar se tem os mesmos elementos e se a proporção de átomos entre a fórmula química

e a fórmula empírica é a mesma para todos os elementos. No exemplo acima, do benzeno, a fórmula química tem seis átomos de carbono e a empírica tem um, numa proporção de 6/1. No caso do hidrogénio é o mesmo, uma proporção de 6/1, por isso a fórmula empírica corresponde a esta fórmula química. Se a fórmula química fosse a do etano, C_2H_6 , não corresponderia porque a proporção de átomos de carbono seria 2/1 e a de hidrogénio de 6/1. Outras, como por exemplo a anilina, C_6H_7N , podia ser logo eliminada porque o azoto só está presente numa das fórmulas.

Encontrando o comando LIST, o programa deve seleccionar todos os compostos na base de dados que contenham o atributo especificado e verificar, para cada um, se é compatível com a fórmula empírica. Deve escrever no ficheiro seleccionado no comando REPORT a linha do comando LIST que está a ser executado e, a seguir a esta e com uma indentação de um tabulador (tab), uma linha por cada composto que corresponda aos critérios pedidos com o nome do composto e a sua fórmula química. Por exemplo, para os comandos ilustrados à esquerda, o resultado no ficheiro de relatório deve ser este mostrado à direita. Notem a indentação das linhas referentes aos compostos encontrados.

```
LIST CH *
LIST CH3 Alkane
LIST CH2O *
LIST CH2O Ester
```

```
LIST CH *
    Benzene;C6H6
LIST CH3 Alkane
    Ethane;C2H6
LIST CH2O *
    Acetic Acid;C2H4O2
    Methyl Formate;C2H4O2
LIST CH2O Ester
    Methyl Formate;C2H4O2
```

- GRAFICO ficheiro;atributo

Criar um gráfico com a relação entre o logaritmo do número de átomos de carbono e a temperatura de ebulição de todos os compostos com o atributo especificado. O gráfico deve apresentar no eixo das abcissas (x) o logaritmo do número de átomos de carbono de cada composto e, no eixo das ordenadas (y), o ponto de ebulição de cada composto. Deve também mostrar a recta de regressão linear para os dados apresentados. Seja esta recta:

$$y = \alpha + \beta x$$

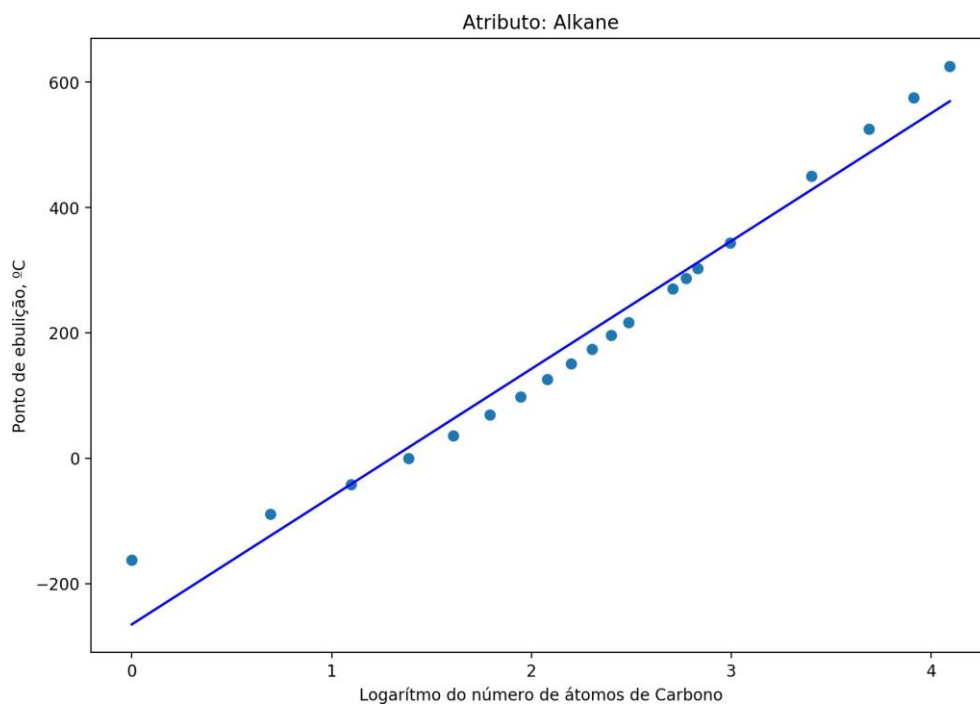
os parâmetros α e β podem ser calculado pela seguintes expressões:

$$\beta = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad \alpha = \frac{\sum y_i}{N} - \beta \frac{\sum x_i}{N}$$

onde x_i e y_i são os valores do logaritmo do número de átomos de carbono e do ponto de ebulição de cada um dos N compostos considerados e \bar{x} e \bar{y} as respectivas médias. A linha da regressão linear deve ser traçada desde o menor valor do logaritmo do número de átomos de carbono e o maior valor do logaritmo do número de átomos de carbono nesse conjunto.

Este gráfico ilustra o resultado esperado para o atributo Alkane.

Átomos C	Ebulição
4	0.0
10	174.0
12	216.0
2	-89.0
17	303.0
7	98.0
60	625.0
16	287.0
6	69.0
20	343.0
1	-162.0
9	151.0
8	126.0
50	575.0
15	270.0
5	36.0
3	-42.0
40	525.0
30	450.0
11	196.0



3 Dados do Projeto

O arquivo dados.zip tem os ficheiros compostos.txt e compostos2.txt, bem como o ficheiro de comandos orders.txt que podem ser usados como exemplos para testar o seu programa. Notem, no entanto, que estes testes não são exaustivos. É aconselhável fazerem também os vossos testes.

