

Centre of mass

Define a class `Atom` that supports the following methods:

1. An initializing method `__init__` to which two arguments must be given: the symbolic name of the atom (a string consisting of an uppercase letter, followed by zero or more lowercase letters) and the position of the atom (represented by a tuple (x, y, z) that represents a point in a three-dimensional area, where $x, y, z \in \mathbb{R}$). The initializing method must appoint these arguments to respectively the attributes `element` and `position` of the newly made object.
2. A method `mas` without arguments that prints the atomic mass of the atom. For the implementation of this method, make optimal use of a predefined dictionary with the symbolic names of all elements from the periodical system as a key, and the atomic masses as the corresponding values.
([click here to watch the predefined dictionary](#))
3. A method `__repr__` without arguments that prints a string representation of the atom. The method `__repr__` prints a syntactically correct Python expression, that — when it were to be evaluated — makes an object that is equal to the object that was originally given to `__repr__`. All real numbers must be given with two digits after the comma.
4. A method `__str__` without any arguments that prints a string representation of the atom. Read the example below to determine what this string representation should look like. All real numbers must be given with three digits after the comma.

Moreover, define a class `Molecule` that supports the following methods:

1. An initializing method `__init__` to which a name (string) for the molecule can be given optionally. Use the empty string as standard value for the name of the molecule. The initializing method must make sure that all object of the class `molecule` have two attributes: *i*) `name`, a string that contains the name of the molecule, and *ii*) `atoms`, a list of all atoms of the molecule. Initially the attribute `atoms` indicates an empty list.
2. A method `addAtom` that adds a given object from the class `Atom` (that is given to the method as an argument) to the bottom of the list of atoms.
3. A method `readPDB` that can read the positions of the atoms of a molecule from a text file in PDB format. To the method `readPDB` the location of the file is given as a string argument. For every `ATOM`-line from the file, the method must add an `Atom` object to the bottom of the list of atoms from the `Molecule` object, with an element name and a position as it is written on that line.

The *Protein Data Bank* (PDB) file format describes the three-dimensional structures of molecules like they are saved in the international PDB data bank. A typical PDB file can consist of hundreds of thousands of lines, as illustrated in the example below (shortened version of a file that describes the structure of a synthetic collagen-like peptide).

```
HEADER  EXTRACELLULAR MATRIX                22-JAN-98  1A3I
TITLE   X-RAY CRYSTALLOGRAPHIC DETERMINATION OF A COLLAGEN-LIKE
TITLE   2 PEPTIDE WITH THE REPEATING SEQUENCE (PRO-PRO-GLY)
...
EXPDTA  X-RAY DIFFRACTION
AUTHOR  R.Z.KRAMER,L.VITAGLIANO,J.BELLA,R.BERISIO,L.MAZZARELLA,
AUTHOR  2 B.BRODSKY,A.ZAGARI,H.M.BERMAN
...
```

```

REMARK 350 BIOMOLECULE: 1
REMARK 350 APPLY THE FOLLOWING TO CHAINS: A, B, C
REMARK 350 BIOMT1  1  1.000000  0.000000  0.000000    0.00000
REMARK 350 BIOMT2  1  0.000000  1.000000  0.000000    0.00000
...
SEQRES  1 A   9  PRO PRO GLY PRO PRO GLY PRO PRO GLY
SEQRES  1 B   6  PRO PRO GLY PRO PRO GLY
SEQRES  1 C   6  PRO PRO GLY PRO PRO GLY
...
ATOM    1  N  PRO A  1      8.316 21.206 21.530  1.00 17.44      N
ATOM    2  CA PRO A  1      7.608 20.729 20.336  1.00 17.44      C
ATOM    3  C  PRO A  1      8.487 20.707 19.092  1.00 17.44      C
ATOM    4  O  PRO A  1      9.466 21.457 19.005  1.00 17.44      O
ATOM    5  CB PRO A  1      6.460 21.723 20.211  1.00 22.26      C
...

```

Every line of a PDB file has a certain type, that is indicated by the first six characters of the line. For this assignment, we are only interested in the lines that describe the positions of the atoms of the molecule. These start with the letters ATOM, followed by two spaces. On these lines we find on positions 31-38, 39-46 and 47-54 respectively the x , y and z co-ordinates from the atom of the molecule (expressed in Ångström). The positions are given as real numbers with three digits after the comma, and may be preceded and/ or followed by spaces. On position 77-78 is the symbolic name of the atom, also possible preceded or followed by a space.

- A method `mass` without any arguments that prints the total mass of the molecule as a result. This is the sum of the masses of the individual atoms from which the molecule is built.
- A method `masscentre` without any arguments that prints the co-ordinates of the mass centre of the molecule as a tuple (x,y,z) . The mass centre (x,y,z) of a molecule that consists of n atoms with mass m_i ($1 \leq i \leq n$) and position (x_i, y_i, z_i) is defined by

$$\begin{aligned} x &= \frac{\sum_{i=1}^n m_i x_i}{\sum_{i=1}^n m_i} \\ y &= \frac{\sum_{i=1}^n m_i y_i}{\sum_{i=1}^n m_i} \\ z &= \frac{\sum_{i=1}^n m_i z_i}{\sum_{i=1}^n m_i} \end{aligned}$$

Voorbeeld

```

>>> oxygen = Atom('O', (4.013, 0.831, -9.083))
>>> oxygen.element
'O'
>>> oxygen.position
(4.013, 0.831, -9.083)
>>> oxygen.mass()
15.9994
>>> print(oxygen)
O-atom with mass 15.999 on position (4.013, 0.831, -9.083)
>>> oxygen
Atom('O', (4.01, 0.83, -9.08))

>>> molecule = Molecule(name='water')
>>> molecule.readPDB('water.pdb')
>>> molecule.atoms
[Atom('O', (4.01, 0.83, -9.08)), Atom('H', (4.94, 0.84, -8.84)), Atom('H', (3.75, -0.07, -9.29))]

```

```
>>> molecule.mass()
18.01528
>>> molecule.masscentre()
(4.0502061994040615, 0.7814290335759422, -9.080985829806696)
```

The interactive Python session above uses the PDB file [water.pdb](#) that describes the positions of the atoms of a water molecule. Below we give the ATOM-lines of that file, preceded by two lines that clarify on which positions the co-ordinates of the atoms are given.

```

1      2      3      4      5      6      7      8
1234567890123456789012345678901234567890123456789012345678901234567890
ATOM   1  OH  OSP3  1    4.013  0.831 -9.083  1.00  0.00      O
ATOM   2  1HH OSP3  1    4.941  0.844 -8.837  1.00  0.00      H
ATOM   3  2HH OSP3  1    3.750 -0.068 -9.293  1.00  0.00      H
```

Definieer een klasse `Atoom` die ondersteuning biedt voor volgende methoden:

1. Een initialisatiemethode `__init__` waaraan twee argumenten moeten doorgegeven worden: de symbolische naam van het atoom (een string bestaande uit een hoofdletter, gevolgd door nul of meer kleine letters) en de positie van het atoom (voorgesteld als een tuple (x, y, z) dat een punt in een driedimensionale ruimte voorstelt, waarbij $x, y, z \in \mathbb{R}$). De initialisatiemethode moet deze argumenten respectievelijk toekennen aan de attributen `element` en `positie` van het nieuw aangemaakte object.
2. Een methode `massa` zonder argumenten die de atoommassa van het atoom teruggeeft. Maak voor de implementatie van deze methode optimaal gebruik van een voordefinieerde dictionary met de symbolische naam van alle elementen uit het periodiek systeem als sleutel, en de atoommassa's als de corresponderende waarden.
([klik hier om voorgedefinieerde dictionary te bekijken](#))
3. Een methode `__repr__` zonder argumenten die een stringvoorstelling van het atoom teruggeeft. De methode `__repr__` geeft een syntactisch correcte Python expressie terug, die — wanneer deze geëvalueerd zou worden — een object aanmaakt dat gelijk is aan het object dat origineel werd doorgegeven aan `__repr__`. Alle reële getallen moeten weergegeven worden met 2 cijfers na de komma.
4. Een methode `__str__` zonder argumenten die een stringvoorstelling van het atoom teruggeeft. Bekijk onderstaand voorbeeld om te achterhalen hoe deze stringvoorstelling er moet uitzien. Alle reële getallen moeten weergegeven worden met drie cijfers na de komma.

Definieer daarenboven nu ook een klasse `Molecule` die ondersteuning biedt voor volgende methoden:

1. Een initialisatiemethode `__init__` waaraan optioneel een naam (string) voor de molecule kan doorgegeven worden. Gebruik de lege string als standaardwaarde voor de naam van de molecule. De initialisatiemethode moet ervoor zorgen dat alle objecten van de klasse `molecule` over twee attributen beschikken: *i*) naam, een string die de naam van de molecule bevat, en *ii*) atomen, een lijst van alle atomen van de molecule. Initieel verwijst het attribuut `atomen` naar een lege lijst.
2. Een methode `atoomToevoegen` die een gegeven object van de klasse `Atoom` (dat als argument aan de methode wordt doorgegeven) achteraan toevoegt aan de lijst van atomen.
3. Een methode `leesPDB` die de posities van de atomen van een molecule kan inlezen uit een tekstbestand in PDB formaat. Aan de methode `leesPDB` wordt de locatie van het bestand als

string-argument doorgegeven. Voor elke ATOM-regel uit het bestand moet de methode een Atoom object achteraan de lijst van atomen van het Molecule object toevoegen, met een elementnaam en een positie zoals die op de regel worden beschreven.

Het *Protein Data Bank* (PDB) bestandsformaat beschrijft de driedimensionale structuren van moleculen zoals die worden opgeslagen in de internationale PDB databank. Een typisch PDB bestand kan bestaan uit honderden of duizenden regels, zoals geïllustreerd in onderstaand voorbeeld (ingekorte versie van een bestand dat de structuur van een synthetische collageenachtige peptide beschrijft).

```

HEADER  EXTRACELLULAR MATRIX                22-JAN-98  1A3I
TITLE   X-RAY CRYSTALLOGRAPHIC DETERMINATION OF A COLLAGEN-LIKE
TITLE   2 PEPTIDE WITH THE REPEATING SEQUENCE (PRO-PRO-GLY)
...
EXPDTA  X-RAY DIFFRACTION
AUTHOR  R.Z.KRAMER,L.VITAGLIANO,J.BELLA,R.BERISIO,L.MAZZARELLA,
AUTHOR  2 B.BRODSKY,A.ZAGARI,H.M.BERMAN
...
REMARK 350 BIOMOLECULE: 1
REMARK 350 APPLY THE FOLLOWING TO CHAINS: A, B, C
REMARK 350 BIOMT1  1  1.000000  0.000000  0.000000      0.00000
REMARK 350 BIOMT2  1  0.000000  1.000000  0.000000      0.00000
...
SEQRES  1 A   9  PRO PRO GLY PRO PRO GLY PRO PRO GLY
SEQRES  1 B   6  PRO PRO GLY PRO PRO GLY
SEQRES  1 C   6  PRO PRO GLY PRO PRO GLY
...
ATOM    1  N  PRO A  1      8.316 21.206 21.530  1.00 17.44      N
ATOM    2  CA PRO A  1      7.608 20.729 20.336  1.00 17.44      C
ATOM    3  C  PRO A  1      8.487 20.707 19.092  1.00 17.44      C
ATOM    4  O  PRO A  1      9.466 21.457 19.005  1.00 17.44      O
ATOM    5  CB PRO A  1      6.460 21.723 20.211  1.00 22.26      C
...

```

Elke regel van een PDB bestand heeft een bepaald type, dat wordt aangegeven door de eerste zes karakters van de regel. Voor deze opgave zijn we enkel geïnteresseerd in de regels die de posities van de atomen van de molecule beschrijven. Deze starten met de letters ATOM, gevolgd door twee spaties. Op dergelijk regels vinden we op posities 31-38, 39-46 en 47-54 respectievelijk de x -, y - en z -coördinaat van een atoom van de molecule terug (uitgedrukt in Ångström). Deze posities worden weergegeven als reële getallen met drie cijfers na de komma, en worden eventueel voorafgegaan en/of gevolgd door spaties. Op positie 77-78 staat de symbolische naam van het atoom, ook eventueel voorafgegaan of gevolgd door een spatie.

4. Een methode `massa` zonder argumenten die de totale massa van de molecule als resultaat teruggeeft. Dit is de som van de massa's van de individuele atomen waaruit de molecule is opgebouwd.
5. Een methode `massacentrum` zonder argumenten die de coördinaten van het massacentrum van de molecule als een tuple (x,y,z) teruggeeft. Het massacentrum (x,y,z) van een molecule die bestaat uit n atomen met massa m_i ($1 \leq i \leq n$) en positie (x_i, y_i, z_i) wordt gedefinieerd door

$$x = \frac{\sum_{i=1}^n m_i x_i}{\sum_{i=1}^n m_i}$$

$$y = \frac{\sum_{i=1}^n m_i y_i}{\sum_{i=1}^n m_i}$$

\quad

$\sum_{i=1}^n m_i z_i$

\]

Voorbeeld

```
>>> zuurstof = Atoom('O', (4.013, 0.831, -9.083))
>>> zuurstof.element
'O'
>>> zuurstof.positie
(4.013, 0.831, -9.083)
>>> zuurstof.massa()
15.9994
>>> print(zuurstof)
O-atoom met massa 15.999 op positie (4.013, 0.831, -9.083)
>>> zuurstof
Atoom('O', (4.01, 0.83, -9.08))

>>> molecule = Molecule(naam='water')
>>> molecule.leesPDB('water.pdb')
>>> molecule.atomen
[Atoom('O', (4.01, 0.83, -9.08)), Atoom('H', (4.94, 0.84, -8.84)), Atoom('H', (3.75, -0.07, -9.29))]
>>> molecule.massa()
18.01528
>>> molecule.massacentrum()
(4.0502061994040615, 0.7814290335759422, -9.080985829806696)
```

Bovenstaande interactieve Python sessie maakt gebruik van het PDB bestand [water.pdb](#) dat de posities van de atomen van een watermolecule beschrijft. Hieronder geven we de ATOM-regels van dat bestand, voorafgegaan door twee regels die verduidelijken op welke posities de coördinaten van de atomen staan weergegeven.

```
      1      2      3      4      5      6      7      8
1234567890123456789012345678901234567890123456789012345678901234567890
ATOM   1  OH  OSP3   1    4.013  0.831 -9.083  1.00  0.00      O
ATOM   2 1HH  OSP3   1    4.941  0.844 -8.837  1.00  0.00      H
ATOM   3 2HH  OSP3   1    3.750 -0.068 -9.293  1.00  0.00      H
```