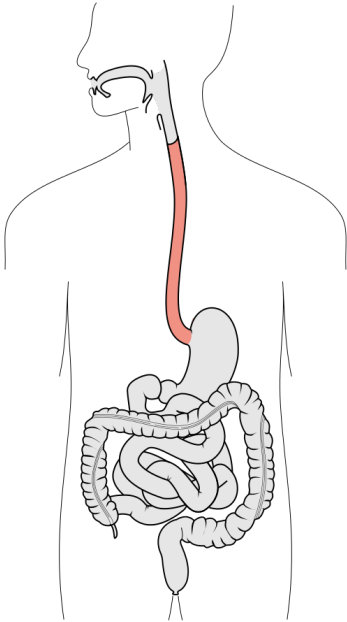


Esophagographers

Esophagographers are members of the medical staff who carry out esophagography, defined in Stedman's Medical Dictionary (6th edition, 2008) as *radiography of the esophagus using swallowed or injected radiopaque contrast media*. The esophagus — commonly known as the foodpipe or gullet — is an organ in vertebrates which consists of a fibromuscular tube through which food passes, aided by peristaltic contractions from the pharynx to the stomach.



Scheme of the human digestive tract, with the esophagus marked in red.

In this exercise, however, we are interested in the word esophagographers for a non-medical reason. Esophagographers, sixteen letters long, is the longest English word in which each letter occurs exactly twice. A fourteen-letter word with this property is **scintillescent**. Twelve-letter words with this property include **happenchance** and **shanghaiings**. Ten-letter words with this property include **arraigning**, **concisions**, **intestines**, and **horseshoer**.

Assignment

Words in which each letter has the same occurrence are called **repetition words**. You are asked to:

- Write a function `occurrences` that takes a word as its argument. The function must return a dictionary that maps each letter in the word onto the number of occurrences of that letter in the word. The function must process given words in a case insensitive way and the returned dictionary must only use lowercase letters as keys.
- Use the function `occurrences` to write a function `isRepetitionWord` that takes a word as its argument. The function also has a second optional parameter `minimal_repetition` (default value 1). The function must return a Boolean value that indicates whether or not the given word is a repetition word. In determining whether the given word is a repetition word, the function must treat the letters in a case insensitive way and must ignore all characters that are not letters. To be a repetition word, the number of occurrences of each letter must also be at least as high as the value passed to the parameter `minimal_repetition`.
- Use the function `isRepetitionWord` to write a function `repetitionWords` that takes the location of a

text file as its argument. The text file must contain a list of words, each on a separate line. The function also has two optional parameters `minimal_repetition` and `minimal_length`, both having 1 as their default value. The function must return the set of repetition words contained in the given file whose length is at least as high as the value passed to the parameter `minimal_length`. In determining whether or not a word is a repetition word, the function must make use of the function `isRepetitionWord`. In doing so, the parameter `minimal_repetition` has the same meaning for the function `repetitionWords` as it has for the function `isRepetitionWord`.

Example

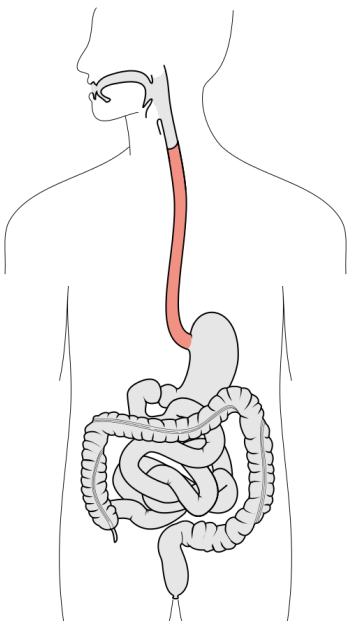
In the following interactive session, we assume that the text file [words.txt](#) is located in the current directory.

```
>>> occurrences('CHACHACHA')
{'a': 3, 'h': 3, 'c': 3}
>>> occurrences('Esophagographers')
{'a': 2, 'e': 2, 'g': 2, 'h': 2, 'o': 2, 'p': 2, 's': 2, 'r': 2}
>>> occurrences('happenstance')
{'a': 2, 'c': 2, 'e': 2, 'h': 2, 'n': 2, 'p': 2}

>>> isRepetitionWord('CHACHACHA')
True
>>> isRepetitionWord('Esophagographers')
True
>>> isRepetitionWord('happenstance', minimal_repetition=3)
False

>>> repetitionWords('words.txt', minimal_repetition=2, minimal_length=10)
{'horseshoer', 'intestines'}
```

Esophagographers is de Engelse term voor personeelsleden van een medisch team die instaan voor het uitvoeren van een oesofagografie. Bij oesofagografie worden opnamen gemaakt van een staande patiënt tijdens het doorlichten van de oesofagus bij het passeren van een slok bariumcontrastmiddel. De slokdarm of oesofagus is een onderdeel van het spijsverteringsstelsel van gewervelden die bestaat uit een holle sluitspier waardoor voedsel passeert, geholpen door peristaltische samentrekkingen, tussen de farynx (keelholte) en de maag.



Het maag-darmstelsel bij de mens, waarbij de oesofagus (slokdarm) wordt aangegeven in het rood.

In deze opgave zijn we echter om een niet-medische reden geïnteresseerd in het woord *esophagographers*. *Esophagographers*, zestien letters lang, is immers het langste Engelse woord waarin elke letter juist twee keer voorkomt. Een veertienletterwoord met deze eigenschap is **scintillesscent**. Twaalfletterwoorden met deze eigenschap zijn onder meer **happenchance** en **shanghaiings**. Onder de tienletterwoorden met deze eigenschap vinden we **arraigning**, **concisions**, **intestines**, en **horseshoer**.

Opgave

Woorden waarvan alle letters even vaak voorkomen worden **herhaalwoorden** genoemd. Gevraagd wordt:

- Schrijf een functie `voorkomens` waaraan een woord moet doorgegeven worden. De functie moet een dictionary teruggeven, die elke letter die in het woord voorkomt, afbeeldt op het aantal voorkomens van die letter in het woord. De functie mag hierbij geen onderscheid maken tussen hoofdletters en kleine letters, en de sleutels van de dictionary mogen enkel uit kleine letters bestaan.
- Gebruik de functie `voorkomens` om een functie `isHerhaalwoord` te schrijven waaraan een woord moet doorgegeven worden. De functie heeft een optionele tweede parameter `minimale_herhaling` (standaardwaarde 1). De functie moet een Booleaanse waarde teruggeven, die aangeeft of het gegeven woord een herhaalwoord is of niet. Bij het bepalen of een woord een herhaalwoord is, mag geen onderscheid gemaakt worden tussen hoofdletters en kleine letters, en moeten alle karakters van het woord die geen letter zijn, genegeerd worden. Om als herhalingswoord in aanmerking te komen moet het aantal voorkomens van elke letter ook groter of gelijk zijn aan de waarde die wordt doorgegeven aan de parameter `minimale_herhaling`.
- Gebruik de functie `isHerhaalwoord` om een functie `herhaalwoorden` te schrijven waaraan de locatie van een tekstbestand moet doorgegeven worden. Dit tekstbestand moet een lijst van woorden bevatten, elk op een afzonderlijke regel. De functie heeft ook nog twee optionele parameters `minimale_herhaling` en `minimale_lengte`, beide met standaardwaarde 1. De functie moet de verzameling herhaalwoorden teruggeven die in het bestand voorkomen en die minstens evenveel karakters bevatten als de waarde doorgegeven aan de parameter `minimale_lengte`. Om te bepalen of een woord al dan niet een herhaalwoord is, moet gebruik gemaakt worden van de functie `isHerhaalwoord`. De parameter `minimale_herhaling` heeft daarbij dezelfde betekenis voor de functie `herhaalwoorden` als voor de functie `isHerhaalwoord`.

Voorbeeld

Bij onderstaande voorbeeldsessie gaan we ervan uit dat het bestand [woorden.txt](#) zich in de huidige directory bevindt.

```
>>> voorkomens('CHACHACHA')
{'a': 3, 'h': 3, 'c': 3}
>>> voorkomens('ronddobberen')
{'b': 2, 'e': 2, 'd': 2, 'o': 2, 'n': 2, 'r': 2}
>>> voorkomens('ballenbakken')
```

```
{'a': 2, 'b': 2, 'e': 2, 'k': 2, 'l': 2, 'n': 2}
```

```
>>> isHerhaalwoord('CHACHACHA')
```

```
True
```

```
>>> isHerhaalwoord('ronddobberen')
```

```
True
```

```
>>> isHerhaalwoord('ballenbakken', minimale_herhaling=3)
```

```
False
```

```
>>> herhaalwoorden('woorden.txt', minimale_herhaling=2, minimale_lengte=10)
```

```
{'doordringing', 'apenpakken'}
```