

Algemeen

Newlines bij het inlezen van een bestand

Als Python de volgende regel moet lezen uit een tekstbestand, dan wordt er gelezen tot en met de eerstvolgende newline ('`\n`') of tot het einde van het bestand. Het laatste karakter van de regel die werd gelezen zal dus doorgaans een newline zijn (tenzij er na de laatste regel geen newline meer stond).

Om het newline karakter achteraan de regel te verwijderen, kan je gebruik maken van de stringmethode `rstrip`. Indien er geen argumenten doorgegeven worden aan deze methode, dan zal de methode alle witruimtekarakters (spaties, tabs en newlines) op het einde van de regel verwijderen. Om er zeker van te zijn dat er enkel newlines verwijderd worden op het einde van de regel, kan je het newlinekarakter doorgegeven als argument aan de methode `rstrip`.

```
>>> regel = bestand.readline()
>>> regel
'Dit is een regel in het bestand.\n'
>>> regel.rstrip('\n')
'Dit is een regel in het bestand.'
```

Tekstbestanden kopiëren naar Eclipse

Als je je oplossingen voor opgaven met tekstbestanden lokaal wilt uittesten, dan is het noodzakelijk dat je ook de tekstbestanden lokaal beschikbaar hebt. Anders kunnen de testgevallen uit de doctest deze tekstbestanden niet vinden. De tekstbestanden die in de gegeven doctest gebruikt worden, staan altijd gekoppeld in de omschrijving boven de doctest. Je kan de inhoud van deze bestanden in je browser bekijken door op de koppeling te klikken.

De meest generieke procedure om een lokale kopie van deze bestanden ter beschikking te hebben in Eclipse verloopt als volgt:

- open het bestand in je browser
- kopieer de inhoud van het bestand (**CTRL-A** + **CTRL-C**)
- maak een nieuw tekstbestand in Eclipse
 - klik rechts op de directory waarin het tekstbestand moet geplaatst worden (je moet er voor zorgen dat het bestand in dezelfde directory geplaatst wordt waar ook je Python script staat)
 - kies het menu item **New** en daarna het menu item **File**
 - geef de correct naam in van het bestand onder **File name**; let hierbij ook op het feit dat je de bestandsextensie moet opgeven (doorgaans `.txt`)
- plak de inhoud van het klembord in het bestand (**CTRL-V**)

Het onderstaande screenshot helpt je alvast op weg.

Als je je oplossingen indient via Dodona, dan zorgt Dodona er zelf voor dat de nodige tekstbestanden beschikbaar zijn in de directory waarin ook het Python script geplaatst wordt.

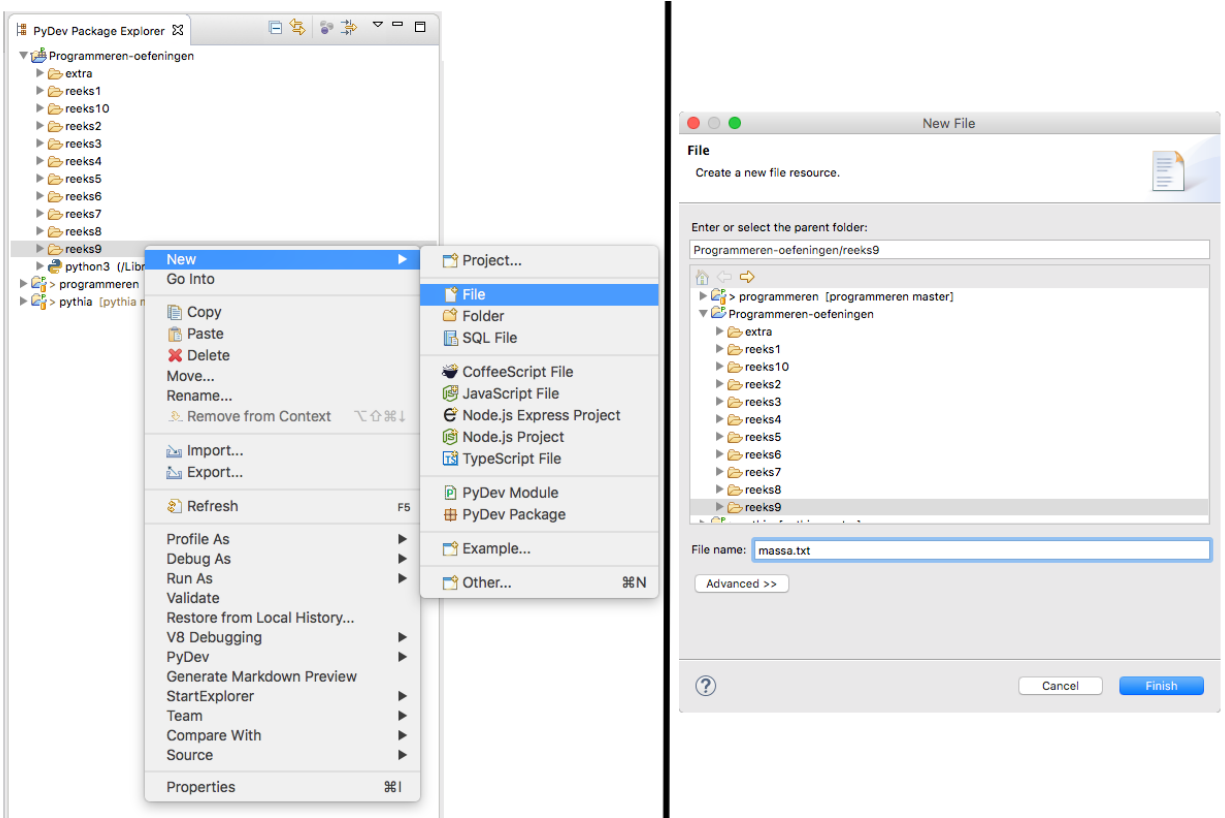


Figure 1: menu nieuw bestand

FASTQ ontdubbeling

Algemene info

Opmerkingen

Resultaten wegschrijven naar een bestand

De ingebouwde functie `print` kan gebruikt worden om de stringvoorstelling van een resultaat weg te schrijven naar een bestand. Hiervoor kan je gebruik maken van de optionele parameter `file` van de functie.

Standaard schrijft de functie `print` het resultaat naar het speciale bestand `sys.stdout` (de standaardwaarde van de parameter `file`) dat bijvoorbeeld aan de Console van Eclipse gekoppeld is. Hetzelfde effect kan je bekomen door de waarde `None` door te geven aan de parameter `file`.

Door een bestandsobject dat geopend werd om te schrijven door te geven aan de parameter `file` van de functie `print`, wordt de stringvoorstelling van het resultaat naar dit bestand weggeschreven.

```
>>> regel1 = 'Dit is de eerste regel.'
>>> regel2 = 'Dit is de tweede regel.'
>>> print(regel1)
Dit is de eerste regel.
>>> print(regel2, file=None)
Dit is de tweede regel.

>>> uitvoer = open('uitvoer.txt', 'w')
>>> print(regel1, file=uitvoer)
>>> print(regel2, file=uitvoer)
>>> uitvoer.close()
>>> bestand = open('uitvoer.txt', 'r')
>>> bestand.readline()
'Dit is de eerste regel.\n'
>>> bestand.readline()
'Dit is de tweede regel.\n'
>>> bestand.readline()
''
```

Specifieke info

De informatie in het tekstbestand staat bij deze oefening gesplitst over meerdere aaneensluitende regels. Daardoor werk je hier best met een bufferlijst. Je leest eerst 4 lijnen in en stopt die in een lijst. Pas daarna verwerk je die 4 lijnen alvorens je verder gaat naar de volgende 4.