Computerphysik Programmiertutorial 11

Prof. Dr. Matteo Rizzi und Dr. Markus Schmitt - Institut für Theoretische Physik, Universität zu Köln

Github: https://github.com/markusschmitt/compphys2022

Inhalt dieses Notebooks: Dateien lesen und schreiben "zu Fuß", Dateien lesen und schreiben mit `JLD`, `do`-Block Syntax, Pipelines von Funktionen, Unicode Zeichen

Dateien lesen und schreiben "zu Fuß"

Bisher haben wir gesehen wie man `DataFrame`s aus `.csv`-Dateien auslesen oder in `.csv`-Dateien schreiben kann.

Julia bietet auch die Möglichkeit Daten unstrukturiert in Dateien zu schreiben.

Die Funktion `write(<Dateiname>, <Variable>)` schreibt den Wert von `<Variable>` in die Datei `<Dateiname>`.

In [1]:
```
write("meine_datei.txt", "Hallo!")
```

Out[1]:
```
6
```

Falls die Datei bereits existiert, wird sie überschrieben:

In [2]:
```
write("meine_datei.txt", "Hallo nochmal!")
```

Out[2]:
```
14
```

Es können auch mehrere Variablen hintereinander geschrieben werden:

In [3]:
```
write("meine_datei.txt", "Hallo nochmal!", "13")
```

Out[3]:
```
16
```

Der Rückgabewert von `write()` gibt an wie viele Bytes geschrieben wurden.

Dateien explizit öffnen

Die Funktion `open(<Dateiname>, <(Schreib-/Lese-)modus>)` öffnet eine Datei und gibt einen `IOStream` zurück. Ein `IOStream` ist eine Datenstruktur, die die geöffnete Datei "verwaltet" und uns ermöglicht in der Datei zu schreiben/lesen.

Bei Öffnen können wir den Schreib-/Lesemodus wählen. Optionen sind:

- `"w"`: Schreiben und Datei überschreiben, falls sie bereits existiert.
- `"a"`: Schreiben und an Datei anhängen, falls sie bereits existiert.
- `"r"`: Lesen.

In [4]:
```
io = open("meine_datei.txt", "w")
write(io, "Hallo!")
close(io)
```

Überschreiben:

In [5]:
```
io = open("meine_datei.txt", "w")
write(io, "Nochmal hallo!")
close(io)
```

Anhängen:

In [6]:
```
io = open("meine_datei.txt", "a")
write(io, " Ich habe noch mehr mitzuteilen")
close(io)
```

In die geöffnete Datei können wir immer weitere Daten schreiben, indem wir `write()` wiederholt aufrufen. Hier schreiben wir eine Reihe von Zufallszahlen in die Datei, jede gefolgt von dem Zeichen `\n`, das ein Zeilenende markiert.

In [7]:
```
io = open("meine_datei.txt", "w")

for i in 1:10
    write(io, "$(rand())\n")
end

close(io)
```

Dateien auslesen

Der Inhalt einer Datei kann mit `read(<Dateiname>, <Datentyp>)` ausgelesen werden. Dabei muss neben dem Dateinamen der `<Datentyp>` angegeben werden. Die Datei liegt auf der Festplatte als eine Folge von Bits, die Julia nur sinnvoll interpretieren kann, wenn wir vorgeben um welchen Datentyp es sich handelt.

In [8]:
```
read("meine_datei.txt", String)
```

Out[8]:
```
"0.8130862265196095\n0.0009358403671910764\n0.12026119337042807\n0.8541726684413435\n0.4643526856048913\n0.2990466184587537\n0.8663329140605435\n0.8711029963
245596\n0.8011191654992112\n0.7382039209285901\n"
```

Die Funktion `readline(<IOStream>)` liest die Datei nur bis zum nächsten Zeilenende, das mit `\n` markiert ist

In [9]:
```
io = open("meine_datei.txt", "r")

readline(io)
```

Out[9]:
```
"0.8130862265196095"
```

Wenn wir eine Datei z.B. mit `readline` nur Stückweise auslesen, "merkt sich" die `IOStream` Datenstruktur die Stelle, bis zu der wir gelesen haben (mit einem "file pointer"). Beim nächsten Aufruf wird dann von dort weitergelesen:

In [10]:
```
readline(io)
```

Out[10]:
```
"0.0009358403671910764"
```

In [11]:
```
close(io)
```

Die Funktion `eof(<IOStream>)` testet ob beim Auslesen das Ende einer Datei erreicht ist. Das können wir nutzen um eine Datei vollständig zeilenweise auszulesen:

In [12]:
```
io = open("meine_datei.txt", "r")

while !eof(io)
    println("nächste Zeile:")
    println(readline(io))
end

close(io)
```

```
nächste Zeile:
0.8130862265196095
nächste Zeile:
0.0009358403671910764
nächste Zeile:
0.12026119337042807
nächste Zeile:
0.8541726684413435
nächste Zeile:
0.4643526856048913
nächste Zeile:
0.2990466184587537
nächste Zeile:
0.8663329140605435
nächste Zeile:
0.8711029963245596
nächste Zeile:
0.8011191654992112
nächste Zeile:
0.7382039209285901
```

Binäre Dateien

Bisher haben wir nur `String`s in Dateien geschrieben. Tatsächlich schreibt `write` die Daten einfach in Binärformat in die Datei. Daher können wir Variablen von beliebigen Datentypen auf die Festplatte schreiben:

In [13]:
```
io = open("meine_datei.txt", "w")

for i in 1:10
    r = rand()
    println(r)
    write(io, r)
end

close(io)
```

```
0.6707462765727293
0.0371559318032757
0.3247081799524083
0.5106944923499416
0.7223203370653651
0.7281796964303447
0.9619154319666288
0.004746582730648874
0.733127274170448
0.9886304082114303
```

Diese Zufallszahlen vom Typ `Float64` können wir auch der Reihe nach wieder auslesen:

In [14]:
```
io = open("meine_datei.txt", "r")

while !eof(io)
    println("nächste Zahl:")
    println(read(io, Float64))
end

close(io)
```

```
nächste Zahl:
0.6707462765727293
nächste Zahl:
0.0371559318032757
nächste Zahl:
0.3247081799524083
nächste Zahl:
0.5106944923499416
nächste Zahl:
0.7223203370653651
nächste Zahl:
0.7281796964303447
nächste Zahl:
0.9619154319666288
nächste Zahl:
0.004746582730648874
nächste Zahl:
0.733127274170448
nächste Zahl:
0.9886304082114303
```

Achtung: Beim Auslesen muss der Datentyp natürlich mit dem Datentyp übereinstimmen, der geschrieben wurde!

Beispiel: Schreibe `String`s

In [15]:
```
io = open("meine_datei.txt", "w")

for i in 1:10
    r = rand()
    write(io, "$r\n")
    println(r)
end

close(io)
```

```
0.30995575978135215
0.27879395542153906
0.9202646998683773
0.482157509143462
0.2991069772074545
0.13091154843632524
0.24033972892032485
0.1651149193894189
0.4106996248101771
0.18741061965045558
```

Lese `Float64`s

In [16]:
```
io = open("meine_datei.txt", "r")

while !eof(io)
    println(read(io, Float64))
end

close(io)
```

```
2.215839804602544e-52
2.003765860008307e-52
8.152382518367192e-43
3.378614040071563e-57
1.803886467200996e-259
3.5360344393164876e-57
8.618356541313796e-43
8.977699424452452e-67
2.743844622858298e-57
5.598611887746306e-67
1.0413769984073557e-42
1.723853075431545e-259
9.74809425586754e-72
7.126022557818426e-67
2.8963080887447744e-57
4.6600099728527127e-33
1.725122117756399e-259
2.7388783742347244e-57
1.1446103173551772e-71
1.4857517366700998e-76
1.369891693170581e-71
3.225229045713317e-86
1.2558104906349353e-71
2.214200513075369e-52
```

```
EOFError: read end of file

Stacktrace:
 [1] read(s::IOStream, T::Type{Int64})
   @ Base ./iostream.jl:412
 [2] read(s::IOStream, #unused#::Type{Float64})
   @ Base ./iostream.jl:421
 [3] top-level scope
   @ ./In[16]:4
 [4] eval
   @ ./boot.jl:373 [inlined]
 [5] include_string(mapexpr::typeof(REPL.softscope), mod::Module, code::String, filename::String)
   @ Base ./loading.jl:1196
```

Es können auch Arrays in Dateien geschrieben werden:

In [17]:
```
A = rand(3,4,2)
```

Out[17]:
```
3×4×2 Array{Float64, 3}:
[:, :, 1] =
 0.673448  0.369113  0.23188   0.0831212
 0.484072  0.882921  0.556548  0.9515
 0.926168  0.94151   0.792746  0.313989

[:, :, 2] =
 0.236231  0.901826  0.664468  0.209012
 0.22784   0.835508  0.114639  0.175087
 0.938654  0.178559  0.589552  0.100255
```

In [18]:
```
io = open("meine_datei.txt", "w")
write(io, A)
close(io)
```

Zum Auslesen legen wir ein leeres Array `B` mit gleicher Größe und gleichem Datentyp an und verwenden die `read!()` Funktion:

In [19]:
```
B = Array{Float64,3}(undef,3,4,2)

io = open("meine_datei.txt", "r")
read!(io,B)
close(io)
```

In [20]:
```
isapprox(A,B)
```

Out[20]:
```
true
```

Daten lesen/schreiben mit `JLD`

Mit dem Paket `JLD` können wir beliebige Datenstrukturen sehr einfach in Dateien schreiben und auslesen (Dokumentation).

In [21]:
```
using JLD
```

Daten verschiedener Datenstrukturen speichern:

In [22]:
```
t = 15
z = Dict("a" => 1, "b" => "petra")
save("meine_andere_datei.jld", "t", t, "z", z, "array", A)
```

Gespeicherte Daten laden:

In [23]:
```
d = load("meine_andere_datei.jld")
```

Out[23]:
```
Dict{String, Any} with 3 entries:
  "array" => [0.673448 0.369113 0.23188 0.0831212; 0.484072 0.882921 0.556548 0…
  "t"     => 15
  "z"     => Dict{String, Any}("b"=>"petra", "a"=>1)
```

In [24]:
```
d["t"]
```

Out[24]:
```
15
```

`do`-Block Syntax

Der `do`-Block ist eine alternative Syntax um einer Funktion als erstes argument eine anonyme Funktion zu übergeben.

Ein Beispiel ist die `map` funktion, die ihr erstes Argument (eine Funktion) elementweise auf das Array anwendet, das als zweites Argument übergeben wird:

In [25]:
```
arr = [17, -16, 0]

map(x->begin
           if x < 0 && iseven(x)
               return 0
           elseif x == 0
               return 1
           else
               return x
           end
       end,
    arr)
```

Out[25]:
```
3-element Vector{Int64}:
 17
  0
  1
```

Mit der `do`-Block Syntax kann das gleich in der folgenden Form geschieben werden:

In [26]:
```
map(arr) do x
    if x < 0 && iseven(x)
        return 0
    elseif x == 0
        return 1
    else
        return x
    end
end
```

Out[26]:
```
3-element Vector{Int64}:
 17
  0
  1
```

Die `do`-Block Syntax wird häufig zum Öffnen von Dateien verwendet, da die `open()`-Funktion die Datei dann automatisch wieder schließt:

In [27]:
```
open("meine_datei.txt", "w") do io
    for i in 1:10
        write(io, "$(rand())\n")
    end
end
```

In [28]:
```
open("meine_datei.txt", "r") do io
    while !eof(io)
        println(readline(io))
    end
end
```

```
0.19662946656929148
0.13587259858781542
0.47558398830806314
0.5462546878133612
0.8790694649334283
0.7975326636617645
0.49270072337851456
0.13526016336912017
0.7468732579123115
0.7787905801101316
```

Pipelines von Funktionen

Verschachtelte Funktionsaufrufe können alternativ als sogenannte *pipeline* geschrieben werden.

In [29]:
```
f(x) = 3*x + 5
g(x) = x^2.4 - x
```

Out[29]:
```
g (generic function with 1 method)
```

Verschachtelter Funktionsaufruf:

In [30]:
```
f(log(g(13)))
```

Out[30]:
```
23.383755485327974
```

Das gleiche mit dem pipeline-Operator `|>`:

In [31]:
```
g(13) |> log |> f
```

Out[31]:
```
23.383755485327974
```

Eine weitere Möglichkeit ist, den Verknüpfungsoperator `∘` zu verwenden:

In [32]:
```
(f ∘ log ∘ g)(13)
```

Out[32]:
```
23.383755485327974
```

In [33]:
```
fun = f ∘ log ∘ g
```

Out[33]:
```
f ∘ log ∘ g
```

In [34]:
```
fun(13)
```

Out[34]:
```
23.383755485327974
```

Unicode Zeichen

Wie oben gesehen sind Unicode Zeichen, wie z.B. `∘` als Teil der Julia Syntax erlaubt.

Diese Zeichen können in Julia Notebooks erzeugt werden, indem man den entsprechenden LaTex code eingibt und anschließend "tab" drückt.

Einige Beispiele:

Bezeichne eine Variable als `α`

In [35]:
```
α = 27
```

Out[35]:
```
27
```

In [36]:
```
α
```

Out[36]:
```
27
```

Das `∈`-Zeichen kann `in` ersetzen

In [37]:
```
for j ∈ 1:3
    println(j)
end
```

```
1
2
3
```

Den Vergleichsoperator `>=` können wir schreiben als `≥`

In [38]:
```
if α ≥ 9
    println("α≥9")
end
```

```
α≥9
```

Der Operator `≈` ist äquivalent zur `isapprox()` Funktion:

In [39]:
```
A≈B
```

Out[39]:
```
true
```

Warnung:

1. Unicode Zeichen sind in anderen Editoren unter Umständen nicht so leicht einzufügen.
2. Die großzügige Verwendung von Unicode Zeichen ist eine Besonderheit von Julia. In vielen anderen Programmiersprachen ist das nicht möglich.