

Übung 9

Arbeitsaufwand insgesamt: 4h

Inhaltsverzeichnis

Übung 9	1
Teil 1 – FASTA-Reader	2
Lösungsidee:	2
Lösung:	3
Testfälle:	4

Teil 1 – FASTA-Reader

Ziel der Übung ist das Implementieren eines FASTA-Readers mit den im Dokument vorgegebenen Funktionen.

Lösungsidee:

Da das Dokument allgemein relativ verwirrend war, möchte ich die Lösungsidee damit starten, dass ich genau festlege, was die folgenden Funktionen tun sollen:

- `initialize(fasta_t& fasta)` – Dadurch dass kein zusätzlicher Parameter gegeben ist, wird nur die `fasta_t` Datenstruktur mit `clear(fasta)` aufgerufen und somit geleert. Eine eigene Funktion dafür anzulegen finde ich aber ziemlich sinnbefreit, weil es grundsätzlich nur Codeverdopplung unter einem anderen Funktionsnamen ist.
- `split(...)` – Ich habe es so verstanden, dass `Split` nach dem Einlesen des FASTA-Files die Entries in verschiedene Files schreibt, anstatt nur in ein einziges (konträr zu `write_to_file`). Alternativ hätte man mit `Split` auch ein File aufteilen können, bevor man die Datenstruktur daraus befüllt.
- `read_from_stream(...)` – Der Formalparameter “max” wurde als Anzahl der einzulesenden Entries betrachtet. Dass die Funktion als FSM (Finite-State-Machine) operieren soll, habe ich verstanden. Ich würde aber noch einige Pfeile (z.B. Sequenz -> Kommentar) hinzufügen, da sonst potentiell wichtige Daten verloren gingen.

Zudem habe ich als Returnwert für die Stream-bezogenen Funktionen einfach den Stream der als Formalparameter vorliegt wieder zurückgegeben. Ich persönlich hätte eher „void“ als Returnwert verwendet, weil es in dieser Situation in meinen Augen keinen Sinn macht, weil der Stream ohnehin bereits außerhalb der Funktion existiert (immerhin wurde er als Aktualparameter mitgegeben).

Nun zur grundsätzlichen Lösungsidee des Algorithmus:

Da der Algorithmus anhand einer FSM (Finite-State-Machine) funktionieren soll, gibt es 4 Zustände:

- Outside
- Header
- Comment
- Sequence

Was in den jeweiligen Phasen geschehen kann, wird auf der nächsten Seite in Form einer Tabelle erläutert.

Dazu gibt es folgendes Verhalten:

State	Verhalten
Outside	Es wird auf einen Header gewartet. Alles andere wird ignoriert. Wird ein Header gefunden, so wird dieser abgespeichert und der State auf „Header“ gewechselt.
Header/ Comment / Sequence	Grundsätzlich wird immer auf den ersten Buchstaben der Zeile geschaut: <ul style="list-style-type: none">• Wird wieder ein „>“ gefunden, so wird der momentane Entry (Datenstruktur) (ohne Kommentar oder Sequence) in der überliegenden Datenstruktur gespeichert. Außerdem wird ein neuer Entry angelegt, der Header gesetzt und der State auf „Header“ gesetzt.• Wird ein „;“ gefunden, so wird dieses im Vektor des Entry abgespeichert. State wird auf „Comment“ gesetzt.• Wird ein dNTP (AGTC) gefunden, so wird die Sequenz in den Entry gespeichert bzw. angehängt und der State auf „Sequence“ gesetzt.• Wird eine Leerzeile oder jegliche andere Zeile gefunden wird der Entry in der überliegenden Datenstruktur gespeichert. State wird auf „Outside“ gesetzt.

Lösung:

Als C++ Projekt „Teil_1“ im Archiv

Testfälle:

- Leeres Eingabefile
- Ungültige Eingabe bei Modusauswahl
- Beliebige Eingaben für die einzulesenden Entries („max“)
- Ungültige Eingabe bei Output Auswahl
- Outputfile existiert bereits
- Eingabe einer ungültigen Zeilenlänge („len“)
- Eingabe von sinnlosen Werten bei Split-Menge („n“)
- Output in File
- Output auf Konsole
- Split der Datei

Leeres Eingabefile:

```
Which file do you want to read from?  
nichtexistent.txt  
File not found! Please enter valid file:
```

Ungültige Eingabe bei Modusauswahl:

```
Which file do you want to read from?  
test.fasta  
  
What do you want to do?  
(1) Output the file.  
(2) Split the file.  
3  
Please select one of the above: 5  
Please select one of the above: a  
Please enter a valid number: @  
Please enter a valid number: 10  
Please select one of the above:
```

Beliebig Eingaben für die einzulesenden Entries („max“):

```
Which file do you want to read from?  
test.fasta
```

```
What do you want to do?
```

```
(1) Output the file.
```

```
(2) Split the file.
```

```
1
```

```
How many entries do you want to read? (0 = all)
```

```
1
```

```
Where do you want to output the data?
```

```
(1) Output to file.
```

```
(2) Output to console.
```

```
2
```

```
>DRB5*02:03
```

```
Which file do you want to read from?  
test_smol.fasta
```

```
What do you want to do?
```

```
(1) Output the file.
```

```
(2) Split the file.
```

```
1
```

```
How many entries do you want to read? (0 = all)
```

```
2000
```

```
Where do you want to output the data?
```

```
(1) Output to file.
```

```
(2) Output to console.
```

```
2
```

```
>DRB5*02:03
```

```
;some comment
```

```
CATGTTTTCTTGACAGCAGGATAAGTATGAGTGTCATTTCTTCAACGGGACGGAGCGGGTGCGGTTCTGCACAGAGGCATC  
TATAACCAAGAGGAGAACGTGCGCTTCGACAGCGACGTGGGGGAGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGACGC  
TGAGTACTGGAACAGCCAGAAGGACATCCTGGAGCAGGCGCGGGCCGCGGTGGACACCTACTGCAGACACAACCTACGGGG  
CTGTGGAGAGCTTCACAGTGCAGCGGCGAG
```

```
>DRB5*02:03
```

```
TGAGTACTGGAACAGCCAGAAGGACATCCTGGAGCAGGCGCGGGCCGCGGTGGACACCTACTGCAGACACAACCTACGGGG  
CTGTGGAGAGCTTCACAGTGCAGCGGCGAG
```

```
>DRB5*02:03
```

```
TGAGTACTGGAACAGCCAGAAGGACATCCTGGAGCAGGCGCGGGCCGCGGTGGACACCTACTGCAGACACAACCTACGGGG
```

Ungültige Eingaben bei der Output Auswahl:

```
Which file do you want to read from?
test.fasta

What do you want to do?
(1) Output the file.
(2) Split the file.
1

How many entries do you want to read? (0 = all)
0

Where do you want to output the data?
(1) Output to file.
(2) Output to console.
3
Please select one of the above: 1
Please enter a valid number: p
Please enter a valid number: 20
Please select one of the above:
```

Outputfile existiert bereits:

```
Which file do you want to read from?
test.fasta

What do you want to do?
(1) Output the file.
(2) Split the file.
1

How many entries do you want to read? (0 = all)
0

Where do you want to output the data?
(1) Output to file.
(2) Output to console.
1

Please enter the output filename:
test.fasta
This will overwrite an existing file.
Continue? (y/n)
```

Eingabe einer ungültigen Zeilenlänge („len“)

```
test.fasta

What do you want to do?
(1) Output the file.
(2) Split the file.
1

How many entries do you want to read? (0 = all)
0

Where do you want to output the data?
(1) Output to file.
(2) Output to console.
1

Please enter the output filename:
test1.fasta

Please enter the linelength: (20-255) (enter 0 for default value)
19
Please enter a valid number: 256
Please enter a valid number: 19
Please enter a valid number: 0
```




Eingabe von sinnlosen Werten bei Split-Menge („n“):

```
Which file do you want to read from?
test.fasta

What do you want to do?
(1) Output the file.
(2) Split the file.
2

Please enter the amount of files to split into:
-1
Value must at least be between 2 and 10: 1
Value must at least be between 2 and 10: 11
Value must at least be between 2 and 10: 𐀀
Please enter a valid number: 3

Please enter the linelength: (20-255) (enter 0 for default value)
0
```

	test_split-1.fasta	06.02.2020 13:48	FASTA-Datei	1 KB
	test_split-2.fasta	06.02.2020 13:48	FASTA-Datei	1 KB
	test_split-3.fasta	06.02.2020 13:48	FASTA-Datei	1 KB

Output in File:

```
Which file do you want to read from?
ti01_HLA_DRBE2r.fna

What do you want to do?
(1) Output the file.
(2) Split the file.
1

How many entries do you want to read? (0 = all)
20

Where do you want to output the data?
(1) Output to file.
(2) Output to console.
1

Please enter the output filename:
ti01_out.fna

Please enter the linelength: (20-255) (enter 0 for default value)
100
```

Das File steht auch im Archiv zur Verfügung. Mithilfe von Notepad++ kann auch schnell die Anzahl der Zeichen pro Zeile kontrolliert werden.

Output auf Konsole:

```
Which file do you want to read from?
drb-ex2.fasta

What do you want to do?
(1) Output the file.
(2) Split the file.
1

How many entries do you want to read? (0 = all)
2

Where do you want to output the data?
(1) Output to file.
(2) Output to console.
2

>DRB5*02:02
CATGTTTCTTGACAGCAGGATAAGTATGAGTGTCATTTCTTCAACGGGACGGAGCGGGTGCGGTTCTGCACAGAGGCATC
TATAACCAAGAGGAGAACTGCGCTTCGACAGCGACGTGGGGGAGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGACGC
TGAGTACTGGAACAGCCAGAAGGACATCCTGGAGCAGGCGCGGGCCGCGGTGGACACCTACTGCAGACACAACCTACGGGG
CTGTGGAGAGCTTCACAGTGCAGCGGCGAG

>DRB5*01:01:01
CACGTTTCTTGACAGCAGGATAAGTATGAGTGTCATTTCTTCAACGGGACGGAGCGGGTGCGGTTCTGCACAGAGACATC
TATAACCAAGAGGAGGACTTGCGCTTCGACAGCGACGTGGGGGAGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGACGC
TGAGTACTGGAACAGCCAGAAGGACTTCCTGGAAGACAGGCGCGCCGCGGTGGACACCTACTGCAGACACAACCTACGGGG
TTGGTGAGAGCTTCACAGTGCAGCGGCGAG
```

Split der Datei:

```
Which file do you want to read from?
drb-ex2.fasta

What do you want to do?
(1) Output the file.
(2) Split the file.
2

Please enter the amount of files to split into:
10

Please enter the linelength: (20-255) (enter 0 for default value)
100
```

Die Files liegen wiederum im Archiv zur Kontrolle.