

# Einführung in die Angewandte Bioinformatik

---

## Kurzeinführung in Unix und verwandte Betriebssysteme

Webseite zur Vorlesung  
<http://bioinfo.wikidot.com/>

Sprechstunde  
Mo 16-17 in OH14, R214  
Sven.Rahmann -at- tu-dortmund.de

# Begriffserklärungen

## Hardware

Oberbegriff für die maschinentechnische Ausrüstung eines Systems - („kann man anfassen“).

## Software

Zusammenfassender Begriff für Programme und Daten in Computern - Dazu gehören Betriebssysteme, Utilities, Anwendungsprogramme.

## Betriebssystem (operating system, OS)

Software, die die Verwendung (den Betrieb) eines Computers ermöglicht. Verwaltet Betriebsmittel wie Speicher, Ein- und Ausgabegeräte. Steuert die Ausführung von Programmen. Beispiele: Linux, Unix, Solaris, Windows, Mac OS, ...

## Arbeitsumgebung und Window-Manager (z.B. KDE mit KWin)

Grundlegende Funktionen für die Arbeit mit dem Computer  
Grundlegende Fensterfunktionen (Minimieren, Vergrößern, Schließen)

## Shell

Textorientierte Eingabe-Schnittstelle zwischen Computer und Benutzer, erlaubt die Eingabe von Befehlen

# Begriffserklärungen

## KDE Desktop mit Dateibrowser (Konqueror) und Shell

The screenshot displays a KDE desktop environment. The desktop background is blue with several icons: Trash, Home, Thunderbird, Firefox, NetBeans IDE 6.0, and Konsole. The Konqueror file browser window is open, showing the directory `/home/rahmann/teaching/ss2008/EinfuehrungAngewandteBioinformatik`. The file list is as follows:

Name	Size	File Type	Modified	Permissions	Owner	Group
accounts.txt	4.5 KB	Plain Text Document	07.02.2008 13:33	-rw-rw----	rahmann	ls11
hilfskraft.doc	100.0 KB	Microsoft Word Document	05.03.2008 17:30	-rw-r--r--	rahmann	ls11
literatur.txt	320 B	Plain Text Document	18.03.2008 15:21	-rw-rw----	rahmann	ls11
termine.txt	2.3 KB	Plain Text Document	19.03.2008 18:01	-rw-rw----	rahmann	ls11
termine.txt~	251 B	Backup File	19.03.2008 17:12	-rw-rw----	rahmann	ls11

The Shell - Konsole window is also open, showing the command `ls | wc` and its output:

```
rahmann@waterman:~$ ls | wc
 29  29 293
rahmann@waterman:~$
```

# Begriffserklärungen

**Browser** (to browse: durchstöbern)

Programm, um Informationen übersichtlich zu betrachten, z.B.

Dateibrowser zeigt Dateien in einem Verzeichnis an.

Webbrowser zeigt Seiten aus dem Internet an.

**Account, home, login, password**

Solaris (das Betriebssystem der Firma Sun) ist ein Mehrbenutzersystem, in dem jeder seinen eigenen Arbeitsbereich hat (home-Verzeichnis).

Damit nur Sie auf Ihre Daten zugreifen können, erhält jeder eine Benutzerkennung und dazu ein geheimes Passwort.

**Verzeichnis** (directory, Ordner)

ein „Behälter“ für verschiedene Dateien

Man benutzt eine hierarchische Ordnung, um Ordnung zu halten

So entsteht ein Verzeichnisbaum.

# Arbeiten mit der Shell

U.U. ist es sinnvoll, jeden Übungszettel in einem eigenen Verzeichnis zu bearbeiten. Für einfache Verwaltungsaufgaben (Anlegen, Löschen von Verzeichnissen) können Sie:

- Einen Dateibrowser (Konqueror) verwenden
- Die Shell (Konsole) verwenden und Befehle eingeben

`pwd`: aktuelles Verzeichnis anzeigen

`mkdir`: neues Verzeichnis anlegen (`mkdir uebung1`)

`cd`: Verzeichnis wechseln (`cd uebung1`)

`cd` ohne Verzeichnisnamen wechselt in Ihr home-Verzeichnis

`rmdir`: Verzeichnis entfernen, wenn leer (`rmdir uebung1`)

Warnung: Löschen kann i.d.R. Nicht rückgängig gemacht werden

`ls`: Verzeichnisinhalt anzeigen

`ls -la`: Verzeichnisinhalt detailliert anzeigen

**Generelle Befehlssyntax:** Befehl Optionen Argumente

„Befehl“ ist der Name des Befehls („`ls`“)

„Optionen“ modifizieren das Verhalten eines Befehls, beginnen mit Minus (`-l -a`)

„Argumente“ geben an, worauf der Befehl angewendet werden soll;

häufig der Name eines Verzeichnisses oder einer Datei (oder mehrere)

# Philosophie hinter Unix-Befehlen

---

Jeder Befehl führt eine einfache, klar umrissene Aufgabe aus.  
Reichhaltige Optionen ermöglichen viele Variationen.  
Komplexe Aufgaben werden durch Aneinanderreihung von Befehlen möglich.

## Hilfe zu Befehlen und Optionen anzeigen lassen

`man befehlsname` („manual pages“)

z.B. `man ls` zeigt die Wirkung und alle möglichen Optionen von `ls`

# Datei-Befehle

## Verwendung von wildcards \* und ?

Häufig darf man statt einer einzelnen Datei eine Liste von Dateien angeben.

Statt diese explizit aufzulisten, kann man \* und ? verwenden:

\* steht für eine beliebige Folge von Zeichen

? steht für ein beliebiges Zeichen

z.B. `rm test*.fasta`

löscht alle Dateien, deren Name mit test anfängt und deren Endung .fasta ist.

Warnung! Löschen mit `rm` oder `rmdir` kann nicht rückgängig gemacht werden!

## Einfache Befehle zur Verwaltung Dateien

`cp Datei1 Datei2` – kopiert Datei1 nach Datei2

`cp Dateiliste Verzeichnis` – kopiert alle Dateien in der Liste ins Verzeichnis

`mv` – wie `cp`, aber löscht Original (move statt copy)

`rm Dateiliste` – löscht alle in der Liste angegebenen Dateien

Vorsicht: Niemals `rm *` ausprobieren!

# Datei-Inhalte

## Befehle zum Untersuchen von Dateien

`cat Dateiliste`: zeigt nacheinander die Inhalte aller Dateien an

`head -N 17 Dateiliste`: zeigt jeweils die ersten 17 Zeilen an

`tail -N 17 Dateiliste`: zeigt jeweils die letzten 17 Zeilen an

`more Dateiliste`: zeigt nacheinander die Inhalte aller Dateien an, wartet auf Tastendruck wenn der Bildschirm voll ist

## `grep` und `wc` (general regular expression matcher und word counter)

`grep Muster Dateiliste`

sucht nach Muster in allen Dateien, gibt alle Zeilen aus, in denen das Muster auftritt

Beispiel: Datei ist ein Telefonbuch, ein Eintrag mit Name, Telefonnummer pro Zeile.

Suche alle Einträge mit Namen Meier: `grep Meier telefonbuch.txt`

Das Muster kann viel komplizierter sein (reguläre Ausdrücke), z.B.

`grep M.er telefonbuch.txt`

Hier steht der Punkt für ein beliebiges Zeichen, man findet Meier, Mayer, Maler, ...

`wc Dateiliste`

gibt für jede Datei 3 Zahlen aus: Anzahl Zeilen, Wörter, Zeichen

# Ein- und Ausgabeumleitung, Pipes

## Ausgabeumleitung >

Man möchte die Meiers aus dem Telefonbuch in einer neuen Datei speichern.

Das > Zeichen leitet die Ausgabe in eine Datei um.

Achtung: Wenn die Datei schon existiert, wird sie überschrieben!

```
grep Meier telefonbuch.txt > meiers.txt
```

## Eingabeumleitung <

Analog kann man die Eingabe zu einem Programm

aus einer Datei (statt z.B. von der Tastatur beziehen) – mit <

## Pipe

Will man direkt die Ausgabe eines Programms als Eingabe eines anderen verwenden, kann man die Pipe (Rohr) | benutzen:

Zähle, wie viele Dateien im aktuellen Verzeichnis sind

```
ls | wc
```

Die Ausgabe von ls (Verzeichnisinhalt) wird direkt als Eingabe für wc verwendet.

Die Ausgabe von wc erscheint im Shell-Fenster.

# Das FASTA-Dateiformat

Biologische Sequenzdaten werden häufig im sogenannten FASTA-Format gespeichert. Dabei können Sequenzen in einer „Titelzeile“ mit zusätzlichen Informationen annotiert werden.

Eine Datei kann aus mehreren Sequenzen (samt Titelzeile) bestehen.

Die Titelzeile (header) beginnt dabei immer mit >

Darauf folgen Sequenzdaten bis zum nächsten header oder Dateiende.

## Beispiel

```
>Lieblingssequenz
```

```
ACGTTGCA
```

```
>andere Sequenz aus dem Internet
```

```
AAAAAAAAAA
```

```
AAAAAAAAAA
```

```
AAAAAAAAAT
```