

Teil 1: Betriebssysteme der anderen Art



Röbbe Wünschiers

Beim Kauf eines Computers braucht man sich eigentlich keine Gedanken um das Betriebssystem machen – was ist das überhaupt? – da fast alle Computer ein Windows-Betriebssystem vorinstalliert haben. Nur wenige Menschen greifen zu einem Apple-Computer mit dem Macintosh-Betriebssystem. Vielleicht hat der eine oder andere schon einmal etwas von Unix oder Linux gehört – als alternatives Betriebssystem kommt es sicher nur wenigen in den Sinn. Aber gerade Linux, eine kostenlose Variante von Unix, bietet insbesondere im wissenschaftlichen Arbeitsumfeld und im Bildungssektor eine Reihe von Vorzügen – und das nicht nur finanziell. Mit diesem Beitrag möchten wir eine Serie beginnen, die in das Arbeiten unter Unix-varianten Betriebssystemen (Linux, Knoppix, Mac OS X, FreeBSD, u.v.m.) einführt. Im Mittelpunkt steht dabei nicht die graphische Oberfläche (die unter Linux mittlerweile mindestens ebenso komfortabel wie unter Windows ist), sondern die Kommandozeile, die viele kleine nützliche Tools bietet. Zudem wird diese Serie in die Programmierung in der Shell und mit der leicht verständlichen Programmiersprache awk einführen. Jeder kann dann seine eigenen Befehle erzeugen und es steht mit awk ein leistungsfähiges Werkzeug zur Verarbeitung und Analyse von Daten zur Verfügung.

Grundlagen & Vorbereitungen

Das Ziel dieses ersten Teils ist es, einen Überblick zu verschaffen und die Grundlagen zu legen, damit Sie in den folgenden Monaten alle Beispiele verfolgen und vor allem ausprobieren können.

Wie wir sehen werden, muss man sich keinen neuen Computer kaufen, um mit Unix arbeiten zu können (stellvertretend für alle Unix-ähnlichen Betriebssysteme wie, Linux und Mac OS X schreibe ich von nun an Unix). Mit Knoppix steht eine Linux-Variante zur Verfügung, die noch nicht einmal auf Ihrem Computer fest installiert werden muss: Es kann von der CD aus gestartet (gebootet) werden. CygWin ist



Der Autor

Dr. Röbbe Wünschiers studierte Biologie in Marburg und promovierte dort über „Eigenschaften, Regulation und Funktion einer Hydrogenase im Wasserstoffmetabolismus der einzelligen Grünalge *Scenedesmus obliquus*“. Seit Januar 2002 ist R. Wünschiers an der Universität Köln, lehrt dort Bioinformatik und am Cologne University Bioinformatics Center (CUBIC) Genetik. Für die CLB schreibt er seit 1997.

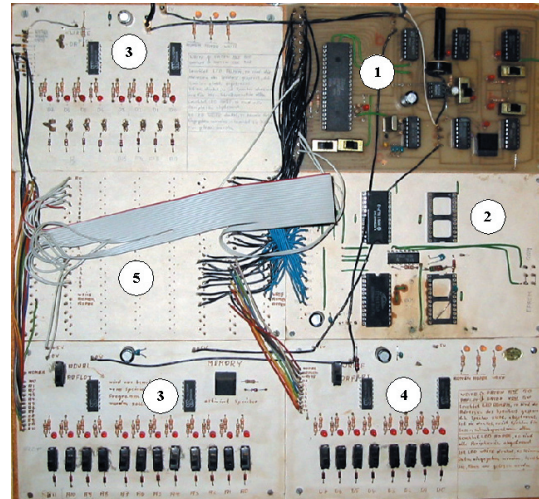


Abbildung 1: Hardware eines Computers. Dies sind die Grundeinheiten eines Computers mit (1) dem Prozessor, (2) dem Speicher, (3) der Adresseingabe, (4) der Datenein- und Ausgabe und (5) dem Bussystem, das alle Einheiten miteinander verbindet. Der Prozessor ist hier ein Zilog Z80 dessen Taktfrequenz zwischen 0-2 Mhz stufenlos eingestellt werden kann. Der Arbeitsspeicher ist 4 kB groß. Der Computer wurde Ende der 80er Jahre vom Autor gebaut und fand vor allem als Lehrmodell Anwendung.

ein Unix-Emulator, der unter Windows läuft. Dies ist besonders komfortabel, da man nicht entweder Windows oder Linux starten muss. Einige Freunde von Apple-Computern werden vielleicht überrascht sein, dass das neue Betriebssystem Mac OS X vollständig auf einer Unix-Version basiert und man daher direkt mit Unix arbeiten kann.

Was ist ein Betriebssystem?

Wie Windows so sind auch Unix, Linux und Mac OS X Betriebssysteme. Was bedeutet das? Um diese Frage zu beantworten ist es notwendig, einen kurzen Überblick über die Hardware (die physischen Bestandteile) eines Computers zu gewinnen. Das Herzstück eines Computers ist der Prozessor. Die meisten Computer haben heute Intel Pentium-Prozessoren. Der Prozessor führt die Rechenoperationen aus, die für das Funktionieren aller Programme notwendig sind.

Dazu nimmt er Daten auf und gibt das Ergebnis der Rechenoperation aus. Die Daten kommen meistens entweder aus dem Speicher oder von einer Dateneingabeeinheit wie der Tastatur. Der Speicher kann der

Arbeitsspeicher sein, die Festplatte, eine CD-ROM oder eine Diskette. Auf den Arbeitsspeicher kann der Prozessor am schnellsten zugreifen, allerdings wird er gelöscht, wenn er nicht mehr mit Strom versorgt wird – der Arbeitsspeicher ist ein flüchtiger Speicher, wie unser Kurzzeitgedächtnis. Die Festplatte, CD-ROM oder Diskette sind permanente Speicher, wie unser Langzeitgedächtnis. Daher sind hier sowohl wichtige Daten als auch Programme und das Betriebssystem gespeichert. Die Ausgabe der Daten erfolgt wiederum entweder in einen Speicher (besser: eine Datei im Speicher) oder auf den Bildschirm (siehe Abbildung 1).

Das Betriebssystem ist nun eine Sammlung von Programmen, die sowohl die Hardware eines Computers als auch die installierte Software, also Programme wie Office, eine Datenbankanwendung oder einen Proteinstrukturbetrachter, steuern. Im Folgenden werden wir uns mit Betriebssystemen beschäftigen, die Abkömmlinge von Unix sind.

Das Unix Betriebssystem

Anfang der sechziger Jahre waren Computer so groß wie Wandschränke, kosteten mindestens 50000 Dollar, konnten nur ein Programm nach dem anderen abarbeiten und nur von einem Benutzer pro Zeit verwendet werden. Dies war Ken Thompson und Dennies Ritchie von der amerikanischen Firma AT&T zu wenig. Von 1969 bis 1971 entwickelten sie Unix (uniplexed information and computer system). Erst 1974 veröffentlichten sie ihre Idee [1] und setzten damit seine Verbreitung in Gang (Abbildung 2). In der Einleitung ihrer Veröffentlichung schreiben sie: "Perhaps the most important achievement of UNIX is to demonstrate that a powerful operating system for interactive use need not be expensive either in equipment or human effort: UNIX can run on hardware costing as little as \$40,000, and less than two man-years were spent on the main system software." [1]. Unix entwickelte sich bald zum Standard-Betriebssystem für Großrechner und Workstations. Eigentlich muss man bei Unix von einer Betriebssystemfamilie sprechen, da fast jeder Großrechnerhersteller sein eigenes Unix heraus brachte. Von Anfang an herrschte bei Unix also die Vielfalt vor.

Linux

Linux wurde 1991 von dem Finnen Linus Torvalds entwickelt, damals Physikstudent an der Universität Helsinki. Die Ankündigung vom 25. August 1991 um 20:57:08 GMT in einer Internet Newsgroup ist noch heute abrufbar. Dort schreibt er: "I'm doing a (free) operating system (just a hobby, won't be big and professional like gnu) for 386(486) AT clones. This has been brewing since april, and is starting to get ready." Seither hat sich das Produkt seines Hobbys explosi-

onsartig ausgebreitet und beginnt sowohl dem guten alten Unix wie auch Windows den Rang abzulaufen.

Der Vor- aber auch Nachteil von Linux ist, dass es von Tausenden von „Freaks“ kostenlos weiterentwickelt wird. Der Vorteil ist, neben der Kostenfreiheit, dass Fehler schnell behoben werden. Das Problem ist nur, diese Updates im Internet zu finden. Außerdem ist es für Softwareentwickler schwierig vorausszusagen, welche Linux-Version der Anwender auf seinem Computer installiert hat. Daher gestaltet sich die Installation der Software oftmals schwierig: sie muss auf dem Zielrechner kompiliert werden. Dies wiederum setzt einige Kenntnis voraus.

Allerdings geht die Entwicklung seit einigen Jahren dahin, dass Firmen Linux-Distributionen vertreiben (etwa Suse oder Redhat), die einfach zu installieren sind und deren Applikationen alle aufeinander abgestimmt sind. Die Weiterentwicklung der Distributionen (z.B. Suse 6.0, Suse 7.3, Suse 8.0) sind somit zu vergleichen mit den unterschiedlichen Windows-Versionen (z.B. Win 3.1, Win 95, Win Me, Win XP).

Unix und Linux in der Chemie und Biologie

Unix war in gewisser Weise von Anfang an auch ein akademisches Betriebssystem, insbesondere seit der Weiterentwicklung von Unix an der Universität von Berkeley in Kalifornien/USA. Dementsprechend wurde Unix von Anfang an insbesondere in den Naturwissenschaften angewendet. So gibt es seit den sechziger Jahren Publikationen aus den Bereichen Chemoinformatik und Bioinformatik. Voraussetzung für diese Entwicklung war, neben dem Zugang zu einem leistungsfähigen Computer (z.B. der IBM 7090 Computer, der auch die amerikanischen Mercury and Gemini Raumflüge unterstützte) und einer verständlichen Programmiersprache (damals FORTRAN (formula translation)), dass chemische bzw. biologische Daten

Abbildung 2: Betriebssystem-Evolution. Schon früh hat sich Unix diversifiziert. Die wichtigsten Unix-Varianten sind BSD (Berkeley Software Distribution), Unix System V und Linux. Mit Mac OS X (durch den Kauf von NextStep durch Apple) ist vor zwei Jahren auch Apple auf Unix umgestiegen. Natürlich entwickeln sich alle Systeme weiter.

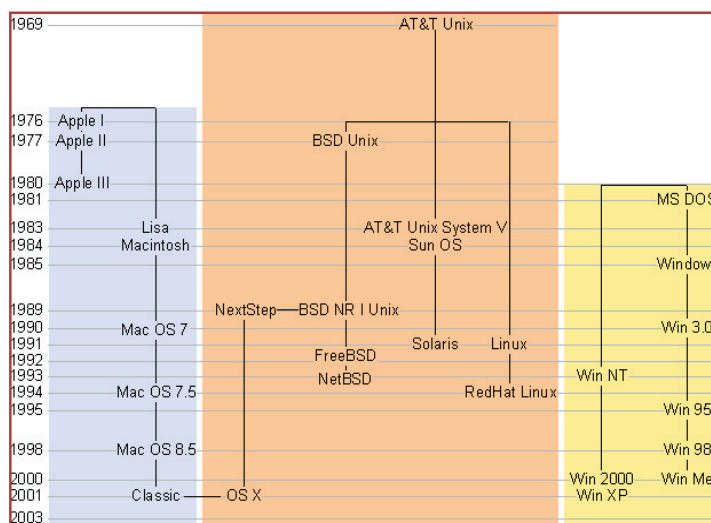
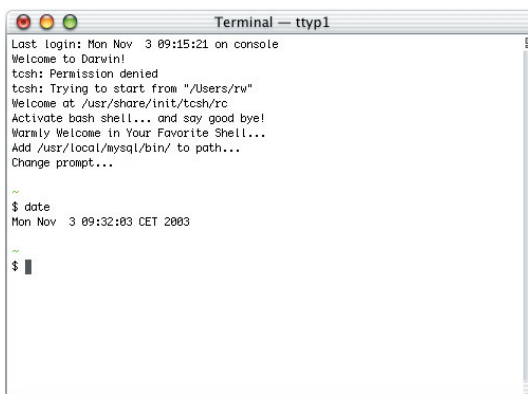


Abbildung 3: Das Terminal Fenster von Mac OS X. Aufgerufen wird es vom Finder über das Menü Go und den Eintrag Applications (deutsche Bezeichnungen im Text). Sie finden dann unter Utilities das Terminal. Das gezeigte Terminal des Verfassers ist bereits angepasst worden. Bei Ihnen wird es anders aussehen! Zu sehen ist das Ergebnis des Befehls `date`.



informatisch „fassbar“ waren. Die ersten Daten dieser Art waren Proteinsequenzen [2] und später Proteinstrukturdaten [3]. In dieser Zeit entstanden auch die ersten Algorithmen, um aus Proteinsequenzen die Evolution von Organismen zu rekonstruieren [4]. Molecular Design Ltd. (chemische Datenbanken, gegründet 1978), Health Design Inc. (toxikologische Vorhersagen, gegründet 1978), Tripos Associates Inc. (molecular modeling und drug design, gegründet 1979) und IntelliGenetics (DNA und Protein Sequenz Analyse, gegründet 1980) waren Pioniere bei der marktwirtschaftlichen Anwendung von Computern (mit Unix) im Bereich der Chemie und Biochemie. Es ist schon erstaunlich: die Grundlage zu dem, was heute als bioinformatische Revolution gefeiert wird, ist schon vor über 30 Jahren gelegt worden.

Linux, Knoppix, Mac OS X, CygWin – wer die Wahl hat...

Um die Beispiele in den kommenden Ausgaben nachvollziehen zu können, müssen Sie Zugang zu Unix haben. Dafür bieten sich eine Reihe unterschiedlicher Möglichkeiten und Systeme an. Wie bereits angesprochen, werden wir ausschließlich im Kommandozeilen-Modus, in der Konsole oder im Terminal, arbeiten.

Mac OS X

Das neue Appel-Betriebssystem Mac OS X (Macintosh Operating System 10) basiert vollständig auf Unix (Abbildung 2). Dies ist der Grund, warum alte Macin-

tosh-Programme nicht auf dem neuen Betriebssystem laufen. Die Kommandozeile von Mac OS X erreicht man über das Terminal (manchmal auch Konsole oder Shell genannt). Wie gelangt man in das Terminal? Öffnen Sie im Finder-Menü *Gehe zu* den Eintrag *Programme*. Sie finden dann unter *Dienstprogramme* das *Terminal*. Nach dem Starten des Terminals sollte ein Fenster wie in Abbildung 3 erscheinen. Das ist das Terminal und unsere zukünftige Spielwiese. Testen Sie, ob sie funktioniert. Geben Sie den Befehl `date` ein und drücken **ENTER**. Der Befehl `date` liefert das aktuelle Datum und druckt es auf den Bildschirm (siehe Terminal 1). Schließen Sie den Terminal mit dem Befehl `exit` oder der Tastenkombination **CTRL** + **D**.

Knoppix

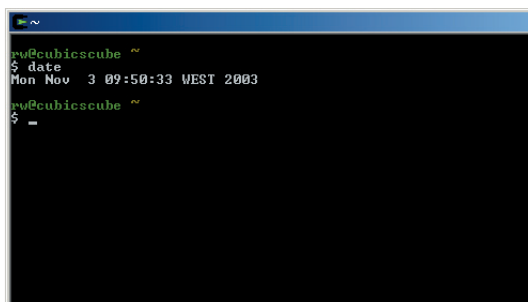
In den vergangenen Monaten ist Knoppix recht bekannt geworden. Knoppix ist eine vollständige Linux-Version, die auf eine CD-ROM komprimiert wurde. Das Tolle ist, dass Knoppix von der CD-ROM gestartet wird und sich dann lediglich im Arbeitsspeicher befindet. Dieser muss daher großzügig ausfallen: Wenigstens 254 MByte sollten es schon sein, um Spaß zu haben. Unsere Serie können Sie aber schon mit einem alten 486er Prozessor und 20 MByte Arbeitsspeicher verfolgen. Dies sind die minimalen Voraussetzungen für den Textmodus. Die Verwendung von Knoppix setzt voraus, dass der Computer zunächst versucht, von der CD-ROM zu booten, bevor er auf der Festplatte nach einem Betriebssystem sucht. Die Reihenfolge der Bootlaufwerke lässt sich im BIOS (Basic Input Output System) einstellen. Um in das BIOS-Setup zu gelangen, muss man kurz nach dem Starten des Computers eine Taste wie **DEL** oder **ESC** oder eine Tastenkombination wie **CTRL** + **ALT** + **ESC** oder **CTRL** + **ALT** + **S** drücken. Meistens wird die benötigte Tastenkombination kurz angezeigt.

Verstellen Sie im BIOS-Setup nichts anderes als die Bootsequenz. Wenn Ihnen die Sache zu heikel ist, dann laden Sie jemanden, der sich damit auskennt, zum Tee ein. Wo gibt es Knoppix? Sie können die neueste Version kostenlos unter <http://www.knopper.net> herunterladen. Dort können Sie auch für einen kleinen Kostenbeitrag eine CD-ROM bestellen (was das Herunterladen spart). Nicht zuletzt können Sie einen genauen Blick auf die CD-ROMs werfen, die häufig Computerzeitschriften beiliegen. Nicht selten ist dort eine komplette Knoppix-Version dabei.

CygWin

CygWin ist eine kostenfreie Unix-Umgebung für Windows (ein Emulator). CygWin wird wie eine normale Applikation unter Windows installiert und gestartet. Nach dem Starten erhält man ein Terminalfenster, wie Sie es für unsere Serie benötigen (Abbildung 4). Wenn Sie über einen Internetanschluß verfügen, dann ist die Installation sehr einfach. Gehen Sie auf die Seite

Abbildung 4: CygWin Terminal. Dies ist Ihr Fenster zur Unix-Welt in der Windowsumgebung. Der Befehl `date` wurde bereits ausgeführt.



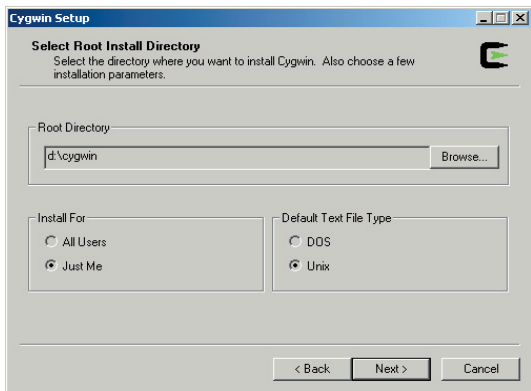


Abbildung 5: CygWin Installation. In diesem Fenster geben Sie an, wohin CygWin auf Ihrer Festplatte installiert werden soll. Außerdem können Sie das Standard-Textdatei-Format festlegen.

<http://www.cygwin.com>. Dort können Sie sich einen etwa 200 kB großes Installationsprogramm (Installer) herunterladen. Dieses starten Sie lokal auf Ihrem Computer. Der Installer leitet Sie durch 6 interaktive Fenster. Zunächst müssen Sie die Installationsquelle angeben. Aktivieren Sie "Install from the Internet". Das darauf folgende Fenster ist in Abbildung 5 gezeigt und sollte entsprechend ausgefüllt werden. Wählen Sie dann "Direct Connection" als Ihren Internet-Anschluss und eine Adresse, von der Sie CygWin installieren möchten. Schließlich werden Sie nach den zu installierenden Paketen gefragt. Verwenden Sie die *default*-Einstellungen. Andernfalls stellen Sie sicher, dass die Pakete *gawk* und *bash* installiert werden. Dies geschieht automatisch, wenn Sie die Grundeinstellungen nicht verändern. Nun erfolgt das Herunterladen der Dateien und Installation.

Nach Doppelklick auf das Desktopsymbol sollten Sie in etwa ein Fenster wie in Abbildung 4 erhalten. Sie können den Installer jede Zeit erneut starten und Pakete installieren oder deinstallieren.

Linux

Schließlich besteht natürlich die Möglichkeit, Linux auf Ihrem Computer zu installieren. Natürlich ist alles frei zugänglich und aus dem Internet herunterzuladen. Dies ist aber sehr mühselig und für den Anfänger ungeeignet. Eine Reihe von Firmen bieten Linux-Distributionen an, die man zum Beispiel in größeren Buchhandlungen für 50 bis 100 Euro kaufen kann. Der Preis ist wirklich niedrig wenn man bedenkt, dass man ein vollständiges Betriebssystem inklusive Handbüchern, Office Applikationen (OpenOffice), dem Photoshop-ähnlichen Programm Gimp, Webserver Software (Apache), Datenbank Applikationen (MySQL) und vieles vieles mehr erhält. Die bekanntesten Distributionen sind wahrscheinlich Suse, Redhat und Mandrake. Aber es gibt viele mehr. Es würde den Rahmen sprengen, hier auf Details einzugehen.

Soviel sei gesagt: Wer schon mal Windows installiert hat, der kann auch Linux installieren. Es gibt auch die Möglichkeit, Windows und Linux parallel auf einem Computer zu installieren. Dies ist meistens in den mit der Linux-Distribution ausgelieferten Handbüchern erläutert. Beim Starten des Computers müssen Sie sich dann das Betriebssystem auswählen, mit dem Sie arbeiten möchten.

Terminal-Beispiel

```

Terminal-1
1      $ date
2      Sat Oct 18 20:09:31 CEST 2003
3      $
    
```

Beim nächsten Mal werden wir loslegen und sehen, wie man mit Dateien und Ordnern umgeht.

Literatur

- [1] Ritchie & Thompson (1974) The Unix time-sharing system. Comm. of the ACM 17:365-375
- [2] Dayhoff & Ledley (1962) Comproteïn: A computer program to aid primary protein structure determination. Proc. Fall Joint Comp. Conf. 22:262-274
- [3] Levinthal (1966) Molecular model-building by computer. Sci. Am. 214:42-52
- [4] Fitch & Margoliash (1967) Construction of phylogenetic trees. Science 155: 279-284

Konventionen

Wenn immer im Text Befehle oder erforderliche Eingaben vorkommen, so werden sie in **Courier** gedruckt. Ebenso werden die Beispielkonsolen (Terminals) und Programme in **Courier** gedruckt und grün bzw. rot hinterlegt. Im Terminal beginnt eine Zeile mit dem Dollarzeichen „\$“. Davor steht je nach Konfiguration und System entweder der Benutzername oder sonst was. In den Terminals im Text wird dies fortgelassen und nur das Dollarzeichen gedruckt. Jede Zeile ist durch eine Zeilennummer gekennzeichnet, auf die im Text Bezug genommen werden kann. Tastaturbefehle werden in **KAPITÄLCHEN** gedruckt bzw. durch Tastensymbole wiedergegeben. Manchmal ist es notwendig, ein Leerzeichen eindeutig zu kennzeichnen um Fehlern vorzubeugen. In diesen Fällen steht das offene Quadrat □ für ein Leerzeichen.