

# grep everything

## Suchen in Anwendungsformaten

Axel Beckert/Frank Hofmann

17. März 2012

- 1 Über die Referenten
- 2 Suchen mit `grep`
- 3 Suchen in Archiven und komprimierten Dateien
- 4 Suchen im PDF mit `pdgrep`
- 5 Suchen in Audio-Dateien mit `taggrepper`
- 6 Suchen in einem Gnumeric-Rechenblatt mit `ssgrep`
- 7 Mailfolder durchsuchen
- 8 XML-Dateien durchforsten
- 9 Referenzen
- 10 Schlussworte

# Axel Beckert – OpenSource-Aktivitäten und Projekte

- Debian-Entwickler
- Vorstandsmitglied Linux User Group Switzerland
- Radiosendung und Podcast “Hackerfunk”

## Frank Hofmann – OpenSource-Aktivitäten und Projekte



2000-2007



seit 2006



seit 2009

Regionales  
LUG-Treffen  
Berlin-  
Brandenburg  
seit 2008



GriCal



# Über Hofmann EDV – Linux, Layout und Satz



Linux, Layout & Satz



WIZARDS OF FOSS  
Open Source Schulungen

- Layout und Satz, Druckvorstufe
- Administration und Service
  - Betreuung von Linux-Systemen (Debian)
  - Programmierung und Automatisierung auf der Basis von PHP und Python
  - Wireless Devices für den Innen- und Außeneinsatz
- Trainings für IT-Spezialisten  
Mitbegründer, Gesellschafter und Trainer

# grep-Grundlagen

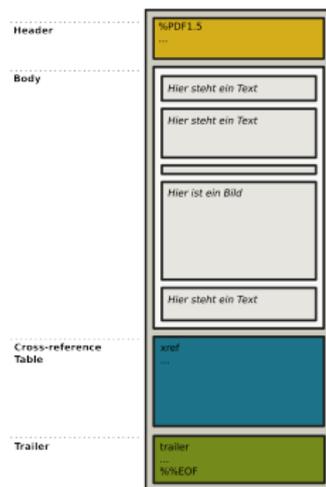
- grep leitet sich ab aus  
g/re/p (**g**lobal / **r**egular **e**xpression / **p**rint)
- grep sucht in Datenströmen nach passenden Mustern
- Muster oder Zeichenfolge ist als Regulärer Ausdruck formuliert
- Beispiel:  
`grep Mikro[tT]ik rechnung201100[1-5]`

Suche in den Dateien `rechnung2011001` bis `rechnung2011005` nach den beiden Mustern „Mikrotik“ und „MikroTik“

# zgrep und Freunde

- `zcat Datei.gz | grep foo` lässt sich abkürzen als `zgrep foo Datei.gz`
- Gibt's für alle Suchmuster-Varianten: `zgrep`, `zfgrep`, `zegrep`
- Gibt's für verschiedene Komprimierungen: `zgrep`, `bzgrep`, `xzgrep`
- Und alle sonstigen Kombinationen: `bzfgrep`, `bzegrep`, `xzfgrep`, `xzegrep`
- `zutils`: `zgrep`-Alternative erkennt Komprimierungsformat automatisch und versteht `gzip`, `bzip2`, `xz` und `lzip`.

# PDFs durchstöbern



- PDF ist ein Datenformat mit einer festgelegten Struktur
- Paket pdfgrep  
<http://pdfgrep.sourceforge.net/>
- Beispiel:  
`pdfgrep -i RouterBoard rechnung*.pdf`

Suche in den Dateien `rechnung*.pdf` nach dem Muster „RouterBoard“ – unabhängig von Groß- und Kleinschreibung (Option „-i“)

# Audio-Dateien durchsuchen

- Paket taggrepper  
<http://gitorious.org/taggrepper/pages/Home>
- Suche in Audiodateien anhand der Tags  
bei MP3: Titel, Interpret, Album, Jahr, Genre, Kommentar, Nummer des Titels, Komponist, Originalinterpret, Copyright und URL
- taggrepper unterstützt die Formate MP3, Ogg Vorbis und FLAC
- Beispiel:  

```
taggrepper --display-artist -a NDR *.mp3
```

Suche in allen MP3-Dateien nach den Stücken, die vom Autor „NDR“ stammen; gib Dateiname und Suchtreffer aus

## ssgrep

	A	B	C	D
1				
2	Datum	Name	Produkt	Auftragswert
3				
4	15.02.2009	Franz Weiler	8xUbiquiti NanoStation M5	632,13
5	06.03.2010	Godfried Schimmer	2xUbiquiti NanoStation 5	178,00
6	13.05.2010	Hartmut Fleischer	2xUbiquiti NanoStation M5	170,00
7	24.05.2010	Georg Strauch	1xUbiquiti PicoStation 2HP	79,00
8	01.07.2010	Hans Grube	1xUbiquiti NanoStation M5	85,00
9				

- Gnumeric-Format ist gzip-komprimiertes XML
- ssgrep ist Bestandteil des Gnumeric-Paketes

Beispiel:

```
$ssgrep -Hn "Nano[Ss]tation" datei.gnumeric
datei.gnumeric:Blatt1!C4:8xUbiquiti NanoStation M5
$
```

Suche in der Dateien datei.gnumeric nach dem Muster „NanoStation“ (mit S oder s). „-Hn“ gibt Zelle und Tabellenblatt aus

# mboxgrep und grepmail

- Durchsuchen E-Mail-Ordner, geben gesamte Mail bei Suchtreffer aus
- Unterstützen das mbox-Format, mboxgrep auch andere Formate (Maildir, MH, etc.)
- Ausgabe ist wieder ein valider E-Mail-Ordner im mbox-Format
- mboxgrep kann Mails mit Treffer direkt löschen
- grepmail kann Ausgabe auf Kopfzeilen oder Inhalt beschränken
- grepmail kann Suchtreffer in Signaturen unterdrücken

# Suche mit xmllgrep

- xmllgrep: grep-Analogon für die Suche in XML-Daten
- gefundene Pakete:  
Perl-Modul XML::Twig, xmlstar, xml-coreutils, xmllclitools, XGrep, xml\_grep2, sgrep
- Lokalisierung eines Knotens im XML-Baum: XPath
- Beispiel für Perl-Modul XML::Twig, Kommando xml\_grep:

```
xml_grep --text_only "//sortiment/buch/isbn" buch.xml
```

Suche in der Datei buch.xml den Unterknoten isbn und gib den Wert des Knotens aus

# Zum Nachlesen

- Übersicht zu verschiedenen grep-Varianten in Axel Beckerts Blog:  
<http://noone.org/blog/English/Computer/Shell/grep%20everything.futile>
- PDF und PostScript durchsuchen, LinuxUser 2/2012
- Suche in komprimierten Dateien und Archiven, LinuxUser 4/2012
- grep in Anwendungsformaten (Teil 1), vorr. in LinuxUser 5/2012
- grep in Anwendungsformaten (Teil 2), vorr. in LinuxUser 6/2012

# Vielen Dank!

## Danke für Ihre Aufmerksamkeit :-)

### Kontakt:

Axel Beckert

E-mail <[abe@deuxchevaux.org](mailto:abe@deuxchevaux.org)>

Web <http://noone.org/abe/>



Linux, Layout & Satz

YES, WE ARE OPEN!  
**büro2.0**  
Open Source Bürogemeinschaft & Expertenetzwerk

Frank Hofmann

E-Mail <[frank.hofmann@efho.de](mailto:frank.hofmann@efho.de)>

Web <http://www.efho.de>