

Using the power of Google - elgooG

Hacking google - hacker's best friend

Michael Prokop / Grazer LinuxTage 2004 / 7. Mai 2004

michael@linuxtage.at / www.michael-prokop.at / www.grml.org





Inhalt 1/2

Einleitung

● Inhalt 1/2

● Inhalt 2/2

Theorie

Über Google

Googlestrings

Generelles zur Googlesuche

Dienste von Google

Googles Ranking

Hacking

Rund um Google

Other stuff

The End...

- Was will man eigentlich?
- Wie geht man an eine Suche heran?
- Wozu?
- Warum Google?
- Hinter den Kulissen von Google
- Strings - "http://www.google.at/search?"
- Googlesuche - Generelles
- Spezialsuchen



Inhalt 2/2

Einleitung

● Inhalt 1/2

● Inhalt 2/2

Theorie

Über Google

Googlestrings

Generelles zur Googlesuche

Dienste von Google

Googles Ranking

Hacking

Rund um Google

Other stuff

The End...

- Dienste von Google
- Googlesuche - Arbeitsweise
- Googlesuche - "Hacking"
- Rund um Google
- Suchmaschinen effizient nutzen
- Probleme von Google/Ausblick
- Q&A



Was will man eigentlich?

suchen wissen

ich was suchen

ich nicht wissen was suchen

ich nicht wissen wie wissen was suchen

ich suchen wie wissen was suchen

ich wissen was suchen

ich suchen wie wissen was suchen

ich wissen ich suchen wie wissen was suchen

ich was wissen

– Ernst Jandl (1978)

Einleitung

Theorie

● Was will man eigentlich?

● Herangehensweise

● Wozu?

● Warum Google?

Über Google

Googlestrings

Generelles zur Googlesuche

Dienste von Google

Googles Ranking

Hacking

Rund um Google

Other stuff

The End...



Herangehensweise

Einleitung

Theorie

● Was will man eigentlich?

● Herangehensweise

● Wozu?

● Warum Google?

Über Google

Googlestrings

Generelles zur Googleuche

Dienste von Google

Googles Ranking

Hacking

Rund um Google

Other stuff

The End...

- Was will man überhaupt erreichen?
- Ist eine Suchmaschine für die Lösung des Problems geeignet?
 - ◆ Welche Suchmaschine?
- Wie viel Zeit möchte ich investieren?
- Wie viel Zeit muss ich für die Suche rechnen?
- Wie viel darf die Suche "kosten"?
- Suchmaschine ist also der Weg zum Ziel?
- Welche Suchbegriffe?
- Welche Suchmuster/Filterbegriffe?



Wozu?

Einleitung

Theorie

- Was will man eigentlich?
- Herangehensweise
- Wozu?
- Warum Google?

Über Google

Googlestrings

Generelles zur Googlesuche

Dienste von Google

Googles Ranking

Hacking

Rund um Google

Other stuff

The End...

- Erarbeiten von Informationskompetenz.
- Informationskompetenz? Literatur, Wissen und Informationen aus elektronischen und konventionellen Quellen auffinden und verarbeiten können.
- Zugang zu Wissen hat sich in den letzten Jahren/Jahrzehnten drastisch verändert.
- Ziel: das Lösen von Problemen.



Warum Google?

Einleitung

Theorie

- Was will man eigentlich?
- Herangehensweise
- Wozu?
- Warum Google?

Über Google

Googlestrings

Generelles zur Googlesuche

Dienste von Google

Googles Ranking

Hacking

Rund um Google

Other stuff

The End...

- grundsätzlich einfache Handhabung
- vielfältige Filtermechanismen
- viele verschiedene Dienste (Vorreiterrolle)
- momentan(?!) noch relativ werbefrei
- großer Datenbestand vorhanden
- gegen Infiltrierungsmechanismen im Vergleich zu anderen Suchmaschinen noch(?) relativ(!) immun



Hinter den Kulissen 1/3

Einleitung

Theorie

Über Google

● Hinter den Kulissen 1/3

● Hinter den Kulissen 2/3

● Hinter den Kulissen 3/3

Googlestrings

Generelles zur Googlesuche

Dienste von Google

Googles Ranking

Hacking

Rund um Google

Other stuff

The End...

- Hauptquartier von Google: Mountain View (Kalifornien)
- Google: zum Synonym für Suchmaschinen geworden (UTFG!)
- "Google": Wortspiel mit "googol" - die von Milton Sirotta geprägte Zahl mit einer 1 und 100 Nullen
- 1998 von Larry Page und Sergey Brin gegründet, ursprünglich ein Forschungsprojekt an Uni Stanford
- Konzernchef Eric Schmidt (ehemals Konzernchef bei Novell)
- > 200 Millionen Anfragen pro Tag, bis zu 2000 pro Sekunde in Rush-Hour (6:00 bis Mittag PST [ab 13:00 MEZ]) [bis zu 4 Tera-Ops]
- offiziell zw. 10.000 und 50.000 Server (Linux)
- mehr als 6 Datacenter



Hinter den Kulissen 2/3

Einleitung

Theorie

Über Google

● Hinter den Kulissen 1/3

● **Hinter den Kulissen 2/3**

● Hinter den Kulissen 3/3

Googlestrings

Generelles zur Googlesuche

Dienste von Google

Googles Ranking

Hacking

Rund um Google

Other stuff

The End...

”offizieller” Stand März-Mai 2004:

- >4 Petabyte Festplattenkapazität
- 3 Milliarden Websites, 4 Milliarden Dokumente indiziert [:= ca. 280 km hohem Papierstapel, 4 Billionen Wörtern]
- 845 Millionen Usenet-Postings (>35.000 Themengebiete)[seit 1981]
- 880 Millionen Bilder
- >1000 Mitarbeiter [AKA Googler]
- 35 verschiedenen Sprachen
[<http://www.google.com/intl/xx-klingon/>]



Hinter den Kulissen 3/3

Einleitung

Theorie

Über Google

- Hinter den Kulissen 1/3
- Hinter den Kulissen 2/3
- Hinter den Kulissen 3/3

Googlestrings

Generelles zur Googlesuche

Dienste von Google

Googles Ranking

Hacking

Rund um Google

Other stuff

The End...

- Suchergebniss meist in weniger als 0.5 Sekunden
- geschätzter Umsatz von 700-1000 Millionen Dollar
- Zeitgeist von Google:
<http://www.google.com/press/zeitgeist.html>
- The Anatomy of a Large-Scale Hypertextual Web Search Engine: <http://www-db.stanford.edu/backrub/google.html>
- Geheimniskrämerei bei Google? - Technology/Heise
- Google-Filesystem [PDF]



Googlestrings - 1/3

Einleitung

Theorie

Über Google

Googlestrings

● Googlestrings - 1/3

● Googlestrings - 2/3

● Googlestrings - 3/3

Generelles zur Googlesuche

Dienste von Google

Googles Ranking

Hacking

Rund um Google

Other stuff

The End...

- `q=` → query/Suchbegriff(e)
- `as_q=` → Suchbegriff(e) [automatisches AND]
- `num=` → Anzahl der auszugebenden Suchergebnisse/Seite
- `hl=` → Sprache für die Benutzeroberfläche von Google
- `ie=` → verwendeter Zeichensatz bei der Eingabe
- `oe=` → ausgegebener Zeichensatz
- `btnG=` → Suchbutton
- `as_epq=` → exakte Phrasensuche



Googlestrings - 2/3

Einleitung

Theorie

Über Google

Googlestrings

● Googlestrings - 1/3

● Googlestrings - 2/3

● Googlestrings - 3/3

Generelles zur Googlesuche

Dienste von Google

Googles Ranking

Hacking

Rund um Google

Other stuff

The End...

- `as_oq=` → Suchbegriffe [Verknüpfung mit OR]
- `as_eq=` → exkludieren von Suchbegriff(en)
- `lr=` → Sprache
- `as_ft=` → Suche auf Dateityp beschränken (=i)/Dateityp exkludieren (=e)
- `as_filetype=` → welcher Dateityp (doc, ppt, pdf,...)



Googlestrings - 3/3

Einleitung

Theorie

Über Google

Googlestrings

● Googlestrings - 1/3

● Googlestrings - 2/3

● Googlestrings - 3/3

Generelles zur Googlesuche

Dienste von Google

Googles Ranking

Hacking

Rund um Google

Other stuff

The End...

- `as_qdr=` → maximales Alter der Suchergebnisse [Zeitpunkt, zu dem Dokument zu Googles Index hinzugefügt wurde]
- `as_occt=` → occurs
- `as_dt=` → Suche auf spezifische URL beschränken (=i) oder URL exkludieren (=e) [siehe `as_sitesearch`]
- `as_sitesearch=` → Suche auf eine Website beschränken/exkludieren
- `safe=` → SafeSearch-Filter (on/off)
- Advanced Search Operators @ Google



Googlesuche - Generelles

Einleitung

Theorie

Über Google

Googlestrings

Generelles zur Googlesuche

● Googlesuche - Generelles

● Googlesuche - Generelles

● Spezialsuchen

Dienste von Google

Googles Ranking

Hacking

Rund um Google

Other stuff

The End...

- Google sucht automatisch nach versch. Schreibweisen
[Disclaimer: Vorsicht]:
 - ◆ case insensitive ("GOOGLE" vs. "google")
 - ◆ Worttrennung irrelevant ("News-Archiv" bs. "Newsarchiv")
 - ◆ versch. Schreibweisen ("Photo" vs. "Foto")
 - ◆ Umlaute ("Wörter" matcht auch "Woerter")
- Was machen, wenn doch bestimmte Schreibweise notwendig?
"[+WORT]"



Googlesuche - Generelles

Einleitung

Theorie

Über Google

Googlestrings

Generelles zur Googlesuche

● Googlesuche - Generelles

● Googlesuche - Generelles

● Spezialsuchen

Dienste von Google

Googles Ranking

Hacking

Rund um Google

Other stuff

The End...

- Google ignoriert leider "Wildcards" (RegEx alla "suck*")
- "Stop-Wörter": z.B. "http" und ".com" werden normalerweise von Google ignoriert, wenn für Suche relevant: "+" vor Suchstring hinzufügen - z.B. "+und"
- Exkludieren von Begriffen via "-": "newsarchiv -tugraz"
- Suche auf bestimmte Website via "site:" reduzieren: "site:tugraz.at"
- Suchergebnisse auf Ergebnisse beschränken, die bestimmten String in URL beinhalten:
<http://www.google.com/search?q=linux+inurl%3Apdf>
Siehe dazu: <http://www.google.com/intl/de/help/basics.html>



Spezialsuchen

Einleitung

Theorie

Über Google

Googlestrings

Generelles zur Googlesuche

● Googlesuche - Generelles

● Googlesuche - Generelles

● **Spezialsuchen**

Dienste von Google

Googles Ranking

Hacking

Rund um Google

Other stuff

The End...

- Definitionen: Was ist ICMP?

<http://www.google.com/search?q=define:ICMP>

- Domain-Einschränkung:

http://www.google.com/search?as_q=100BASE-T&as_sitesearch=ieee.or

- Linked: <http://www.google.com/search?q=link:www.orf.at>

- Print:

<http://www.google.com/search?q=site:print.google.com+shakespeare>

- Rechner: Ergebnis von “ $\pi + ((2 * 2) / 3)$ ”?

<http://www.google.com/search?q=pi%2B%282%2A2%29%2F3>

- Rechner: Maßeinheiten umrechnen

<http://www.google.com/search?q=10meter+to+inch>

Siehe auch: <http://www.google.com/intl/de/features.html>



Dienste von Google 1/5

[Einleitung](#)

[Theorie](#)

[Über Google](#)

[Googlestrings](#)

[Generelles zur Googlesuche](#)

[Dienste von Google](#)

● [Dienste von Google 1/5](#)

● [Dienste von Google 2/5](#)

● [Dienste von Google 3/5](#)

● [Dienste von Google 4/5](#)

● [Dienste von Google 5/5](#)

[Googles Ranking](#)

[Hacking](#)

[Rund um Google](#)

[Other stuff](#)

[The End...](#)

- <http://groups.google.com/> - Google Groups (Usenet [ehemals Dejanews])
- <http://froogle.google.com/> - Google Froogle (Produktverkauf)
- <http://catalogs.google.com/> - Google Catalogs (mail-order catalogs)
- <http://images.google.com/> - Google Image Search
- http://www.google.com/language_tools - Google Language Tools
- <http://answers.google.com/> - Google Answers (kostenpflichtig)



Dienste von Google 2/5

Einleitung

Theorie

Über Google

Googlestrings

Generelles zur Googlesuche

Dienste von Google

● Dienste von Google 1/5

● **Dienste von Google 2/5**

● Dienste von Google 3/5

● Dienste von Google 4/5

● Dienste von Google 5/5

Googles Ranking

Hacking

Rund um Google

Other stuff

The End...

- <http://news.google.com/> - Google News ("aktuelle Indizierung")
- <http://directory.google.com/> - Google Web Directory (Kategorien browsen)
- <http://groups.google.at/options/universities.html> - Google University Search
- <http://groups.google.at/options/wireless.html> - Google Wireless (GNS, PDA)



Dienste von Google 3/5

Einleitung

Theorie

Über Google

Googlestrings

Generelles zur Googlesuche

Dienste von Google

● Dienste von Google 1/5

● Dienste von Google 2/5

● **Dienste von Google 3/5**

● Dienste von Google 4/5

● Dienste von Google 5/5

Googles Ranking

Hacking

Rund um Google

Other stuff

The End...

- <http://groups.google.at/options/specialsearches.html> - Google Special Searches:
 - ◆ <http://groups.google.at/unclesam> - U.S. Government
 - ◆ <http://groups.google.at/linux> - Linux
 - ◆ <http://groups.google.at/bsd> - *BSD
 - ◆ <http://images.google.com/mac.html> - Apple Macintosh
 - ◆ <http://images.google.com/microsoft.html> - Microsoft



Dienste von Google 4/5

Einleitung

Theorie

Über Google

Googlestrings

Generelles zur Googlesuche

Dienste von Google

● Dienste von Google 1/5

● Dienste von Google 2/5

● Dienste von Google 3/5

● **Dienste von Google 4/5**

● Dienste von Google 5/5

Googles Ranking

Hacking

Rund um Google

Other stuff

The End...

- Google Labs (<http://labs.google.com/>)
 - ◆ Google Glossary - Definitionen von Begriffen, Abkürzungen und Akronymen
 - ◆ Google Sets - Liste mit passenden Ergänzungen
 - ◆ Google Keyboard Shortcuts - Navigieren ohne Maus in den Suchergebnissen
 - ◆ Google Voice Search
 - ◆ Google Webquotes - Kommentare zu Websites
 - ◆ Google Viewer - Diaschau



Dienste von Google 5/5

Einleitung

Theorie

Über Google

Googlestrings

Generelles zur Googlesuche

Dienste von Google

● Dienste von Google 1/5

● Dienste von Google 2/5

● Dienste von Google 3/5

● Dienste von Google 4/5

● Dienste von Google 5/5

Googles Ranking

Hacking

Rund um Google

Other stuff

The End...

- Google Web APIs - : SOAP, WSDL um Google innerhalb von Applikationen (Java, Perl, Visual Studio .NET,...) zu nutzen.
- Überblick über Dienste von Google? Klar:
 - ◆ <http://www.google.at/sitemap.html>
 - ◆ <http://www.google.com/help/features.html>



Googlesuche - Arbeitsweise 1/2

Einleitung

Theorie

Über Google

Googlestrings

Generelles zur Googlesuche

Dienste von Google

Googles Ranking

● Googlesuche - Arbeitsweise 1

● Googlesuche - Arbeitsweise 2/2

Hacking

Rund um Google

Other stuff

The End...

- PageRank-Technologie: Bewertung der Wichtigkeit von Websites → Gleichung mit 500 Millionen Variablen und über 3 Milliarden Begriffen; Analyse der Linkstruktur → Angriffspunkt durch Selbstreferenzialität!
- hypertextbasiert: Analyse des gesamten Inhaltes (Schriftart, Unterteilungen, genaue Position aller Begriffe im Text)



Googlesuche - Arbeitsweise 2/2

Einleitung

Theorie

Über Google

Googlestrings

Generelles zur Googlesuche

Dienste von Google

Googles Ranking

● Googlesuche - Arbeitsweise 1/2

● Googlesuche - Arbeitsweise 2

Hacking

Rund um Google

Other stuff

The End...

- Analyse von benachbarten Websites
- Werbung: auf Inhalt bezogen
- Google Search Appliance: integrierte Suchlösung für Geschäftskunden
- Literatur:
 - ◆ <http://www.kso.co.uk/de/pagerank.html>
 - ◆ http://www.webmasterpro.de/groups_tip-365.php



Googlesuche - "Hacking" 1/3

Einleitung

Theorie

Über Google

Googlestrings

Generelles zur Googlesuche

Dienste von Google

Googles Ranking

Hacking

● Googlesuche - "Hacking" 1/3

● Googlesuche - "Hacking" 2/3

● Googlesuche - "Hacking" 3/3

Rund um Google

Other stuff

The End...

- Erkennen von Standardpattern:
<http://images.google.com/images?q=DSCF0005>
- Verknüpfungen herstellen:
<http://images.google.com/images?q=bill+gates+torte>
- Suche nach "Unbekanntem":
[http://\[images|www\].google.com/images?q=windows+nachfolger+xp](http://[images|www].google.com/images?q=windows+nachfolger+xp)



Googlesuche - "Hacking" 2/3

[Einleitung](#)

[Theorie](#)

[Über Google](#)

[Googlestrings](#)

[Generelles zur Googlesuche](#)

[Dienste von Google](#)

[Googles Ranking](#)

Hacking

● [Googlesuche - "Hacking" 1/3](#)

● [Googlesuche - "Hacking" 2/3](#)

● [Googlesuche - "Hacking" 3/3](#)

[Rund um Google](#)

[Other stuff](#)

[The End...](#)

- google.com/search?q=völlige+inkompetenz&ie=utf-8
- Google-Whacking: Erzielen von "Resultate 1-1 von 1"
- Google-Bombing: Verlinken via Linkbeschreibung; Motiv: oft politisch ('völlige Inkompetenz', 'miserable failure')
- google.at/search?q=answer+to+life+the+universe+and+everything%3D



Googlesuche - "Hacking" 3/3

Einleitung

Theorie

Über Google

Googlestrings

Generelles zur Googlesuche

Dienste von Google

Googles Ranking

Hacking

● Googlesuche - "Hacking" 1/3

● Googlesuche - "Hacking" 2/3

● Googlesuche - "Hacking" 3/3

Rund um Google

Other stuff

The End...

- <http://www.heise.de/newsticker/meldung/46634>

- <http://www.heise.de/...>

- <http://www.heise.de/...>

-

google.com/search?q=news://E1BFgEN-0001xf-00@gadolinium.btinterne

- Hinweis auf torrent (google: torrent search) → <http://isohunt.com/>



Rund um Google 1/3

Einleitung

Theorie

Über Google

Googlestrings

Generelles zur Googlesuche

Dienste von Google

Googles Ranking

Hacking

Rund um Google

● Rund um Google 1/3

● Rund um Google 2/3

● Rund um Google 3/3

Other stuff

The End...

- <http://www.alltooflat.com/geeky/elgoog/>
- <http://www.google.com/intl/xx-hacker/>
- <http://www.eaves.org:28080/google/howcool>
- <http://google.blogspot.com/>
- <http://www.googleblaster.com/>
- <http://www.google.com/googlefriends/archive.html>



Rund um Google 2/3

Einleitung

Theorie

Über Google

Googlestrings

Generelles zur Googlesuche

Dienste von Google

Googles Ranking

Hacking

Rund um Google

● Rund um Google 1/3

● Rund um Google 2/3

● Rund um Google 3/3

Other stuff

The End...

- <http://google.weblogsinc.com/>
- <http://www.queryster.com/google/>
- <http://www.freshgoo.com>
- <http://www.google-watch.org/>
- <http://www.searchlores.org/>



Rund um Google 3/3

Einleitung

Theorie

Über Google

Googlestrings

Generelles zur Googlesuche

Dienste von Google

Googles Ranking

Hacking

Rund um Google

● Rund um Google 1/3

● Rund um Google 2/3

● Rund um Google 3/3

Other stuff

The End...

- <http://www.suchfibel.de/>
- <http://www.google-watch.org/>
- <http://www.searchenginewatch.com/>
- <http://www.googlefight.com/>
- <http://www.geocities.com/googlepubsupgenfaq/>



Suchmaschinen effizient nutzen

Einleitung

Theorie

Über Google

Googlestrings

Generelles zur Googlesuche

Dienste von Google

Googles Ranking

Hacking

Rund um Google

Other stuff

● Suchmaschinen effizient nutz

● Probleme von Google

● Ausblick

The End...

- Kennenlernen der gängigen Suchmechanismen
- eigene Suchseite mit häufig gebrauchte Suchmaschinen erstellen und nutzen (siehe z.B. www.michael-prokop.at/suche/)
- Browserfeatures nutzen:
 - ◆ Opera → Shift+F8
 - ◆ Konqueror "gg:suchbegriff"
 - ◆ Mozilla: www.mozilla.org/docs/end-user/keywords.html
 - ◆ Internet Explorer → grundsätzlich vermeiden, wenn schon dann Google-Toolbar



Probleme von Google

[Einleitung](#)

[Theorie](#)

[Über Google](#)

[Googlestrings](#)

[Generelles zur Googlesuche](#)

[Dienste von Google](#)

[Googles Ranking](#)

[Hacking](#)

[Rund um Google](#)

[Other stuff](#)

● Suchmaschinen effizient nutzen

● Probleme von Google

● Ausblick

[The End...](#)

- Cloaking: verschiedene Seiten werden passend angeboten
- Blogs
- Referers
- Spamming
- Produktlinks
- Quasi-Monopol(!)



Ausblick

[Einleitung](#)

[Theorie](#)

[Über Google](#)

[Googlestrings](#)

[Generelles zur Googlesuche](#)

[Dienste von Google](#)

[Googles Ranking](#)

[Hacking](#)

[Rund um Google](#)

[Other stuff](#)

● Suchmaschinen effizient nutzen

● Probleme von Google

● **Ausblick**

[The End...](#)

- Bessere Unterstützung von audiovisuellen Inhalten (MPEG7, AudioID)
- Personalisierung von Diensten (spezielle Vorlieben) und des Interfaces (personalisierte Startseite, History-Funktion)
- "All-in-One-Interface ohne speziellen Kenntnissen"
- Börsengang und Microsoft-Suchmaschine? ...



Und nun?

- Geschafft.. puh ;-)



Q&A: Gibt es Fragen?

[Einleitung](#)

[Theorie](#)

[Über Google](#)

[Googlestrings](#)

[Generelles zur Googlesuche](#)

[Dienste von Google](#)

[Googles Ranking](#)

[Hacking](#)

[Rund um Google](#)

[Other stuff](#)

[The End...](#)

Und nun?

The End...



Und nun?

- Geschafft.. puh ;-)



Q&A: Gibt es Fragen?

- google.at :-)

Einleitung

Theorie

Über Google

Googlestrings

Generelles zur Googlesuche

Dienste von Google

Googles Ranking

Hacking

Rund um Google

Other stuff

The End...

● Und nun?

● The End...



The End...

Online: www.linuxtage.at/prokop/
Danke für die Aufmerksamkeit!
Feedback ist willkommen!
michael@linuxtage.at

Viel Spaß noch auf den Grazer Linuxtagen!

[Einleitung](#)

[Theorie](#)

[Über Google](#)

[Googlestrings](#)

[Generelles zur Googlesuche](#)

[Dienste von Google](#)

[Googles Ranking](#)

[Hacking](#)

[Rund um Google](#)

[Other stuff](#)

[The End...](#)

Und nun?

The End...