

Watch out, Google.

© 2003 by Sebastian Wolfgarten

# Table of contents

- How Google works
- Basics
- Advanced usage
- Introduction to web security /  
Examples
- Google is DANGEROUS / Examples
- Future prospects
- Summary

# How google works (I)

- Google does not spider the whole web like previous search engines did
- It uses a distributed network of servers (>10000) to follow links and cache the contents of the responding web sites
- The overall search results are based on the number of referring links a site has
- New intelligent search concept but makes room for abuse.

# How google works (II)

- Site A has a link to site B because it has some useful information about a special topic. Google crawls the pages of A and finds the link to B.
- It visits site B and that site gets listed in the index under the keyword A defined.
- Depending on the overall number of links a site has, it gets listed at the top of google's index.

# Basics (I)

- The usage of Google is very simple. Just point your browser to `http://www.google.com` and define your search pattern:



# Basics (II)

- Google supports simple logic operations like AND (+) and NOT (-).
- It also supports searching for given phrases by using inverted commas.
- Matching any character is possible by using a single period.
- Additionally matching a space and a wildcard word is possible with the dash (-) and the asterisk (\*) character (when used in quotes).

# Advanced usage (I)

- Beside these very basic commands, Google has a lot more to offer (most of these features are more or less undocumented!):
- Date restricted search: Searches for documents that were published within a specific date range.
- Title search: Searches for a specific keyword in the title of a page.

# Advanced usage (II)

- **Url search:** Enables you to look for a specific pattern within the URL.
- **Filetype search:** For a long period of time Google is able to search the content of binary documents (e.g. Microsoft Excel or PDF files).
- **Site specific search:** Google is able to limit a search only to a given site (e.g. ba-stuttgart.de) or a certain top level domain (e.g. .de).

# Introduction to web security (I)

- A web server is a server that serves information that are stored on a local disk (e.g. .txt or .html files).
- These information can be requested by clients using a small piece of software (=browser, e.g. Internet Explorer) and a special protocol (HTTP).
- Famous software products used to run a web server are Apache and IIS.

# Introduction to web security / (II)

- The amount of information (e.g. images, movies, documents etc.) published by the web server is either controlled by the admins or other people that are in charge of that.
- Most of the information should be accessible by everyone but some information (e.g. log files, special contents) should usually only be visible for authorized people/customers.

# Introduction to web security / (III)

- In the beginning of an attack a remote system (e.g. web server) is very often a black box to the attacker.
- At this point the attacker has no or just little knowledge about the architecture, software used or the file/directory structure on the remote system.
- How does this all regard to Google?

# Google is DANGEROUS (I)

- Due to misconfiguration or unintentional links Google is even able to find sensible information that should NOT be published to the public.
- So Google becomes dangerous and abets misuse by the sheer amount of (sensitive) information it holds.
- What do you think can be found by using Google?

# Google is DANGEROUS (II)

- Using the advanced commands Google is offering an attacker is able to find:
  - ✓ Valid usernames and passwords
  - ✓ Credit card numbers
  - ✓ Vulnerable systems
  - ✓ Error messages or certain types of servers
  - ✓ Sensitive information like hidden directories or files (e.g. backups)

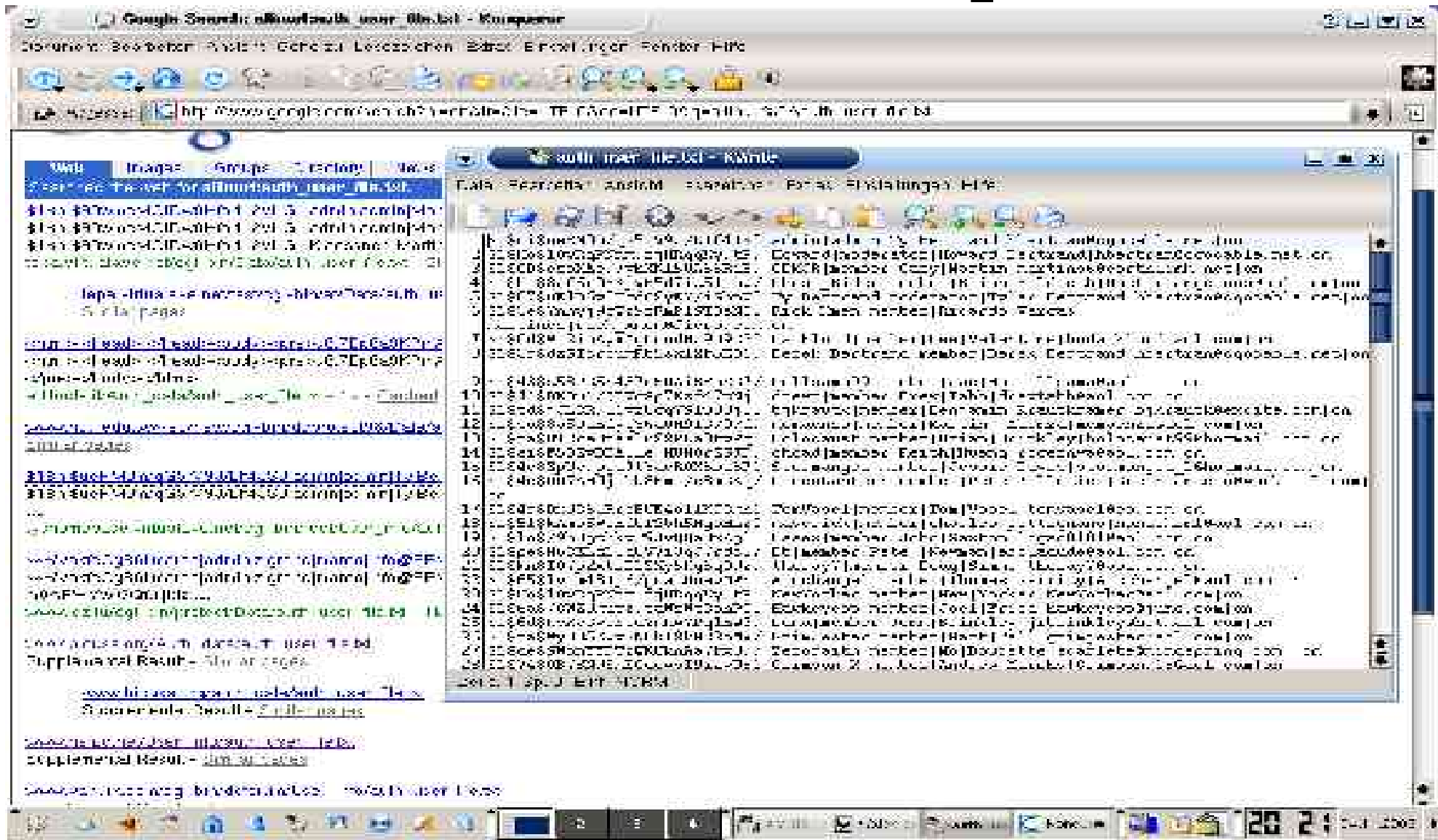
# Google is DANGEROUS (III)

- Valid usernames and passwords can be found by searching for...
  - `intitle:"Index of" htpasswd.bak`
  - `intitle:"index.of.etc" passwd`
  - `allinurl:auth_user_file.txt`
  - ...

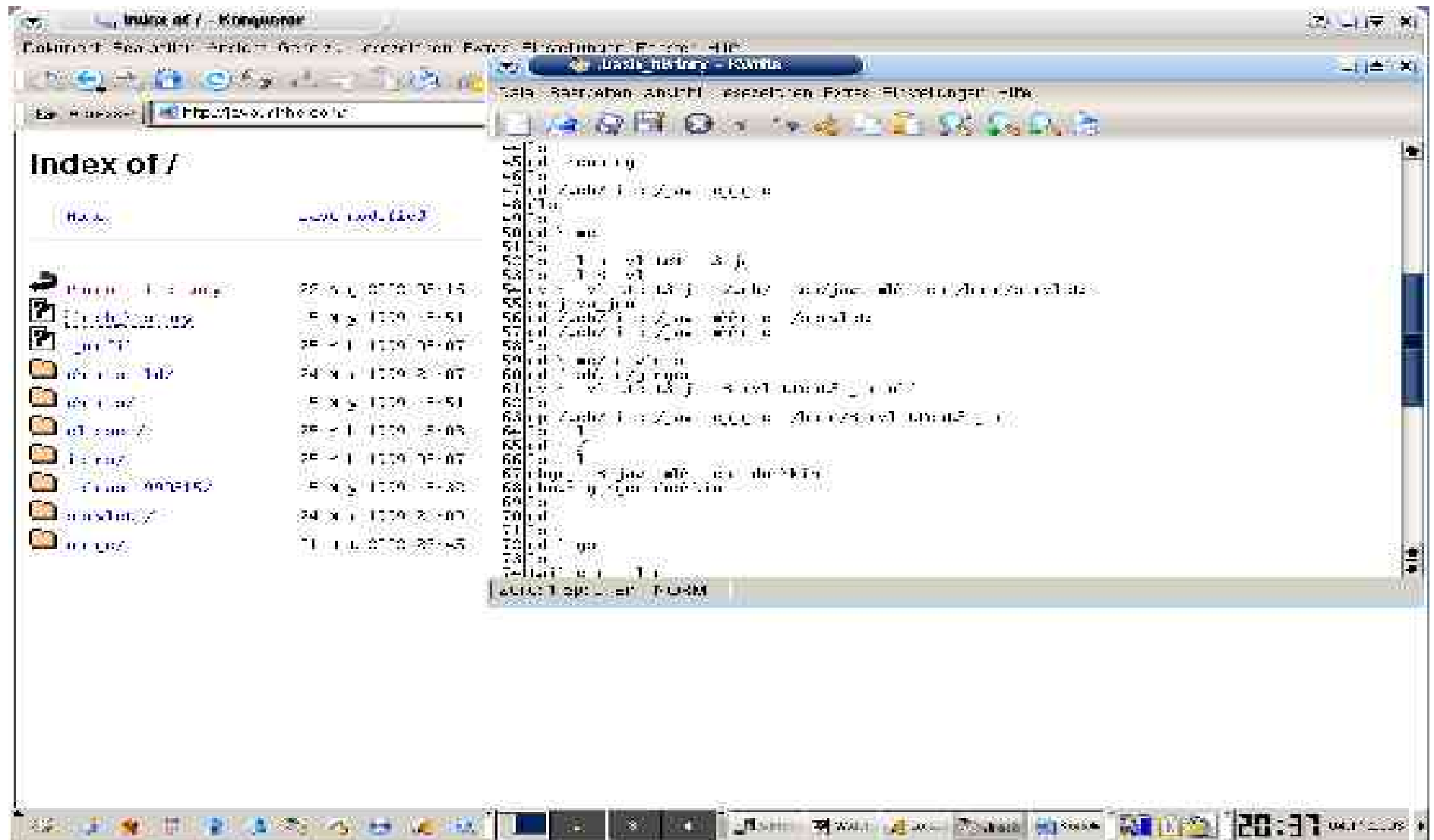
# Google is DANGEROUS (IV)

- Other interesting stuff (backup files, error messages, credit card numbers etc.):
  - "Index of /backup"
  - intitle:"Index of" .bash\_history
  - intitle:admin intitle:login
  - intitle:index.of.secret
  - i\_index.shtml "Ready"

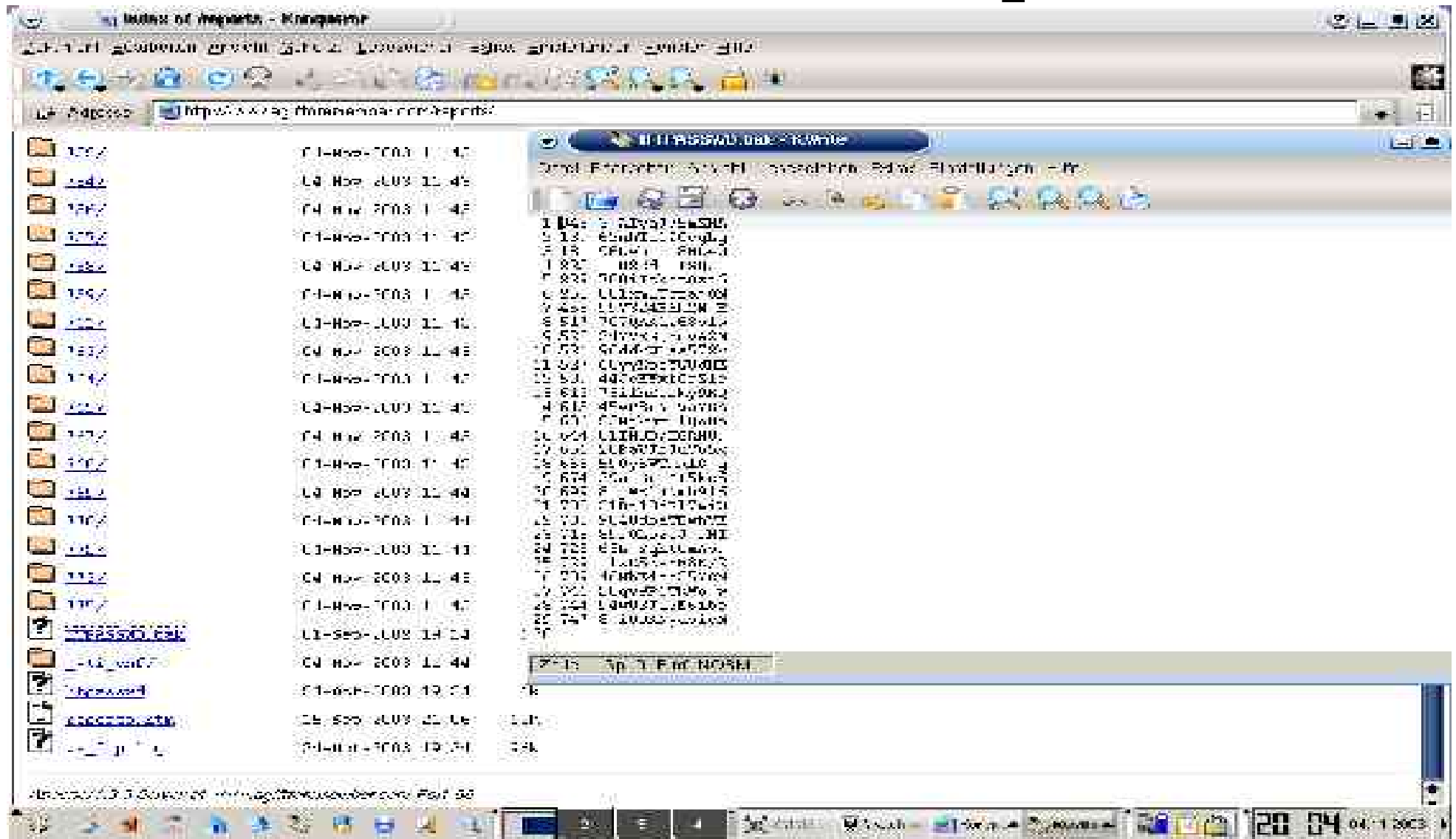
# Google is DANGEROUS/Examples (V)



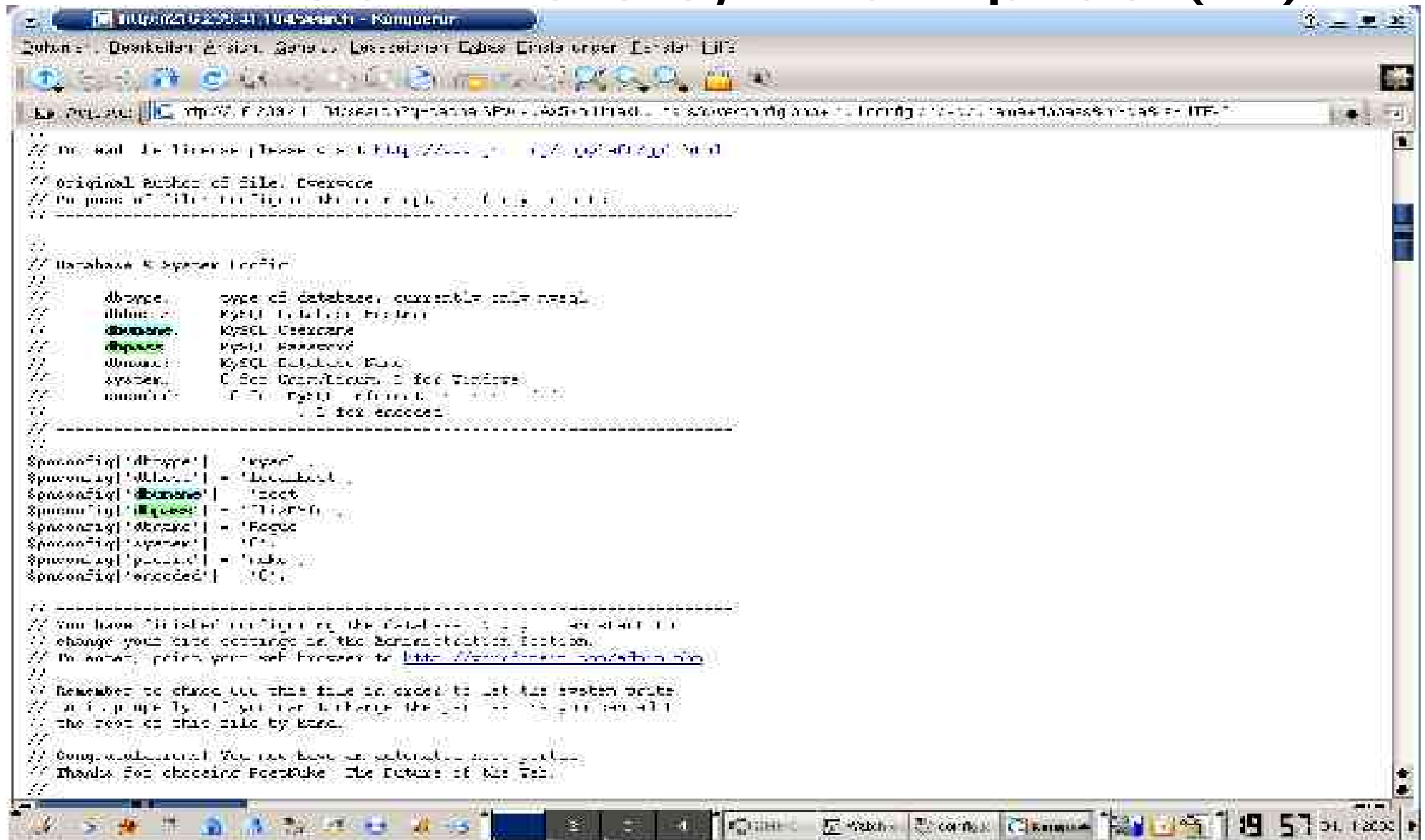
# Google is DANGEROUS/Examples (V)



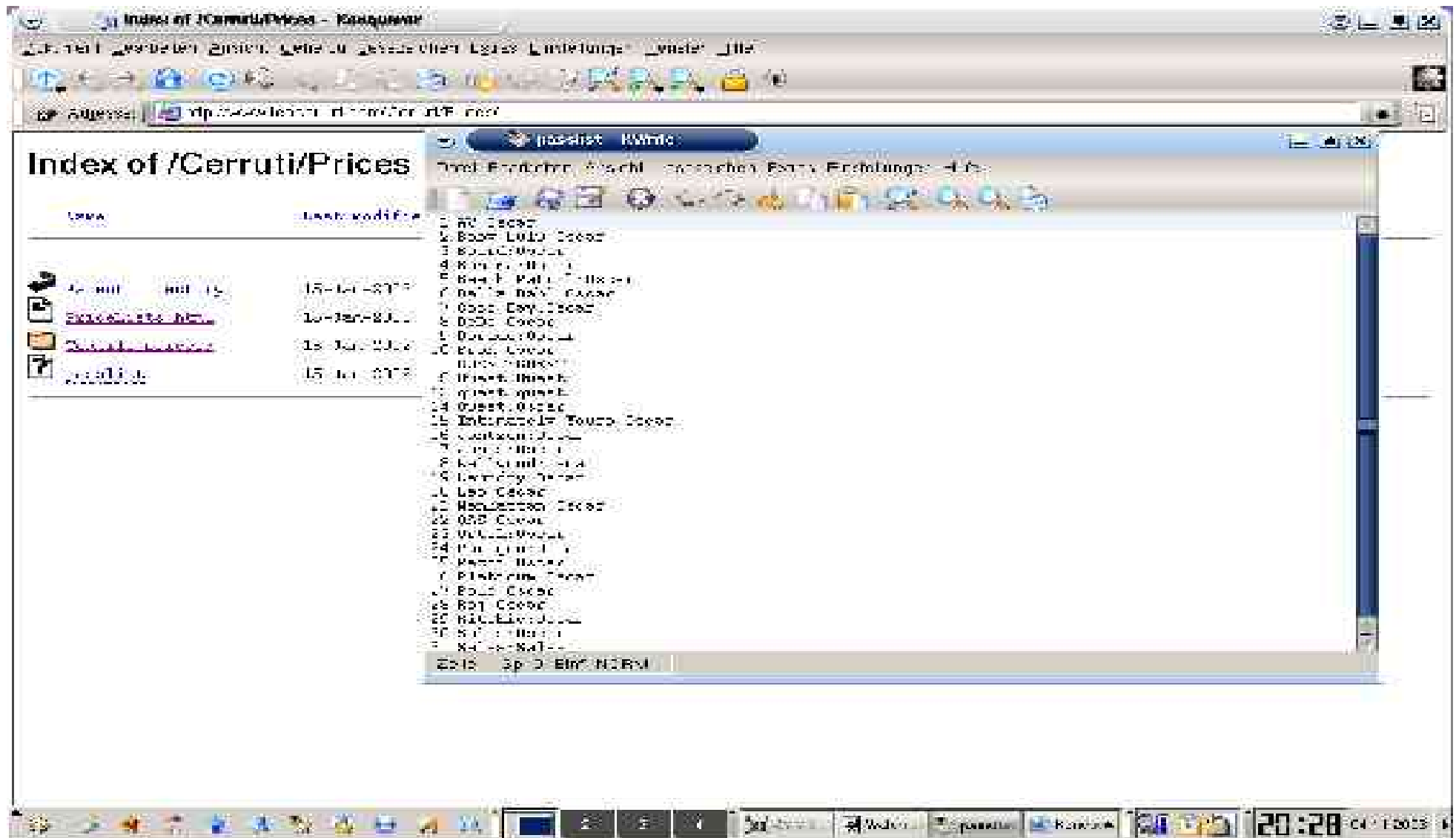
# Google is DANGEROUS/Examples (V)



# Google is DANGEROUS/Examples (V)

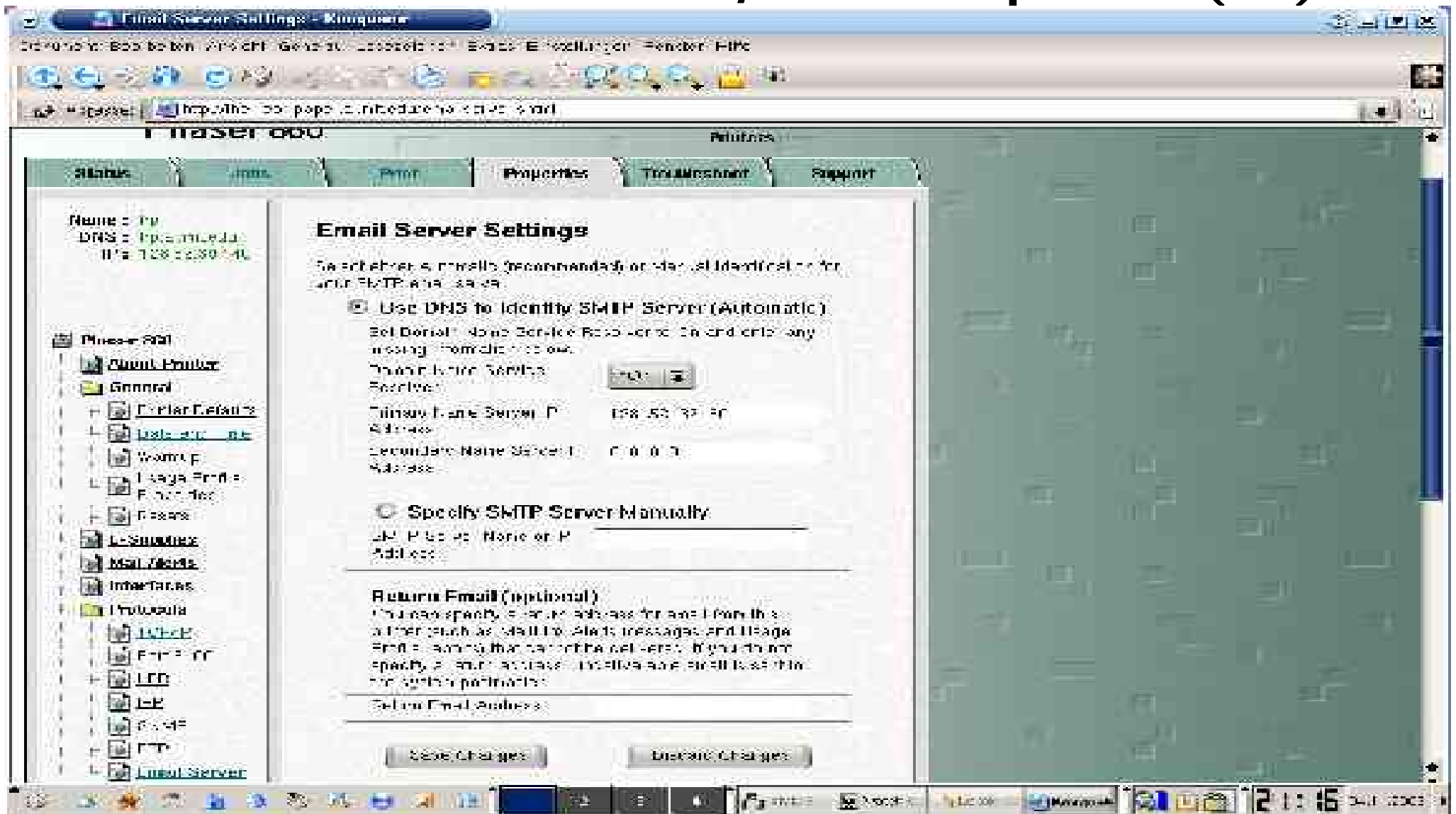


# Google is DANGEROUS/Examples (V)

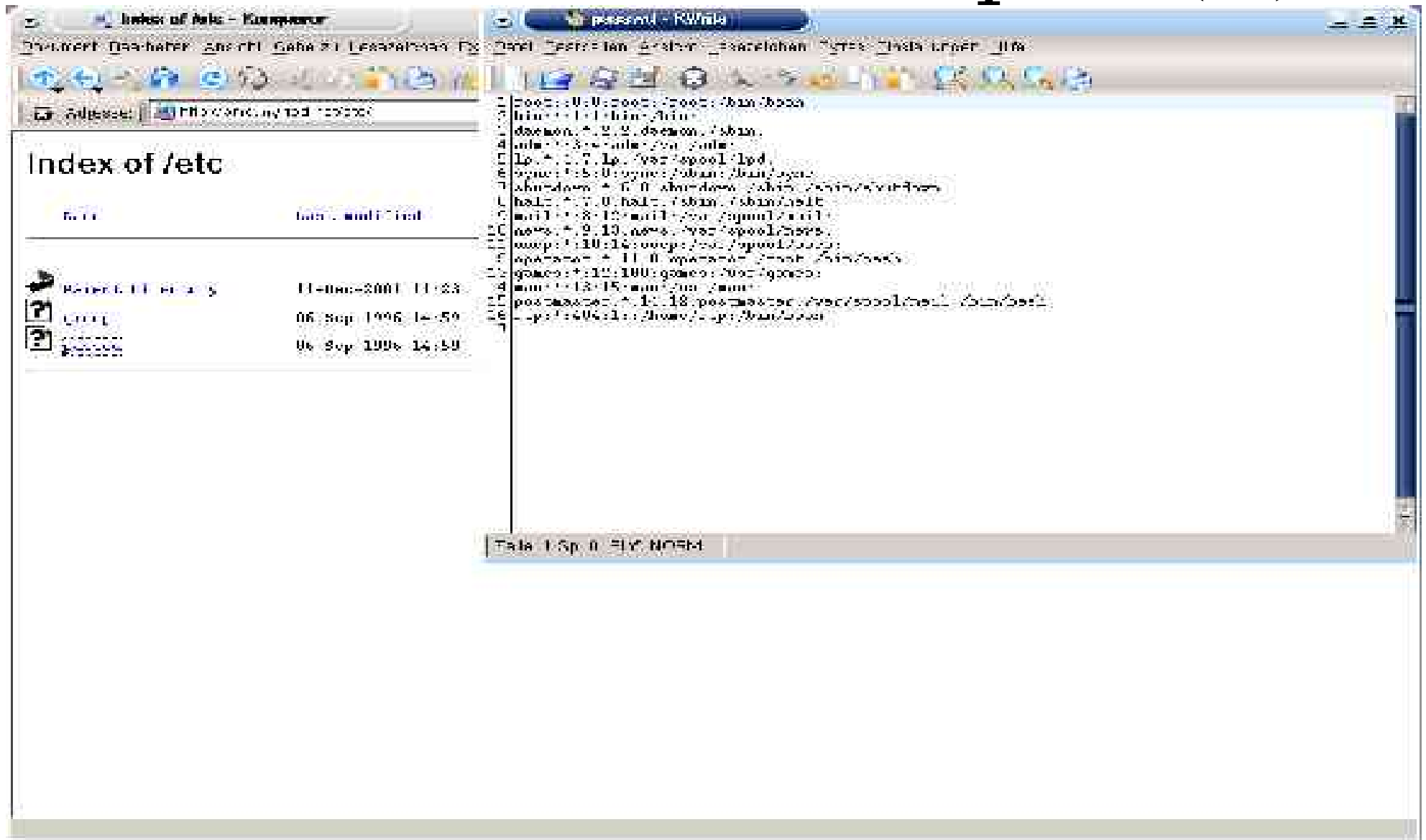




# Google is DANGEROUS/Examples (V)



# Google is DANGEROUS/Examples (V)



# Future prospects

- Google will be used to actively exploit a remote system
- Google will be misused as a transparent proxy
- Google's ranking system will be more and more misused
- ...to be continued.

# Summary

- Google has very much information stored in its databases.
- Most of them should be publicly available on the web but some don't.
- Protect you and your company from accidentally storing sensitive information in Google by controlling the information published
- Hopefully Google will install control mechanisms to prevent misuse.

# Watch out, Google.

That's all folks. Thanks for your (long) patience and attention.

I would now like to answer your questions.

# Watch out, Google.

References/More information:

Johnny Long “Googledorks”,  
<http://johnny.ihackstuff.com>, first  
presented at DefCon 2003 in Las  
Vegas/USA.