What’s this talk about?

- Show the current state of web critics.
- Show what we have to circumvent this.
- Give ideas how we can make the web a better place.

Christoph Lohmann <20h@r-36.net>
"The WorldWideWeb (W3) is a wide-area hypermedia information retrieval initiative aiming to give universal access to a large universe of documents."

- Tim Berners-Lee

Keywords:
- "hypermedia"
- "information retrieval"
- "universal access"
Hypermedia
==========

HTML - parser
CSS - parser
Javascript - parser and runtime
3D and substandards - hardware-dependency
HTTP is too complex to parse but fits its usage well.

WebSocket? Another ghost in the town. But Google is the only market power that could change HTTP.
Universal Access

If you need a quad core to retrieve basic information this is not true anymore.
What is needed to draw a pixel through the web

1.) Parse some URI.
2.) Connect to the webserver.
3.) Send a state-less request adding pixel state through cookies.
4.) Retrieve the HTML.
5.) Parse the HTML.
6.) Do the same steps for all the dependencies.
7.) Parse CSS/JavaScript/whatever.
8.) Run the required dependencies in your runtime and interact via the big DOM model.
9.) Somewhen draw something for the user. Do premature optimisation by outputting half-drawn layouts because dependencies haven’t been downloaded yet.
10.) ...
11.) Be the main cause of global warming. (Not really.)
What’s the suckless answer?
================================

Not just one. Many.
surf

- A compatibility layer to the web.
- Works and is tested.
- Needs more integration with ideas of this talk.
Idea #1: A real commandline interface?
======================================

    cmd -a val1 -b -c val3

    gets

    GET /cmd?-a=val1&-b&-c=val3

Output: Simple text stream of command output.

Problem: Mass adoption does not happen. Sheep follow the shepherd.
Idea #2: Write a script for every website.

- google(1) # Google search
- ddg(1)    # DuckDuckGo search
- cclive(1) # Play flash videos
- translate(1) # Use Google Translate for translations
- curconv(1) # Use Google Finance to convert currencies.

They are domain-specific but work.

Problem: Every website needs a new script. Website providers can easily change their API.
Idea #2: Abandon all hope, use Gopher.

- Gopherfs works
- easy to parse
- superb structurized directory tree

Problems:
  - no MIME types (yeah, just 255 possibilities of selector types)
  - no images
  - will not get any mass adoption
Idea #3: Wait for Google to know what we want.

* It won’t be what we want. We will have to adapt to Google.
Idea #4: Make the web parseable.
=================================
Python mechanize // Perl mechanize
  o works very well
  o add huge dependencies due to the prototyping language

libxml2 // xmlstarlet:
  o they removed the HTML parsing compatibility so you are forced to add
    DTDs - hell

sed + curl:
  o does not parse structured data that well

==> Write suckless tools to handle commandline web sessions.
  o surf integration?
Idea #4: Javascript is ruining our life.

- Most Javascript has a backend which can be accessed and parsed more easily than running a Javascript VM.
- We are running Python, Perl and Ruby in our daily life, so this won’t hurt speed too much.
- Some people use node.js to run daemons.

==> Easy Javascript / DOM access from the commandline is needed.
Idea #4: My theory.

Write example scripts.
Make the level of writing your own scripts lower.
Get the masses of people to work for us.
Make a simple public repository for scripts.
  - Search for domain scripts with a simple utility.
    - Automatic download?
    - surf integration?
Idea #4: Websites that need scripts.

- Wikipedia
- Google News
- suckless.org
Example of web script usage

```
  ddg youtube atb what about us \ 
  |  head -n 1 \ 
  |  xargs quvi --exec "mplayer %%u"
```
Discussion

- Anyone has done similar work?
  - You want to share your scripts?
- Anyone knows some Javascript VM internals to add some commandline interface for DOM?
- Who wants to lead this project?