

Michigan Library Consortium  
Annual Meeting  
Lansing, Michigan  
October 3<sup>rd</sup> 2008

# DansGuardian

---

## Open Source Web Filtering

John Rucker  
Branch District Library

1

Hi, I'm John Rucker, the systems administrator, and recently assistant director, of the Branch District library in Coldwater.

We're a rural library system of 6 locations across Branch county, which is straight down I-69 from here on the border with Indiana.



When I started at Branch District Library a little over five years ago, I was fresh out of university and full of idealism. So I was both surprised and disappointed to learn that we censored our patrons Internet access...



“FILTERED” :-)

...I mean, “Filtered”. We had covered the issues relating Internet filtering in library school, of course, but I guess I just really never thought about having to deal with filtering in real life. Five years later, and a parent myself now, I'm no less disappointed that we censor the Internet, but I do understand why we have to, like it or not.

And, basically, we have to because It's The Law.

As a public library in Michigan which receives Universal Service Fund discounts, Branch District Library has a double mandate to use some sort of mechanism to prevent minors from viewing certain kinds of content.

# Relevant Legislation

- Children's Internet Protection Act (CIPA)
  - Applies to public libraries receiving USF discounts for computer or Internet access
  - Requires a "technology protection measure"
  - Concerned with visual materials
  - Must filter staff and public access computers

4

The first is the Children's Internet Protection Act (CIPA). CIPA was introduced as a bill in the Senate in 1999 by John McCain, and signed into law in 2000 by Bill Clinton.

CIPA applies to all Public libraries and schools which receive Universal Service Fund discounts that go toward the purchase of computers or Internet access.

These institutions are required to use a technological method to filter their own computers for visual materials that are obscene, child pornography--and in the case of children's computers--harmful to minors. CIPA doesn't provide funds to cover the purchase of the filter, of course.

CIPA doesn't specifically state that the filtering should be only on public computers. This has been interpreted to mean, then, that we must filter staff workstations, too.

Librarians were allowed to disable the filter for an adult (defined as 17 or older) to view obscene material, provided that it was for research or other lawful purposes.

The American Library Association challenged CIPA soon after it became law. The case made it to the Supreme Court in 2003. That summer the court upheld CIPA as constitutional, provided that librarians disable filter immediately if asked by an adult patron. And after the court's ruling librarians can't inquire of the patron why they want the filter turned off.

# Relevant Legislation

- Michigan Library Privacy Act
  - All “matter”, not just visual depictions
  - Networks, not just Library PCs
  - A “policy” and “method”, not necessarily a technological solution.
  - Applies to all public libraries

5

At the state level, we have the Michigan Library Privacy Act. In 2000 Section 6 was amended to the Privacy Act. This is a provision to impose restrictions on minors' Internet access in public libraries.

The Privacy Act is concerned with “matter” that is obscene and harmful to minors, not just visual depictions as in CIPA.

It also says that the “computer networks” libraries provide to the public must be restricted. So we're talking not just the library's own computers here, but any wireless or wired networks the library provides to the public, too.

The good news is that the Privacy Act doesn't require any expenditure on technological filtering methods. Libraries just need a policy in place and a method to implement it. As with CIPA, adults may use unfiltered computers, though here adults are defined as those 18 and older.

The bad news is that it applies to all public libraries, not just those receiving some sort of financial subsidy.

(For all you academic libraries out there at this point, obviously neither CIPA nor Section 6 of the Michigan Library Privacy Act apply to you. But stick around, I still think you'll find it interesting to see how free software can come to the aid of libraries in this regard.)

Disclaimer:

IANAL

OK, Now for the disclaimer: I Am Not A Lawyer. These are the facts as I understand them after much research and advice from talented presenters at other MLC and MLA functions on this topic, and from Branch District Library's lawyer.

As always, you should consult with your library's own legal counsel before embarking on any change of your policy or procedure.

# The Requirements

- Filter all PCs, whether owned by library or patron
- Filter text and images
- Disable on request

7

However, it doesn't take a lawyer to parse the difficulties imposed on a small library like us by these two overlapping laws.

We must have a filter in place on all staff and public computers, and any wireless or wired network the public can use their own devices on.

We must filter both text and images falling under these laws' definitions of obscene and harmful (and we're definitely not going down the path of interpreting those definitions in this talk...).

We must disable the filter immediately if an adult requests it, no questions asked.

# Commercial Filters

- NetNanny
- CYBERSitter
- CyberPatrol
- Smart Filter (Formerly N2H2)
- Websense (Formerly SurfControl)
- Norton and McAfee products

8

Earlier I said I was surprised that “we” were censoring the Internet at Branch District Library... Well, specifically \*we\* didn’t do the censoring, some company did.

When I started at BDL, we used a shall-remain-nameless bit of software which didn’t really accomplish any of the three requirements we just discussed.

It was also difficult to administer, it had to be installed individually on all public computers, and had an annoying tendency to block plenty of content that was completely inoffensive.

This product, and pretty much all of the commercial filtering products and services out there, some of which are listed here, give only an indication of how or what they are going to block. Broad categories, really, but finding an actual list to browse is difficult or impossible.

In the cases where some of the proprietary blacklists have been leaked, it has been discovered that the lists often have a tendency to further a political point of view under the guise of protecting children from porn.



# Costs:

- BDL PCs: 60 Public, 30 Staff
  - Workstation installed: \$30 - \$60 each/year
    - \$2,700 - \$5,400 annually
    - Can't filter wireless network
  - Network-wide filters = \$\$\$

9

And to top it all off, the cost of these products can be daunting, even for a small library like Branch District Library with only about 60 public and about 30 staff computers. (Remember, all workstations must be filtered.)

The filters that need to be installed on each workstation retail from \$30-\$60 per computer per year. So that's \$2700 - \$5400 a year. Even if you negotiate a volume discount, the cost will still be significant.

But it doesn't matter, because we want to offer wireless access to our patrons, and we can't very well install (and pay for) a copy of a filter for every patron laptop that comes in the door.

The products that can completely integrate into your network without individual PC installs don't actually tell you how much they cost up front. Call a sales rep and get a quote.

At Branch District Library, my annual tech budget is woefully small, and mostly goes to hardware anyway, not software. So, for us, the network filters were definitely a case of "if you have to ask, you can't afford it."



open source web filter

Search

Web

Results 1 - 10 of about 951,000 for

## [DansGuardian - True Web Content Filtering for All](#)

Offers phrase matching, url filtering and PICS filtering. News, FAQ, mailing list, screenshots. [Unix, Linux, BSD]

[dansguardian.org/](#) - 15k - [Cached](#) - [Similar pages](#) - [Note this](#)

[Docs & HOWTO](#)

[Extras & Add-ons](#)

[Download DansGuardian](#)

[Screen Shots](#)

[What is DansGuardian](#)

[Introduction](#)

[Blacklists](#)

[SmoothWall](#)

[More results from dansguardian.org »](#)

So, none of the commercial offerings seemed very palatable to me. As a techie librarian at a not-particularly-wealthy library, I needed a system that was effective, open and transparent in its functioning, under our control, and as inexpensive as possible.

I've been a longtime supporter and user of open source software, so I figured that the community must have stepped up to the plate in this regard like it had in so many other areas of computing. So I hit the search engines.

Thankfully, I didn't have to search long. Even 5 years ago, the first hit in Google for "open source web filter" was DansGuardian.



“DansGuardian is designed to be completely flexible and allows you to tailor the filtering to your exact needs. It can be as draconian or as unobstructive as you want. The default settings are geared towards what a primary school might want but DansGuardian puts you in control of what you want to block.”

- <http://dansguardian.org/?page=whatisdg>

DansGuardian had all the features we needed, and at free, the price was right.

DansGuardian is a software package first written in 2001 by Daniel Baron, a Briton, because he felt the other open source filters out there at the time were “rubbish”.

From the moment I got DansGuardian installed it was obvious that indeed it was not rubbish. And, as Dan’s own words on the screen indicate, it can be as strict or as lenient as you want it to be.

The amount of control that DansGuardian gives you in filtering content is truly amazing. Primarily, you can use it to block pages with pornographic content. But you can do so much more if you’re inclined.

You can block any file type you want, to prevent music and movie downloads. You can disable Flash and Java applets. You can have it remove all popup windows, or any other annoying JavaScript. You can block a wide array of non-porn sites, including games and chat sites. You can censor-out swear words. You can scan everything anyone downloads for viruses. And so much more.

And all of those features I just listed... well, you can have multiple instances of DansGuardian running to have different levels of filtering. So, having the bare minimum for your adult computers, and a stricter filter for your children’s computers, is a simple thing.

On features, DansGuardian comes up equal or better than all the commercial offerings. In price, it wins hands down. That it’s open source and fully in your control is just gravy.

# DansGuardian uses:

- URL pattern matching
- Community-provided blacklists
- PICS filtering
- term weighting



DansGuardian uses several methods for determining if a page should be blocked. The major ones are:

URL pattern matching, community-provided blacklists with support for 3-rd party commercial lists, PICS filtering (a sort of self-policing by honest web sites), and primarily term weighting, a concept many of us here are very familiar with.

Because the term weighting deals with words, obscene text will get blocked as well as visual materials, keeping us in compliance with the Michigan Library Privacy Act.

The term weighting algorithm DansGuardian uses takes into account the medical and educational uses of terms that might otherwise accompany pornography in an effort to avoid over-blocking.



  
SOLARIS

The Solaris logo features a stylized sun with orange rays above the word "SOLARIS" in a blue, sans-serif font.

DansGuardian runs only on Unix-like systems, so Linux, Solaris, the BSDs, and Macintosh OS X, and so on.

So with the possible exception of the Mac, the barrier to entry was a bit high in terms of knowledge or hardware cost for many librarians, especially those in small, poorer libraries like ours that I wanted to help.

My first installation of DansGuardian was on OpenBSD. It worked great, but it wasn't going to get the attention of the average librarian any time soon.

# ubuntu



14

By 2006, things had changed in a big way. Ubuntu, a Linux company launched in late 2004 by South African billionaire Mark Shuttleworth, had in a very short time made quite a name for itself as a highly-polished, user-friendly operating system.

After trying it, I found that Ubuntu actually lived up to the hype. (As an aside, Ubuntu continues to grow. You can now buy computers pre-loaded with Ubuntu from Dell, for example, among other places.)



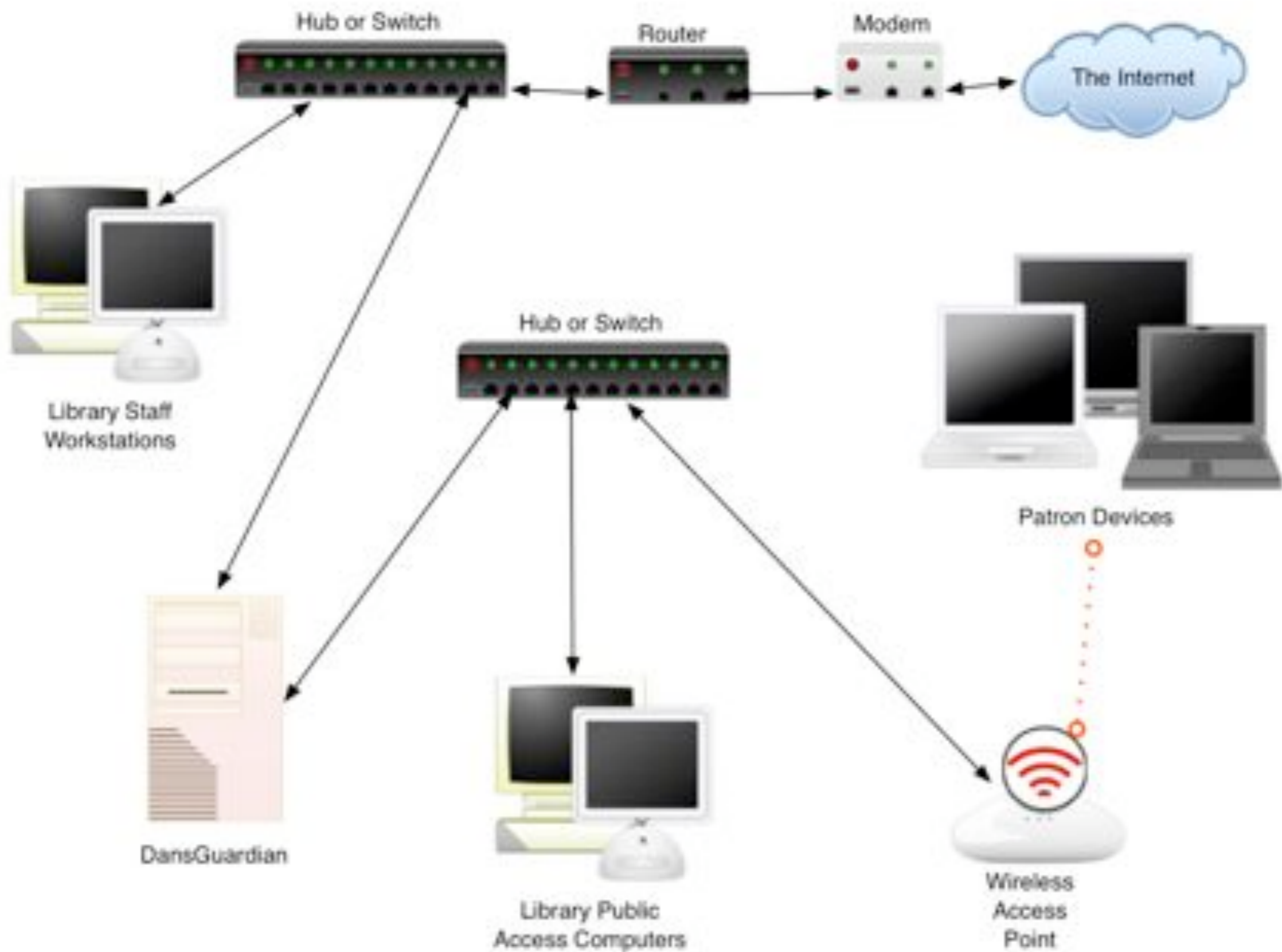
- 8 years old
- Pentium III  
800 MHz
- 256 MB RAM



So, I had finally found the combination that would work best for us:  
Ubuntu + DansGuardian made for one happy librarian.

We had a number of computers left over from the Gates Foundation grant back in 2000. The terms of the grant were fulfilled so we could do whatever we wanted with them. These machines didn't have a lot of horsepower and didn't run Windows XP well enough for us to want to keep them around for our Windows computers. But they ran Ubuntu just fine.

I added some wireless access points, and for very little money we now had a good infrastructure in place.



With my DansGuardian server sitting between the public and staff parts of our network, we could use it to filter content for both sides.

And since all Internet traffic from our patrons must go through the DansGuardian server, anything anyone brings in, from iPhone to Laptop, gets filtered without us doing a thing.

I could have just used one server for all our locations, but I installed them at each branch because, as a side effect of the way DansGuardian works, popular web pages are cached on the local network. This speeds up the Internet for everyone and lowers our bandwidth usage.

After every thing was in place and working well, I made my first tutorial to help other librarians do what I had done. That was about two years ago.



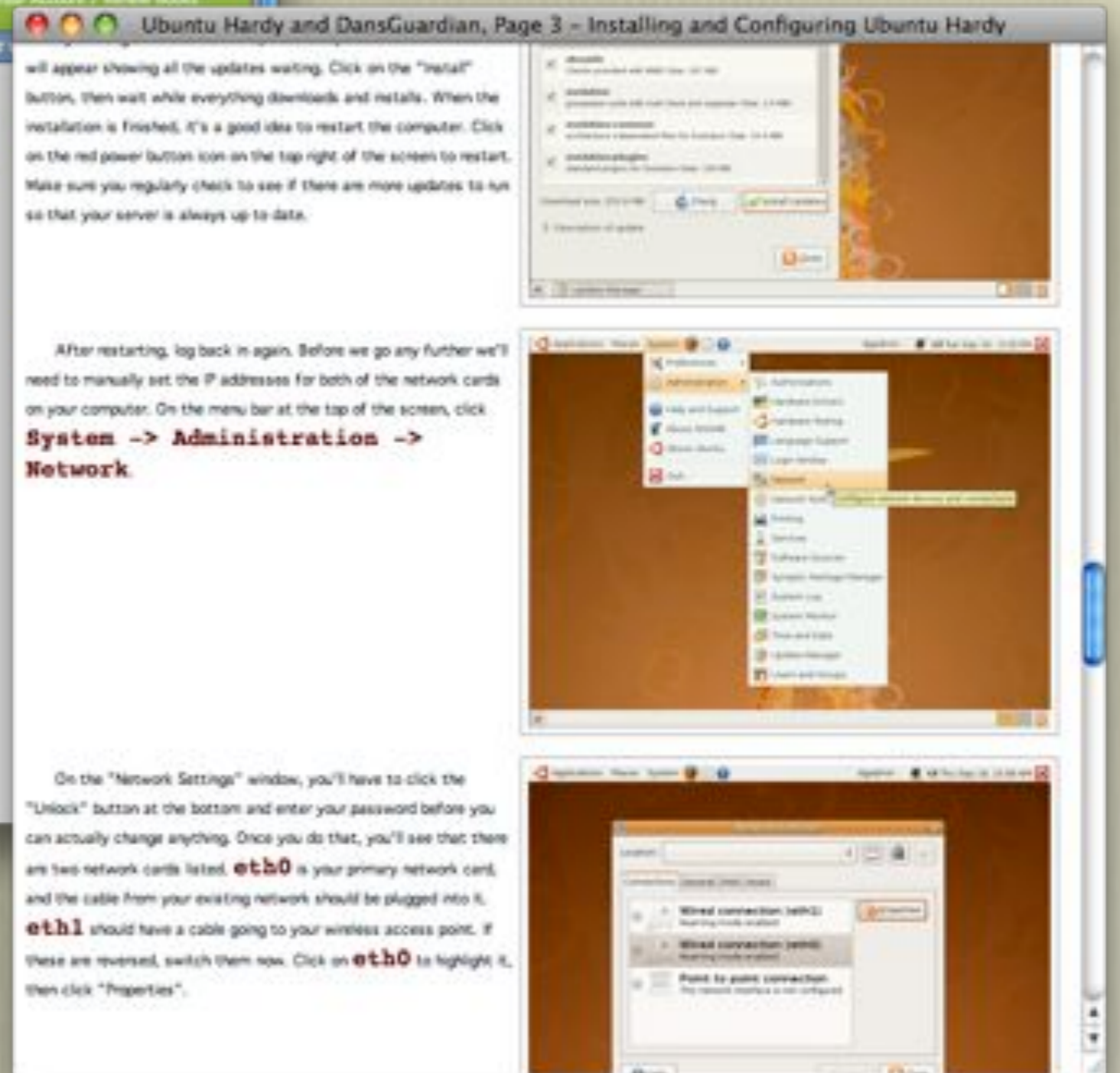
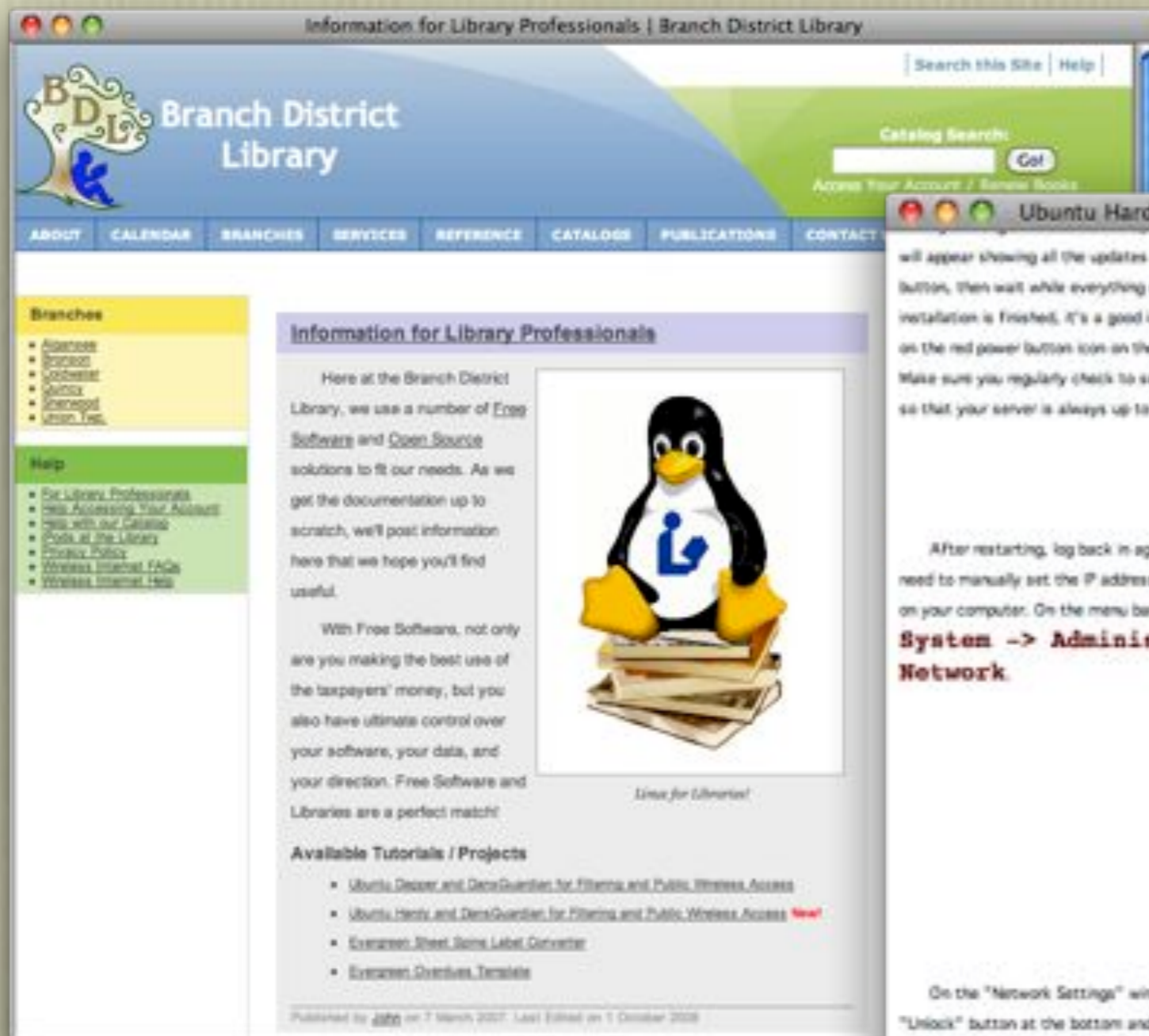
```
root@filter: /etc/dansguardian1 — ssh
root@filter: /etc/dansguardian1# ls
bannedextensionlist  dansguardianfl.conf  filtergroupslist
bannediplist         exceptioniplist      greysitelist
bannedmimetyplist   exceptionphraselist  greyurlist
bannedphraselist    exceptionsitelist    languages
bannedregexpurllist exceptionurllist     generalrate
bannedsitelist      exceptionuserlist    pics
bannedurllist       exceptionvirusextensionlist transparent1x1.gif
banneduserlist      exceptionvirusmimetyplist weightedphraselist
contentregexplist  exceptionvirussitelist
dansguardian.conf  exceptionvirusurllist
root@filter: /etc/dansguardian1# vi contentregexplist
root@filter: /etc/dansguardian1# /etc/init.d/dansguardian1 restart
* Shutting down dansguardian1...
                                [ ok ]
* Starting dansguardian1...
                                [ ok ]
root@filter: /etc/dansguardian1# vi contentregexplist
root@filter: /etc/dansguardian1# /etc/init.d/dansguardian1 restart
* Shutting down dansguardian1...
                                [ ok ]
* Starting dansguardian1...
                                [ ok ]
root@filter: /etc/dansguardian1#
```

My first tutorial didn't actually use the polished interface that Ubuntu was famous for, though. For ease of administration by those more technically inclined, I opted to go with the command-line, or terminal, interface. Though it looks sparse, it's really very quick and powerful, once you are comfortable with it.

Pretty soon I started getting positive feedback on my tutorial from all over, and people found it useful. My tutorial is actually the first hit if you Google "ubuntu and dansguardian". The feedback I got was mostly from tech-savvy parents, though, who were using DansGuardian for their own homes.

Feedback that I got from librarians was often in the realm of, "Well, DansGuardian is nice and all, but this terminal business is still too scary for us."

So this year, after Ubuntu had another major release this past April, I decided to revise my tutorial using the latest Ubuntu and all the familiar graphical interfaces that most of us expect and are comfortable with.



[www.branchdistrictlibrary.org/professional](http://www.branchdistrictlibrary.org/professional)

And just in time for this meeting, I finished it!

Of course, you are not here today to get a description of this process. And there's not time to do so, anyway. So for the next few minutes I'll just give you a quick peek of how everything works. But if you like what you see and want to try it out yourself, be sure to read my tutorial.

Here's the address. I'll leave it up at the end of the presentation.

My tutorial is a detailed, step-by-step walkthrough from the planning stages on up. It covers concepts, hardware choices, firewall issues, setting up and customizing the web filter, and more.

It's loaded with pictures every step of the way. And as for that "scary" terminal: you'll have to use it for only one small command, which you can even copy and paste.

# Director:



“Wait!” you say, “still sounds scary.” Let me reassure you that it’s not! If you can follow some step-by-step directions, you can have a powerful web filter up and running before one work day is out.

How do I know? Earlier this week I sat my director down with a computer and this new tutorial and asked him to go through it, start to finish, without my help.

He’s not the type who’s afraid of computers...but he is a Mac user, so... :-)

A couple of hours later, he had a system working perfectly.

So, if a library director can do it, anyone can!

(Thanks to my director, Bruce Mills, for being such a good sport.)

# Access has been Denied!

Access to the page:

<http://www.google.co.th/search?q=porn&ie=utf-8&oe=utf-8&aq=t&rls=com.ubuntu:en-US:unofficial&client=firefox-a>

... has been denied for the following reason:

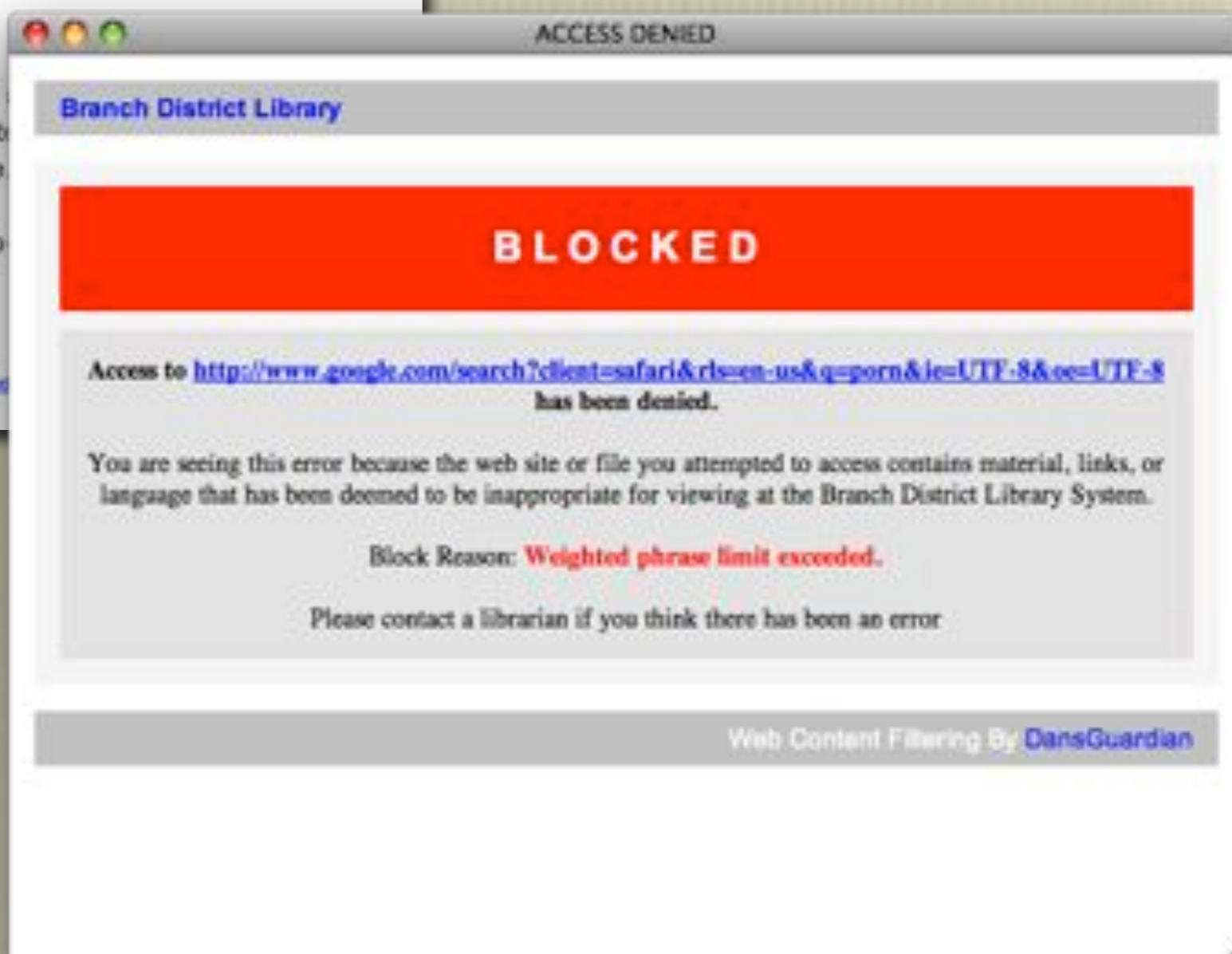
**Weighted phrase limit exceeded.**

You are seeing this error because what you contain, or is labeled as containing, may be inappropriate.

If you have any queries contact your ICT Co

YOUR ORG NAME

Powered by [DansGuardian](#)



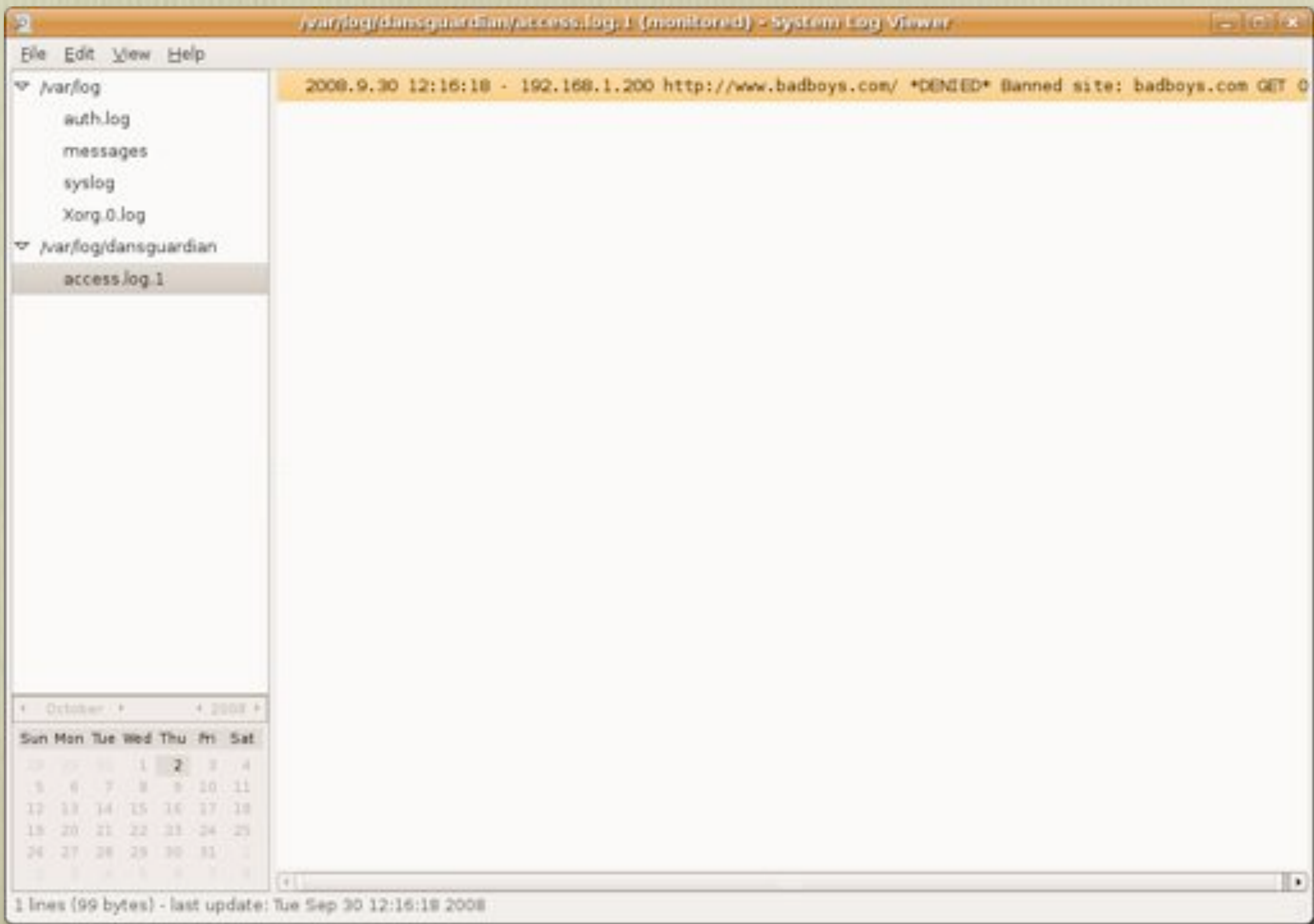
20

So after a couple of hours to set everything up, the system just works and you can forget about it.

When patrons try to view blockable material, they will see a banner stating the page was blocked.

The default is on the left and our customization is on the right.

You can make your blocked notice look any way you want, and say as much or as little as you like as to why the page was blocked.

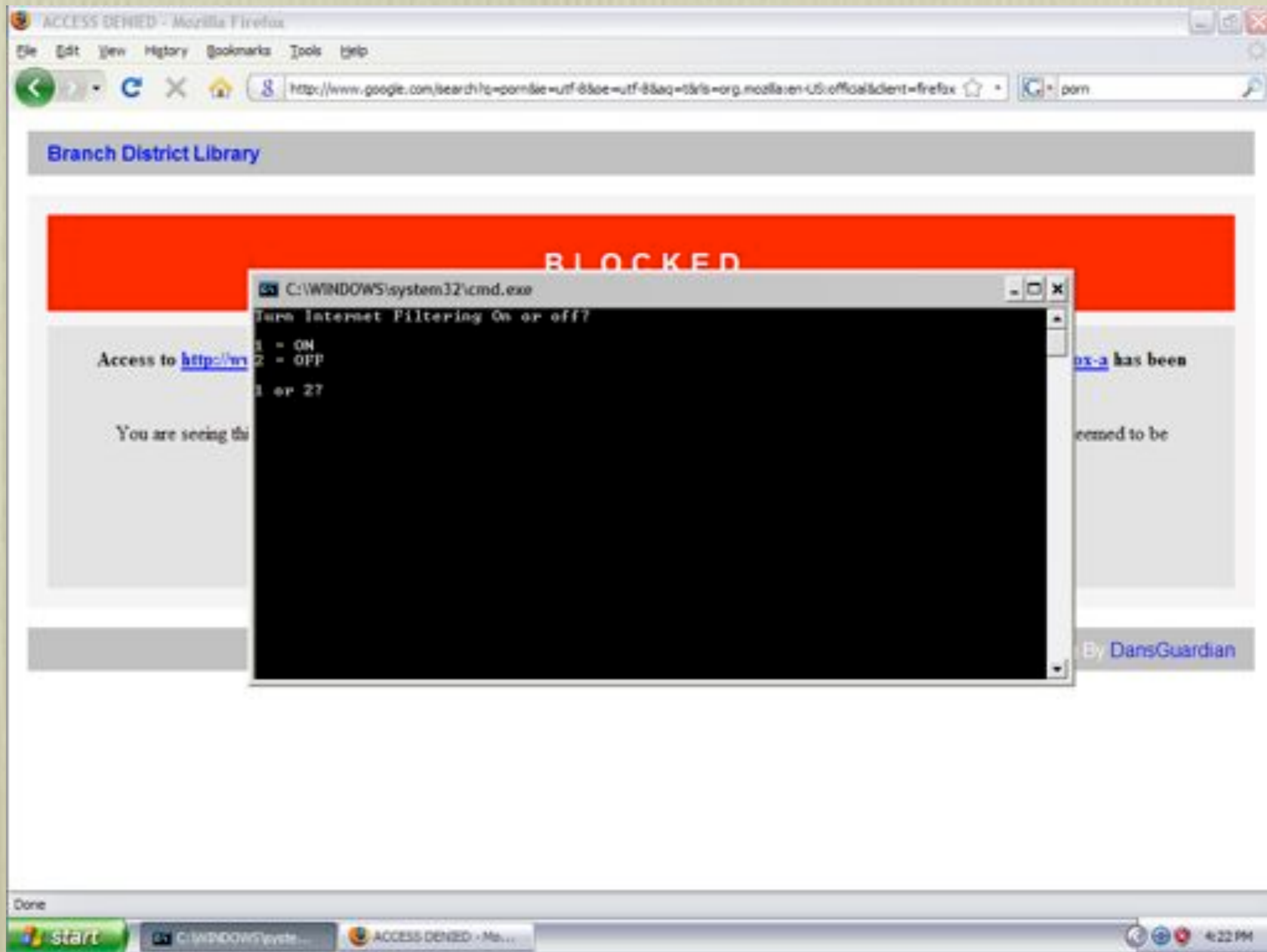


If the page shouldn't have been blocked, you can add it to a white list in seconds, and come back later if you want to tweak things further to prevent similar errors.

You can do the same if a page gets through that should have been blocked.

Throughout this whole process, patron privacy is safeguarded: As shown here on the screen, logs are kept only of what gets blocked.

# Easy for Staff to Disable



22

If an adult patron requests the filter to be turned off, you can do so quickly, just for that specific computer from within the DansGuardian controls, whether a library computer or a patron's device.

Or, you can use a cheap USB flash drive as a "key" to quickly disable the filter on the library's own computers, and then enable it again later.

The tutorial also details steps you can take to prevent patrons from disabling the filter themselves on your library computers.

# With DansGuardian, You Will:

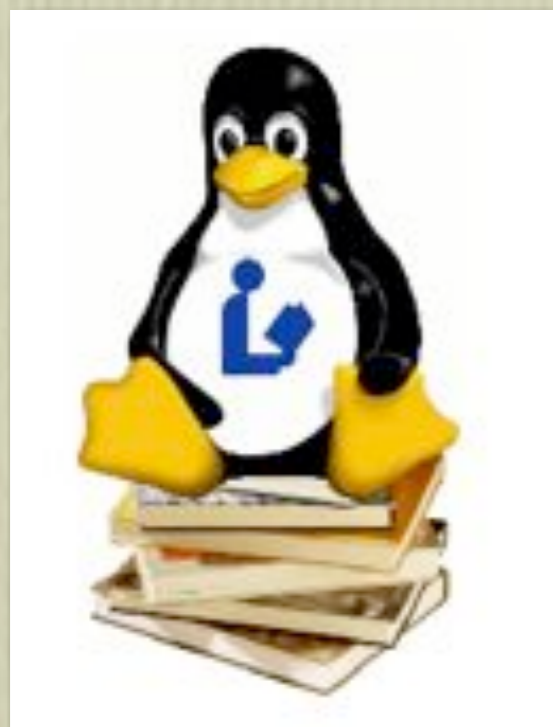
- Comply with the law
- Be in control
- Be in the know
- Save money



So, to sum up, with DansGuardian you will:

- \* Be in compliance with CIPA and the Michigan Library Privacy Act
- \* Completely control the level of Internet access you allow for your patrons and staff
- \* Know exactly what is getting blocked and why.
- \* Get quite a bargain with \$0 cost for software and the potential to give old hardware a new lease on life

Yet you're not getting a cheap product. DansGuardian will make you a happy librarian.



*Linux for Libraries!*

John Rucker

[www.branchdistrictlibrary.org/professional](http://www.branchdistrictlibrary.org/professional)