

# Building AltLaw.org



Stuart Sierra

Columbia Law School Program on Law & Technology

<http://columbialawtech.org/>

**A long time ago in  
a galaxy far, far away...**

# ProjectPosner.org



**Tim Wu**

Prof. of Law, Columbia University  
Former Posner clerk

**Project  
Posner**

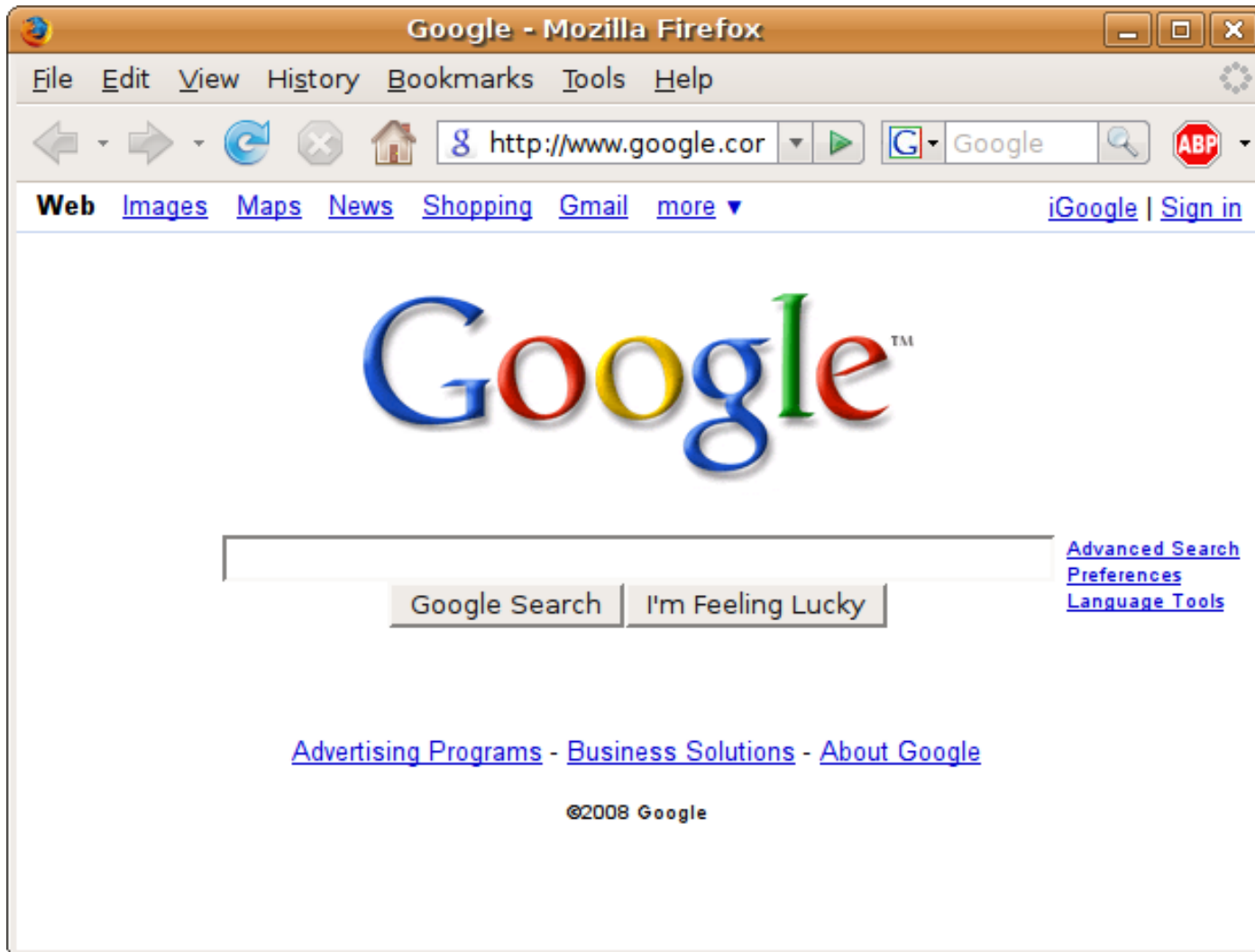
*The opinions of Judge Richard A. Posner*

# That Was Then

The screenshot shows a Mozilla Firefox browser window titled "Welcome to Westlaw - Law School - Mozilla Firefox". The address bar displays "http://web2.westlaw.com/welcome/LawSchool". The browser's menu bar includes "File", "Edit", "View", "History", "Bookmarks", "Tools", and "Help". The Westlaw website header features the logo and navigation links: "FIND & PRINT", "KEYCITE", "DIRECTORY", "KEY NUMBERS", "COURT DOCS", "SITE MAP", "HELP", and "SIGN OFF". Below the header are tabs for "Law School", "Westlaw", "Business & News", and "New York", along with an "Add/Remove Tabs" link. The main content area is divided into several sections:

- Shortcuts**: Includes "ALR - A Westlaw Exclusive" with a link to "American Law Reports" and a description: "In-depth analysis of all caselaw relevant to your specific point of law." It also features a "Find by citation:" search box with a "Go" button and a "Find using a template Publications List" link.
- Finding Tools**: Contains a "Find a Case by Party Name" link.
- KeyCite this citation:**: Includes a search box with a "Go" button.
- Search for a database:**: Includes a search box with a "Go" button and dropdown menus for "Recent Databases" and "Favorite Databases".
- Resources**: A central column with sections for:
  - My Personal Databases**: Includes a description and links for "U.S. Supreme Court Cases".
  - Cases**: Includes links for "All Federal", "All States", "Cases by State", and "Additional materials".
  - Statutes**: Includes links for "US Constitution", "State Constitutions for the 50 states and D.C.", "All Federal", "All States", "Statutes by State", "Additional materials", and "50 State Surveys".
  - Administrative Materials**: Includes a link for "Code of Federal Regulations".
- Secondary Sources**: Includes links for "Black's Law Dictionary", "American Jurisprudence (Am Jur)", "Am Jur Proof of Facts", "American Law Reports - ALR", "Causes of Actions", "Journals and Law Reviews", "Restatements", and "Additional materials".
- Forms**: Includes links for "All Forms", "Am Jur Legal Forms", "Am Jur Pleading and Practice Forms", "Annotated Federal Procedural Forms", "National Pleading and Practice Forms", "West's Federal Forms", "West's Legal Forms", and "Additional materials".
- News**: Includes links for "All News", "New York Times", and "Thomson Financial News".

# This Is Now



**“Making the Common Law  
a bit more common.”**

# Crawling Court Web Sites



Paul Ohm, U. of Colorado Law School

<http://paulohm.com/>

<http://www.supremecourtus.gov/>

<http://www.ca1.uscourts.gov/>

<http://www.ca2.uscourts.gov/>

<http://www.ca3.uscourts.gov/>

<http://www.ca4.uscourts.gov/>

<http://www.ca5.uscourts.gov/>

<http://www.ca6.uscourts.gov/>

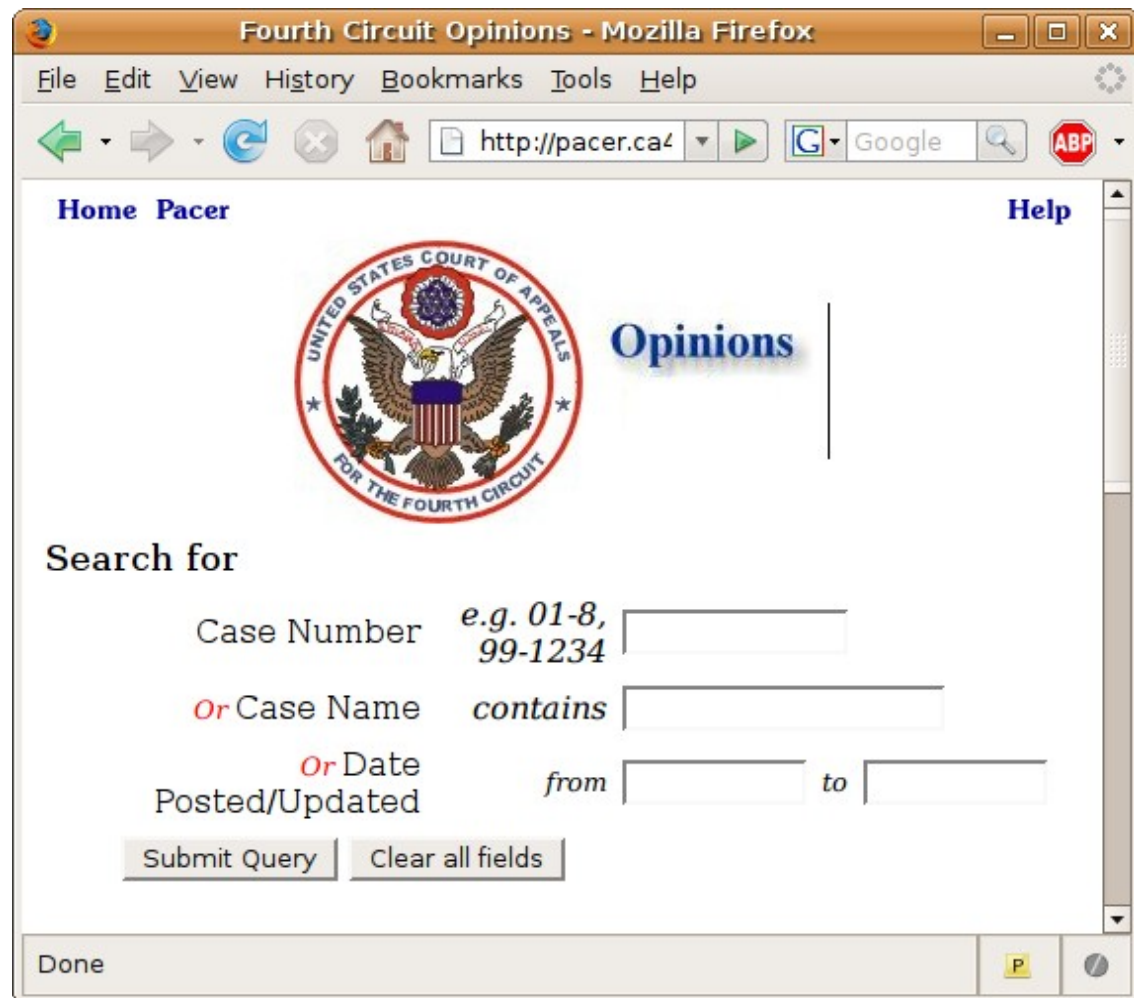
<http://www.ca7.uscourts.gov/>

<http://www.ca8.uscourts.gov/>

<http://www.ca9.uscourts.gov/>

<http://www.ca10.uscourts.gov/>

<http://www.ca11.uscourts.gov/>



Fourth Circuit Opinions - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://pacer.ca4

Home Pacer Help

UNITED STATES COURT OF APPEALS  
FOR THE FOURTH CIRCUIT

Opinions

Search for

Case Number *e.g. 01-8, 99-1234*

*Or* Case Name *contains*

*Or* Date Posted/Updated *from*  *to*

Submit Query Clear all fields

Done

# A Little Help From Our Friends



Cornell University  
Law School

<http://www.law.cornell.edu/>

LII / Legal Information Institute



<http://justia.com/>



<http://bulk.resource.org/courts.gov/>



**Building a Search Engine**

**Easy, right?**

# altlaw.org, Take 1

Web server



Apache

Application



Database  
& Search



# You don't need a database

- Any Java servlet container, e.g. *jetty://*

The Solr logo is written in a bold, orange, sans-serif font.

- Caching
- Maintains pool of Lucene threads
- Tokenization, stemming, synonyms

The Lucene logo is written in a green, cursive, script font.

- Searching
- Indexing

# altlaw.org, Take 2

Web server



Apache

Application



Database  
& Search

Solr

**Big Data, Big Love**

# The Big Join

## CASES

435 F.2d 1, United States v. Dean, ...  
313 F.2d 236, United States v. Walker, ...  
452 F.2d 93, John v. Smith, ...  
452 F.2d 1135, United States v. Augello, ...  
501 F.2d 88, Smith v. Jones, ...

## CITATION LINKS

435 F.2d 1 cites 378 F.2d 359  
435 F.2d 1 cites 313 F.2d 236  
435 F.2d 1 cites 168 U.S. 382

## REVERSE CITATIONS

435 F.2d 1 *is cited by* U.S. v. Chagra, 669 F.2d 241  
435 F.2d 1 *is cited by* U.S. v. Haley, 452 F.2d 391

# How *not* to process a million documents

XML



Berkeley DB

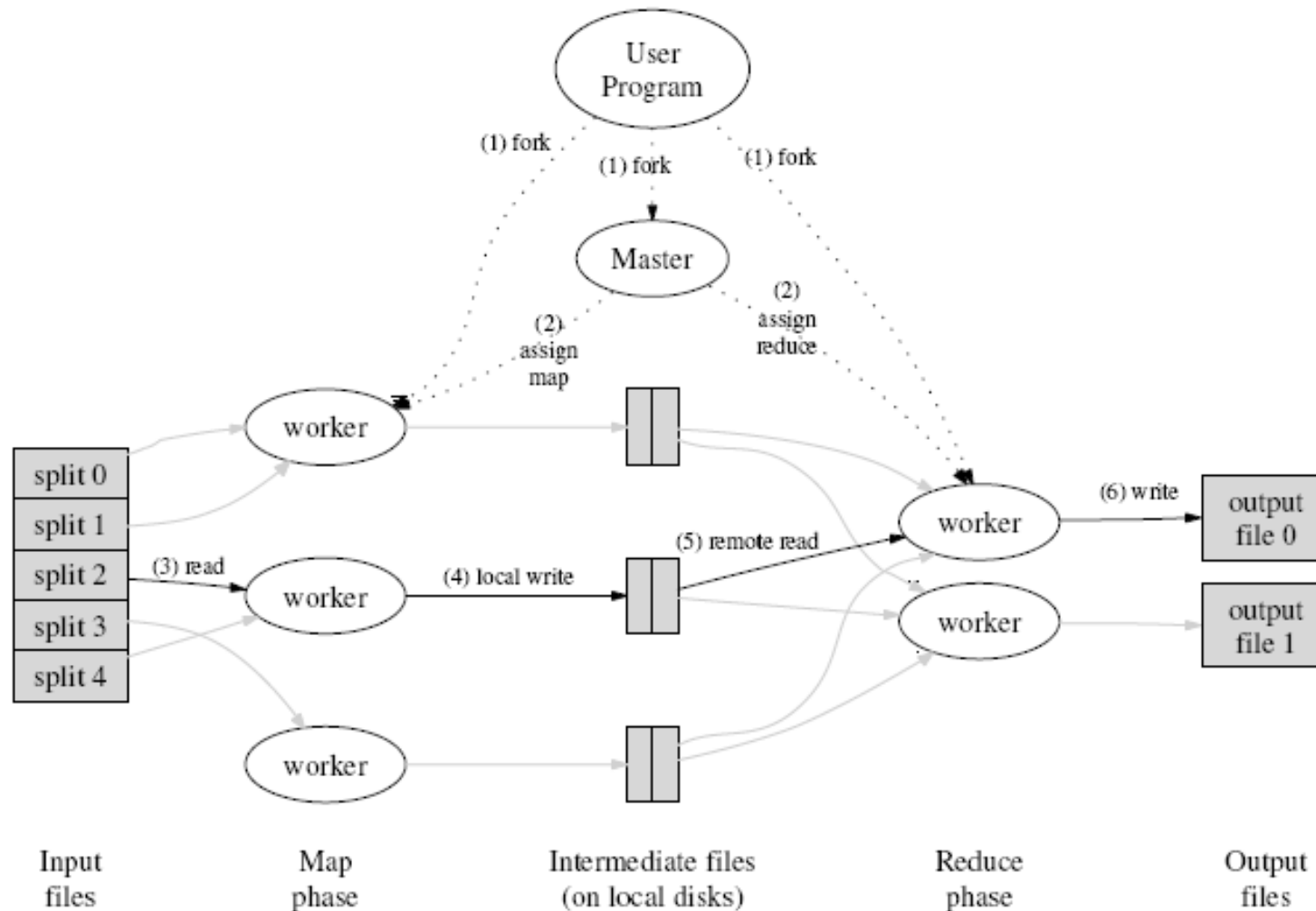


*Lucene*

Solr



# MapReduce: it works for Google...



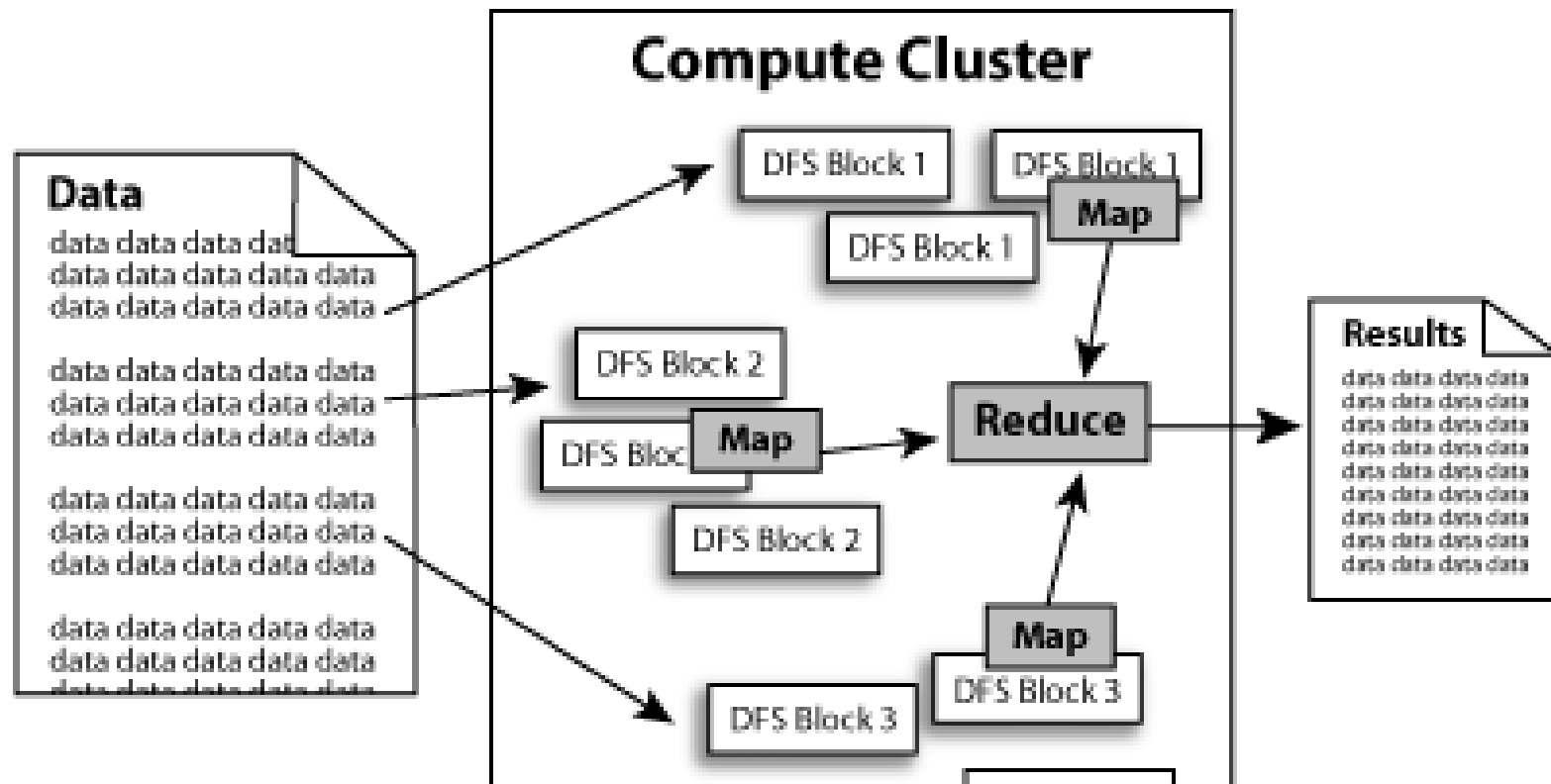
- <http://code.google.com/edu/parallel/>



# Hadoop: it works for Yahoo...



<http://hadoop.apache.org/>



# AWS: it works for Amazon...



<http://www.amazon.com/gp/browse.html?node=3435361>

- Elastic Compute Cloud (EC2):  
\$0.10 per hour, per machine
- Simple Storage Service (S3):  
\$0.15 per GB, per month

**Issues**

**(We've all got them.)**

# Medium-Neutral Citation

- West:  
123 F.2d 456
- public.resource.org:  
[4.F3d.26.93-1236.93-1171.html](#)
- AALL Universal Citation Guide:  
1989 US App. (4th) 104 ¶ 52  
<http://www.aallnet.org/committee/citation/ucg/>
- court + date + docket
- AltLaw:  
<http://altlaw.org/v1/cases/685561>

## Ideal Neutral Citation

- unique
- unchanging
- medium-neutral
- discernible from the case text alone
- available as soon as case is decided
- public-domain
- machine-readable

# Privacy

- “Almost the only results on Google now for my mother's name are your site .... She has been applying for new jobs lately, and I have no doubt that this has hurt her prospects.”
- “[W]e have received several threatening message regarding the materials you have put at your website regarding our case. ... This materials needs to be remove from your website ASAP.”
- “This serves no purpose but to open up my own ancient history via the search engines and cause me to suffer unduly all over again. ... I stumbled into a business opportunity that lasted only 7 months in 1987, and it spiraled into a Kafkaesque nightmare ....”
- Policy: <http://altlaw.org/v1/about/why>

# The Future

- More cases  
(district courts, state supreme courts, ...)
- PageRank applied to citations
- Accept corrections
- Commentary from legal scholars
- More research tools

[LawCommons.org](http://LawCommons.org)



# Law

## BETA

*The free legal search engine — over 700,000 documents.*

Enter a case name, citation, or key words and phrases:

[About AltLaw](#)

[Advanced Search](#)

[Coverage](#)

[Browse Cases](#)

[Browse U.S. Code](#)