



Free Law Project

<http://freelawproject.org>

Brian W. Carver
&
Michael Lissner

Columbia University Web Archiving Conference
June 4, 2015

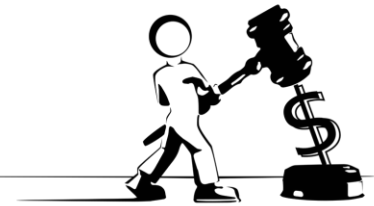
Free Law Project



- Free Law Project is a California non-profit that seeks to:
 - provide free online access to primary legal materials,
 - develop online legal research tools, and
 - support academic research on legal corpora.
-
- Co-founders:
 - Michael Lissner &
 - Brian W. Carver



Free Law Project



- Free Law Project sponsors the development of:
- CourtListener, <https://www.courtlistener.com>

COURT LISTENER
About Coverage Sign in / Register

Advanced Search

Refine Your Query

Select Jurisdictions
All jurisdictions selected

Case name

Judge

Precedential Status

- Precedential (2,221,450)
- Non-Precedential (285,915)
- Errata (291)
- Separate Opinion (84)
- In-chambers (12)
- Relating-to orders (88)
- Unknown Status (5)

Filed After

Filed Before

2,517,835 Results Newest first

30ms

Ranjit Singh v. Eric Holder, Jr. (6th Cir. 2014)
Date Filed: May 13th, 2014 Status: Non-Precedential Docket Number: 13-3380
NOT RECOMMENDED FOR PUBLICATION File Name: 14a0362n.06 No. 13-3380 FILED May 13, 2014 DEBORAH S. HUNT, Clerk UNITED STATES COURTS OF APPEALS ...

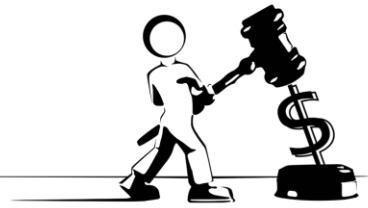
Ice House America, LLC v. Charles Cardin (6th Cir. 2014)
Date Filed: May 13th, 2014 Status: Precedential Docket Number: 13-5764
RECOMMENDED FOR FULL-TEXT PUBLICATION Pursuant to Sixth Circuit I.O.P. 32.1(b) File Name: 14a0099p.06 UNITED STATES COURT OF APPEALS FOR THE SIXTH CIRCUIT ICE HOUSE AMERICA, LLC, Appellant, | ...

United States v. Quavon Peterson (6th Cir. 2014)
Date Filed: May 13th, 2014 Status: Non-Precedential Docket Number: 13-3542
NOT RECOMMENDED FOR FULL-TEXT PUBLICATION File Name: 14a0358n.06 No 13-3542 FILED UNITED STATES COURT OF APPEALS May 13, 2014 FOR THE SIXTH CIRCUIT DEBORAH S. HUNT, Clerk UNITED STAT...

Xiu Ling Chen v. Eric H. Holder, Jr. (8th Cir. 2014)
Date Filed: May 13th, 2014 Status: Precedential Docket Number: 13-1103
United States Court of Appeals For the Eighth Circuit No. 13-1103 Xiu Ling Chen Petitioner v. Eric H. Holder, Jr., Attorney General of the United States ...

Gregory Bickley v. Dish Network LLC (6th Cir. 2014)
Date Filed: May 13th, 2014 Status: Precedential Docket Number: 13-5979
RECOMMENDED FOR FULL-TEXT PUBLICATION Pursuant to Sixth Circuit I.O.P. 32.1(b) File Name: 14a0098n.06 UNITED

Free Law Project



•Free Law Project sponsors the development of:

•Juriscraper,

<https://github.com/freelawproject/juriscraper>, a software system for harvesting the court opinions provided on appellate court websites.

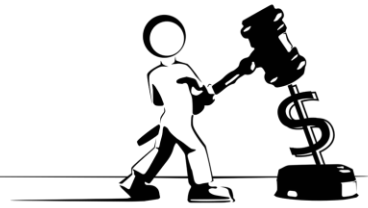
- Covers all federal circuit courts and Supreme Court.

- Now covers all state courts of last resort and most all intermediate appellate state courts.

- Covers various special jurisdictions, such as Court of Federal Claims, Veterans Claims, Tax, Int'l Trade, etc.

- This amounts to watching over 200 different webpages for new documents, every weekday.

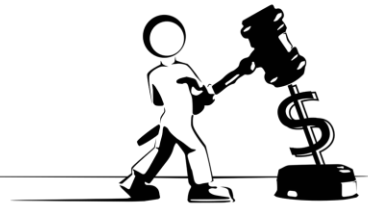
Free Law Project



- Free Law Project partners with Princeton's Center for Information Technology Policy on:
- RECAP, <http://recapthelaw.org>, browser extensions that improve the experience of browsing the federal electronic document system, PACER, while helping PACER users build a free and open repository of public court records.
- Michael will give a lightning talk about RECAP tomorrow.



Free Law Project



- Free Law Project sponsors the development of:
 - Online legal research tools on the CourtListener platform, such as a citator, table of authorities for each document, sophisticated search filtering options, improvements to relevancy search results, etc.
 - An Application Programming Interface (API) to provide access to all court opinions and key related metadata in our database, to enable others to programmatically utilize our collections.

Free Law Project



□ What did you do with our money?!

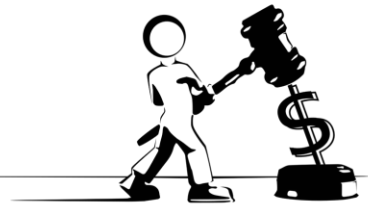
Select Jurisdictions Check All Clear All Check Current Tab Clear Current Tab ×

Select Jurisdictions by Typing

FEDERAL APPELLATE FEDERAL DISTRICT BANKRUPTCY **STATE** SPECIAL

<input checked="" type="checkbox"/> Supreme Court of Alabama	<input checked="" type="checkbox"/> Supreme Court of Iowa	<input checked="" type="checkbox"/> North Dakota Supreme Court
<input checked="" type="checkbox"/> Alabama Court of Appeals	<input checked="" type="checkbox"/> Court of Appeals of Iowa	<input checked="" type="checkbox"/> North Dakota Court of Appeals
<input checked="" type="checkbox"/> Court of Criminal Appeals of Alabama	<input checked="" type="checkbox"/> Supreme Court of Kansas	<input checked="" type="checkbox"/> Ohio Supreme Court
<input checked="" type="checkbox"/> Court of Civil Appeals of Alabama	<input checked="" type="checkbox"/> Court of Appeals of Kansas	<input checked="" type="checkbox"/> Ohio Court of Appeals
<input checked="" type="checkbox"/> Alaska Supreme Court	<input checked="" type="checkbox"/> Kentucky Supreme Court	<input checked="" type="checkbox"/> Ohio Court of Claims
<input checked="" type="checkbox"/> Court of Appeals of Alaska	<input checked="" type="checkbox"/> Court of Appeals of Kentucky	<input checked="" type="checkbox"/> Supreme Court of Oklahoma
<input checked="" type="checkbox"/> Arizona Supreme Court	<input checked="" type="checkbox"/> Supreme Court of Louisiana	<input checked="" type="checkbox"/> Court of Civil Appeals of Oklahoma
<input checked="" type="checkbox"/> Court of Appeals of Arizona	<input checked="" type="checkbox"/> Louisiana Court of Appeal	<input checked="" type="checkbox"/> Court of Criminal Appeals of Oklahoma
<input checked="" type="checkbox"/> Arizona Tax Court	<input checked="" type="checkbox"/> Supreme Judicial Court of Maine	<input checked="" type="checkbox"/> Oklahoma Judicial Ethics Advisory Panel
<input checked="" type="checkbox"/> Supreme Court of Arkansas	<input checked="" type="checkbox"/> Court of Appeals of Maryland	<input checked="" type="checkbox"/> Court on the Judiciary of Oklahoma
<input checked="" type="checkbox"/> Court of Appeals of Arkansas	<input checked="" type="checkbox"/> Court of Special Appeals of Maryland	<input checked="" type="checkbox"/> Oklahoma Attorney General Reports
<input checked="" type="checkbox"/> California Supreme Court	<input checked="" type="checkbox"/> Massachusetts Supreme Judicial Court	<input checked="" type="checkbox"/> Oregon Supreme Court
<input checked="" type="checkbox"/> California Court of Appeal	<input checked="" type="checkbox"/> Massachusetts Appeals Court	<input checked="" type="checkbox"/> Court of Appeals of Oregon
<input checked="" type="checkbox"/> Supreme Court of Colorado	<input checked="" type="checkbox"/> Michigan Supreme Court	<input checked="" type="checkbox"/> Supreme Court of Pennsylvania
<input checked="" type="checkbox"/> Colorado Court of Appeals	<input checked="" type="checkbox"/> Michigan Court of Appeals	<input checked="" type="checkbox"/> Superior Court of Pennsylvania
<input checked="" type="checkbox"/> Supreme Court of Connecticut	<input checked="" type="checkbox"/> Supreme Court of Minnesota	<input checked="" type="checkbox"/> Commonwealth Court of Pennsylvania
<input checked="" type="checkbox"/> Connecticut Appellate Court	<input checked="" type="checkbox"/> Court of Appeals of Minnesota	<input checked="" type="checkbox"/> Judicial Discipline of Pennsylvania
<input checked="" type="checkbox"/> Connecticut Superior Court	<input checked="" type="checkbox"/> Mississippi Supreme Court	<input checked="" type="checkbox"/> Supreme Court of Rhode Island
<input checked="" type="checkbox"/> Supreme Court of Delaware	<input checked="" type="checkbox"/> Court of Appeals of Mississippi	<input checked="" type="checkbox"/> Supreme Court of South Carolina
<input checked="" type="checkbox"/> Court of Chancery of Delaware	<input checked="" type="checkbox"/> Supreme Court of Missouri	<input checked="" type="checkbox"/> Court of Appeals of South Carolina
<input checked="" type="checkbox"/> Superior Court of Delaware	<input checked="" type="checkbox"/> Missouri Court of Appeals	<input checked="" type="checkbox"/> South Dakota Supreme Court
<input checked="" type="checkbox"/> Delaware Court of Common Pleas	<input checked="" type="checkbox"/> Montana Supreme Court	<input checked="" type="checkbox"/> Tennessee Supreme Court
<input checked="" type="checkbox"/> Delaware Family Court	<input checked="" type="checkbox"/> Nebraska Supreme Court	<input checked="" type="checkbox"/> Court of Appeals of Tennessee
<input checked="" type="checkbox"/> Court on the Judiciary of Delaware	<input checked="" type="checkbox"/> Nebraska Court of Appeals	<input checked="" type="checkbox"/> Court of Criminal Appeals of Tennessee
<input checked="" type="checkbox"/> District of Columbia Court of Appeals	<input checked="" type="checkbox"/> Nevada Supreme Court	<input checked="" type="checkbox"/> Texas Supreme Court
<input checked="" type="checkbox"/> Supreme Court of Florida	<input checked="" type="checkbox"/> Supreme Court of New Hampshire	<input checked="" type="checkbox"/> Court of Appeals of Texas
<input checked="" type="checkbox"/> District Court of Appeal of Florida	<input checked="" type="checkbox"/> Supreme Court of New Jersey	<input checked="" type="checkbox"/> Court of Criminal Appeals of Texas
<input checked="" type="checkbox"/> Supreme Court of Georgia	<input checked="" type="checkbox"/> New Jersey Superior Court	<input checked="" type="checkbox"/> Texas Special Court of Review
<input checked="" type="checkbox"/> Court of Appeals of Georgia	<input checked="" type="checkbox"/> New Jersey Tax Court	<input checked="" type="checkbox"/> Utah Supreme Court
<input checked="" type="checkbox"/> Hawaii Supreme Court	<input checked="" type="checkbox"/> New Mexico Supreme Court	<input checked="" type="checkbox"/> Court of Appeals of Utah
<input checked="" type="checkbox"/> Hawaii Intermediate Court of Appeals	<input checked="" type="checkbox"/> New Mexico Court of Appeals	<input checked="" type="checkbox"/> Supreme Court of Vermont
<input checked="" type="checkbox"/> Idaho Supreme Court	<input checked="" type="checkbox"/> New York Court of Appeals	<input checked="" type="checkbox"/> Supreme Court of Virginia
<input checked="" type="checkbox"/> Idaho Court of Appeals	<input checked="" type="checkbox"/> Appellate Division of the Supreme Court of New York	<input checked="" type="checkbox"/> Court of Appeals of Virginia
<input checked="" type="checkbox"/> Illinois Supreme Court	<input checked="" type="checkbox"/> New York Family Court	<input checked="" type="checkbox"/> Washington Supreme Court
<input checked="" type="checkbox"/> Appellate Court of Illinois	<input checked="" type="checkbox"/> N.Y. Surrogate's Court	<input checked="" type="checkbox"/> Court of Appeals of Washington
<input checked="" type="checkbox"/> Indiana Supreme Court	<input checked="" type="checkbox"/> Supreme Court of North Carolina	<input checked="" type="checkbox"/> West Virginia Supreme Court
<input checked="" type="checkbox"/> Indiana Court of Appeals		<input checked="" type="checkbox"/> Wisconsin Supreme Court

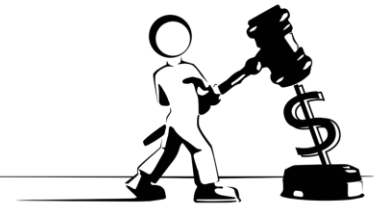
Free Law Project



▣ Scrapers

- ▣ We want the metadata about the documents we collect to be accurate.
- ▣ A one-size-fits-all solution cannot guarantee this across hundreds of different court websites.
- ▣ We do have an extensible general core.
- ▣ The details for each site are handled independently, and this creates a very low barrier to entry for new developer-contributors.

Free Law Project



```
56 lines (45 sloc) | 1.717 kB
Raw Blame History
1 """Scraper for Supreme Court of Maine
2 CourtID: me
3 Court Short Name: Me.
4 Author: Brian W. Carver
5 Date created: June 20, 2014
6
7 History:
8 2014-06-25 (est): Added code for additional date formats.
9 2014-07-02: Was receiving InsanityException and tweaked date code to get some
10 missing dates.
11 2014-12-15: Fixes insanity exception by tweaking the XPath.
12 """
13
14 from datetime import datetime
15
16 from lxml import html
17 from juriscraper.OpinionSite import OpinionSite
18
19
20 class Site(OpinionSite):
21     def __init__(self):
22         super(Site, self).__init__()
23         self.court_id = self.__module__
24         self.url = 'http://www.courts.maine.gov/opinions_orders/supreme/publishedopinions.shtml'
25
26     def _get_download_urls(self):
27         path = '//table//tr/td[2]/a[1]/@href'
28         return list(self.html.xpath(path))
29
30     def _get_case_names(self):
31         case_names = []
32         for e in self.html.xpath('//table//tr/td[2]/a[1]'):
33             s = html.tostring(e, method='text', encoding='unicode')
34             case_names.append(s)
35         return case_names
36
37     def _get_case_dates(self):
38         path = '//table//tr/td[3]/text()'
39         date_styles = ['%B %d, %Y', '%B %d,%Y']
40         dates = []
41         for s in self.html.xpath(path):
42             for date_style in date_styles:
43                 try:
44                     d = datetime.strptime(s.strip(), date_style).date()
45                 except ValueError:
46                     continue
47             dates.append(d)
48         return dates
```

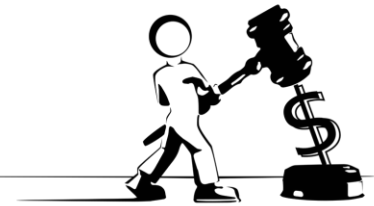
Free Law Project



▣ Scrapers

- ▣ When we started this grant, we had 87 state court scrapers.
- ▣ When we finished work under this grant, we had 154 state court scrapers.
- ▣ Since completing this work we have added over 58,000 state court opinions to our archive.

Free Law Project



□ What did you do with our money?!

COURT LISTENER About FAQ Coverage Profile

Search Query Help

Opinions **10,308 Results** Newest First

Oral Arguments

Refine Your Query

Select Jurisdictions All Jurisdictions Selected

Case Name

Judge

Argued After Argued Before

Docket Number

Search

Create an Alert

United States v. Gamache (1st Cir. 2015)
Date Argued: May 6th, 2015 Docket Number: Case:14-1546 Duration: 10:18

United States v. Miguel Osuna-Alvarez (9th Cir. 2015)
Date Argued: May 6th, 2015 Docket Number: 13-50636 Duration: 31:42
Judge: Noonan, Wardlaw, Murguia

Patricia Stockman-Sann v. Robert McKnight, Jr. (9th Cir. 2015)
Date Argued: May 6th, 2015 Docket Number: 13-55892 Duration: 42:05
Judge: Noonan, Wardlaw, Murguia

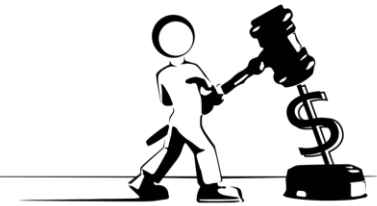
United States v. Cruz Torres-Gonzalez (9th Cir. 2015)
Date Argued: May 6th, 2015 Docket Number: 14-50017 Duration: 39:30
Judge: Noonan, Wardlaw, Murguia

Alaska Airlines v. Judy Schurke (9th Cir. 2015)
Date Argued: May 6th, 2015 Docket Number: 13-35574 Duration: 36:41
Judge: Wallace, Kleinfeld, Christen

Nadia Naffe v. John Frey (9th Cir. 2015)
Date Argued: May 6th, 2015 Docket Number: 13-55666 Duration: 30:11
Judge: Tashima, Tallman, Nguyen

Zagfly, Inc. v. CIR (9th Cir. 2015)
Date Argued: May 6th, 2015 Docket Number: 13-71566 Duration: 19:00
Judge: Fisher, Bea, Friedland

Free Law Project



□ What did you do with our money?!

COURT LISTENER [About](#) [FAQ](#) [Coverage](#) [Profile](#)

[Back to Search Results](#) ☆ **EEOC v. Abercrombie & Fitch Stores, Inc.**

Supreme Court of the United States

Date Argued: February 25th, 2015
Duration: 59:01
Docket Number: 14-86
Fingerprint: de8e5e0d0b15c8051c0f05a93d9235c9a3c3a860

Admin

[Docket](#) [Audio File](#)

Download

[Original](#) | [Our Version](#)

Share this Oral Argument

[Email](#) [Facebook](#) [Twitter](#)

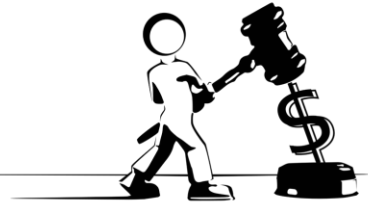
Sponsored Links

CourtListener is a project of [Free Law Project](#), a federally-recognized 501(c)(3) non-profit. We rely heavily on donations for our financial security. To turn off ads, we ask for a \$10 donation each year.

[Donate Now](#)

An audio player interface showing a play button, a progress bar at 00:00, and a volume icon. Below the player is the text "EEOC v. Abercrombie & Fitch Stores, Inc."

Free Law Project



□ Oral Arguments

- We preserve the original audio file.
- We create an improved .mp3 with metadata, album covers, etc.
- These are also accessible via our API and bulk downloads.

Free Law Project



□ Oral Arguments

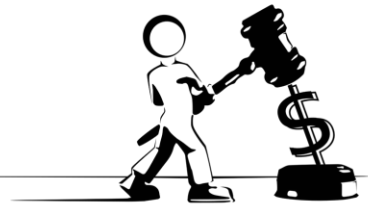
- When we started this grant, we had 0 oral argument audio files.
- Today we host more than 10,000 oral argument audio files. (*257 days* of listening.)
- By adding a new document/entity type we began the process of generalizing our infrastructure to accommodate new document types in the future.
- DC Legal Hackers named it a Legal Hack of the Year for 2014.

Free Law Project



- What did we do under this grant?
 - Complete our daily coverage of all 50 states, even the state courts with terrible websites, by building lots of web scrapers.
 - Add oral argument audio scrapers for the federal appellate courts.
 - Add an open source audio player to our site.
 - Design our system to recognize a new type of document/media.

Free Law Project



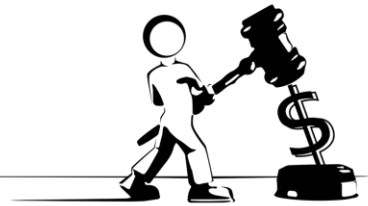
□ Pain Points

- PDFs are horrible.
- Would like to automate a no-cost audio-to-text transcription solution. Does it exist?
- Ongoing maintenance and monitoring.

□ Future (Related) Work

- Increasingly, courts are providing oral argument *video*.

Free Law Project



- Thinking beyond the law:
 - Juriscraper has taught us a lot about how to harvest material from lots of websites.
 - One could change some variable names and use the same library to harvest anything.
 - Harvesting audio taught us a lot about audio processing and audio metadata that is useful to anyone archiving audio files.

Free Law Project



- Get in touch!
 - <http://freelawproject.org>
 - <https://www.courtlistener.com>
 - <http://recapthelaw.org>
 - @freelawproject
 - @courtlistener
 - @recapthelaw
 - @brianwc
 - @mlissner