LECTURE16: GOOD PRACTICES OF SITE DESIGN

CS418/518: WEB PROGRAMMING

BY DR. JIAN WU

COURTESY: DR. JUSTIN BRUNELLE

LAYOUT

- Consistent header, logo, navigation https://www.semanticscholar.org/
- Clear logo/title/site name
- Consistent footer with contact info
- Header and navigation =~ 0.10-0.25 of viewport real-estate (e.g., https://time.com/)
- Sitemap, help, … (see https://time.com/robots.txt)

site name

contact

header

logo

navigation

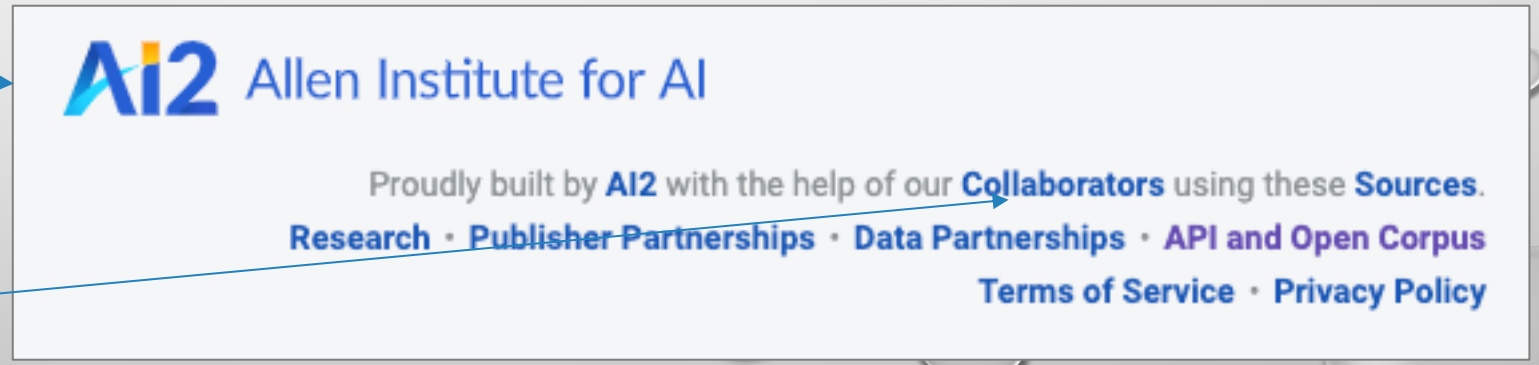
navigation

title

logo

footer

navigation

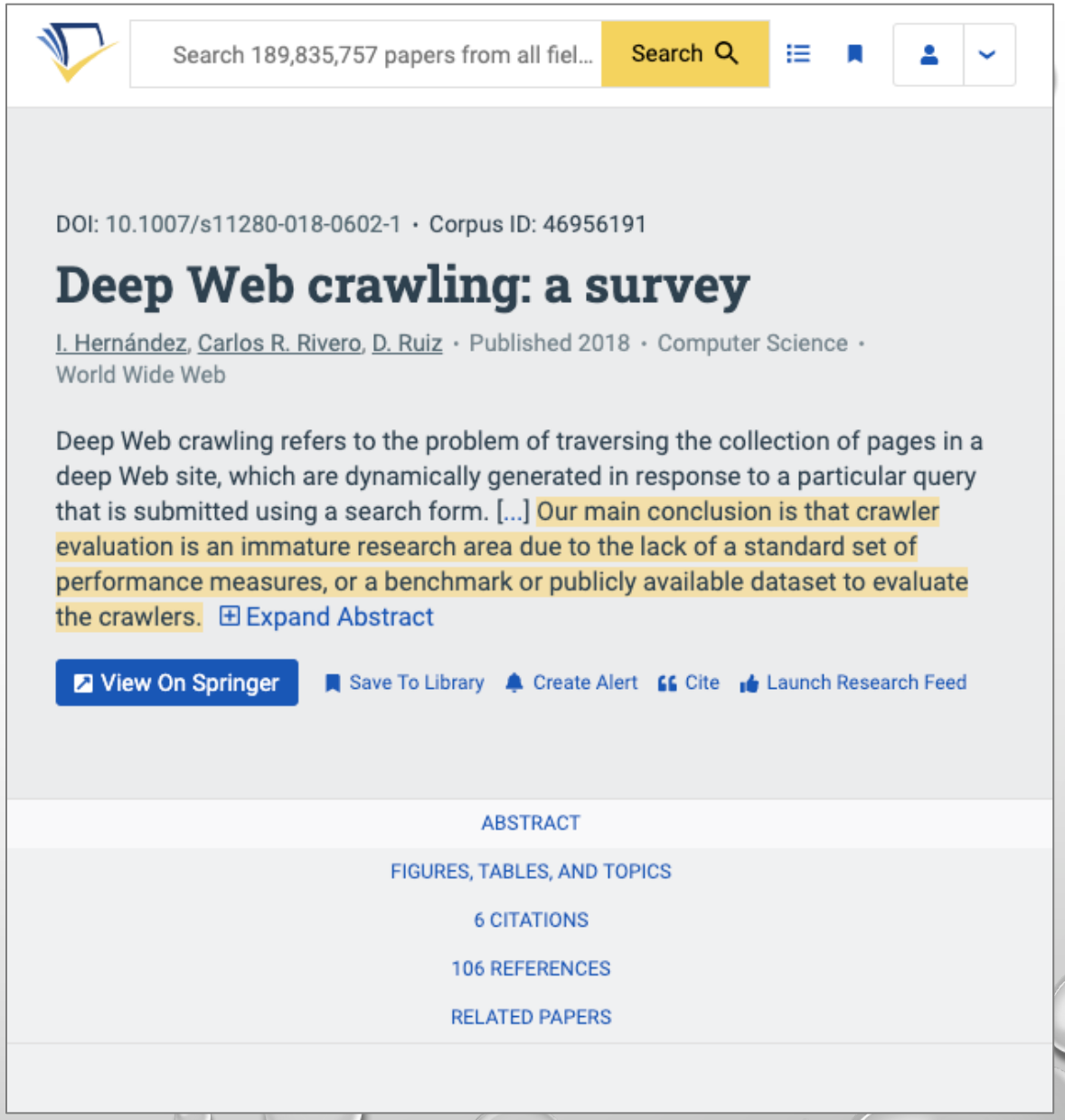


COMPATIBILITY

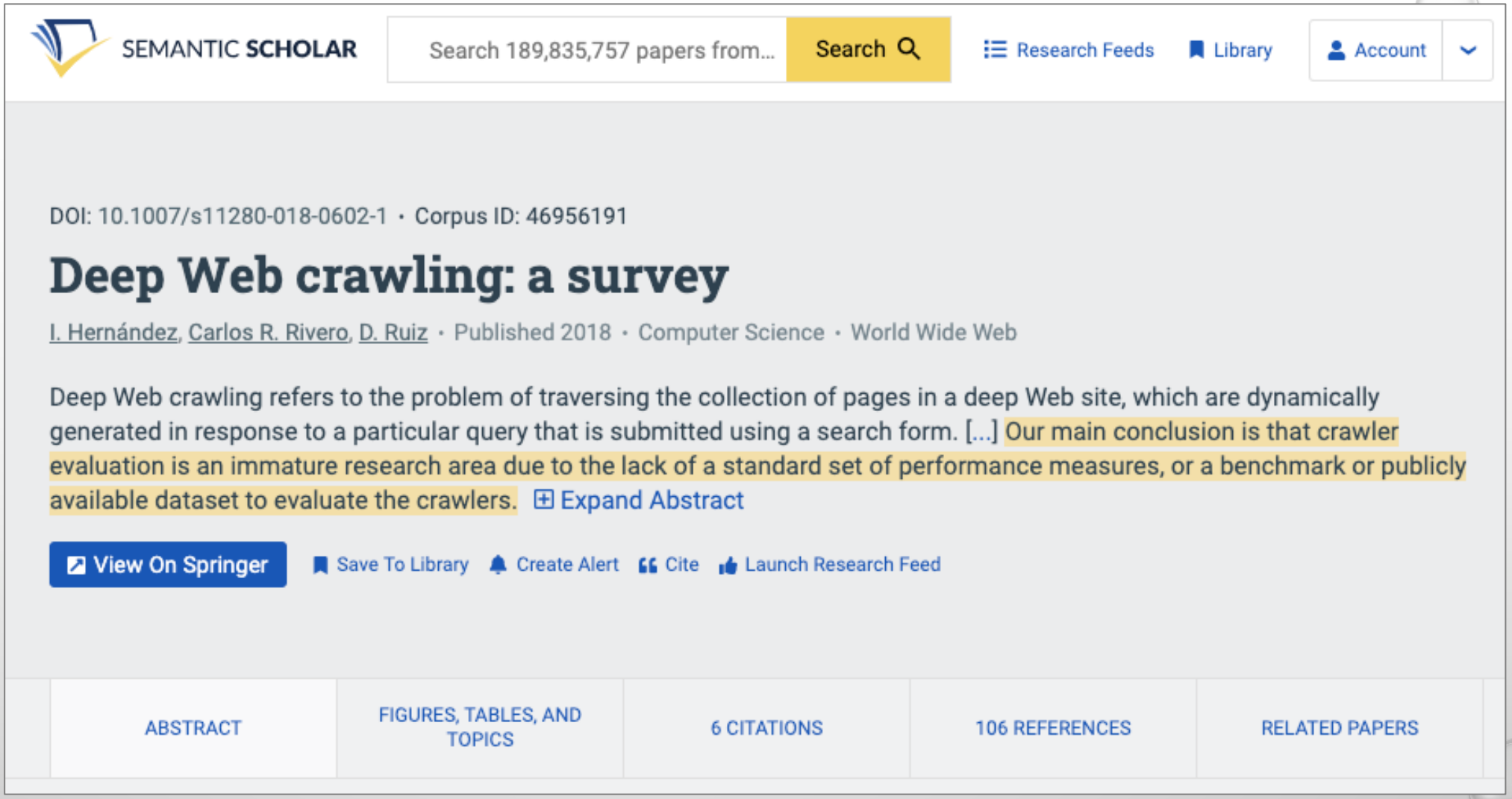
- Visible with vary screen sizes https://www.codeproject.com/Articles/762201/Adjusting-your-websites-to-fit-all-types-of-resolu
- All major browsers: Mozilla, Safari, Edge, etc.
  - HTML Responsive Web Design
  - CSS Responsive Web Design
  - Tutorial on Creating Cross-Browser Compatible HTML and CSS by Harshit Paul
  - How to make a cross browser compatible website? by Deeksha Agarwal
- Captions/alt text
- Structured language for machines, natural language for humans
  - Error codes and messages, paired!


Search 189,835,757 pape...
DOI: 10.1007/s11280-018-0602-1 • Corpus ID: 46956191
Deep Web crawling: a survey
I. Hernández, Carlos R. Rivero, D. Ruiz • Published 2018 •
Computer Science • World Wide Web
Deep Web crawling refers to the problem of traversing the collection of pages in a deep Web site, which are dynamically generated in response to a particular query that is submitted using a search form. [...] Our main conclusion is that crawler evaluation is an immature research area due to the lack of a standard set of performance measures, or a benchmark or publicly available dataset to evaluate the crawlers. Expand Abstract
View On Springer
ABSTRACT
FIGURES, TABLES, AND TOPICS
6 CITATIONS
106 REFERENCES
RELATED PAPERS

wider


Search 189,835,757 papers from all fiel...
Search
DOI: 10.1007/s11280-018-0602-1 • Corpus ID: 46956191
Deep Web crawling: a survey
I. Hernández, Carlos R. Rivero, D. Ruiz • Published 2018 • Computer Science •
World Wide Web
Deep Web crawling refers to the problem of traversing the collection of pages in a deep Web site, which are dynamically generated in response to a particular query that is submitted using a search form. [...] Our main conclusion is that crawler evaluation is an immature research area due to the lack of a standard set of performance measures, or a benchmark or publicly available dataset to evaluate the crawlers. Expand Abstract
View On Springer
Save To Library
Create Alert
Cite
Launch Research Feed
ABSTRACT
FIGURES, TABLES, AND TOPICS
6 CITATIONS
106 REFERENCES
RELATED PAPERS

even wider


SEMANTIC SCHOLAR
Search 189,835,757 papers from...
Search
Research Feeds
Library
Account
DOI: 10.1007/s11280-018-0602-1 · Corpus ID: 46956191
Deep Web crawling: a survey
I. Hernández, Carlos R. Rivero, D. Ruiz · Published 2018 · Computer Science · World Wide Web
Deep Web crawling refers to the problem of traversing the collection of pages in a deep Web site, which are dynamically generated in response to a particular query that is submitted using a search form. [...] Our main conclusion is that crawler evaluation is an immature research area due to the lack of a standard set of performance measures, or a benchmark or publicly available dataset to evaluate the crawlers.
Expand Abstract
View On Springer
Save To Library
Create Alert
Cite
Launch Research Feed
ABSTRACT
FIGURES, TABLES, AND TOPICS
6 CITATIONS
106 REFERENCES
RELATED PAPERS

Share This Paper

DOI: 10.1007/s11280-018-0602-1 · Corpus ID: 46956191

Deep Web crawling: a survey

I. Hernández, Carlos R. Rivero, D. Ruiz · Published 2018 · Computer Science · World Wide Web

Deep Web crawling refers to the problem of traversing the collection of pages in a deep Web site, which are dynamically generated in response to a particular query that is submitted using a search form. [...] Our main conclusion is that crawler evaluation is an immature research area due to the lack of a standard set of performance measures, or a benchmark or publicly available dataset to evaluate the crawlers. Expand Abstract

View On Springer Save To Library Create Alert Cite Launch Research Feed

Top 3 of 6 Citations

View All

Advanced Web Crawler For Deep Web Interface Using Binary Vector & Page Rank

Vishal. V. Mahale, Mahesh T. Dhande, Amruta V. Pandit · 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2018 2nd International Conference on · 2018

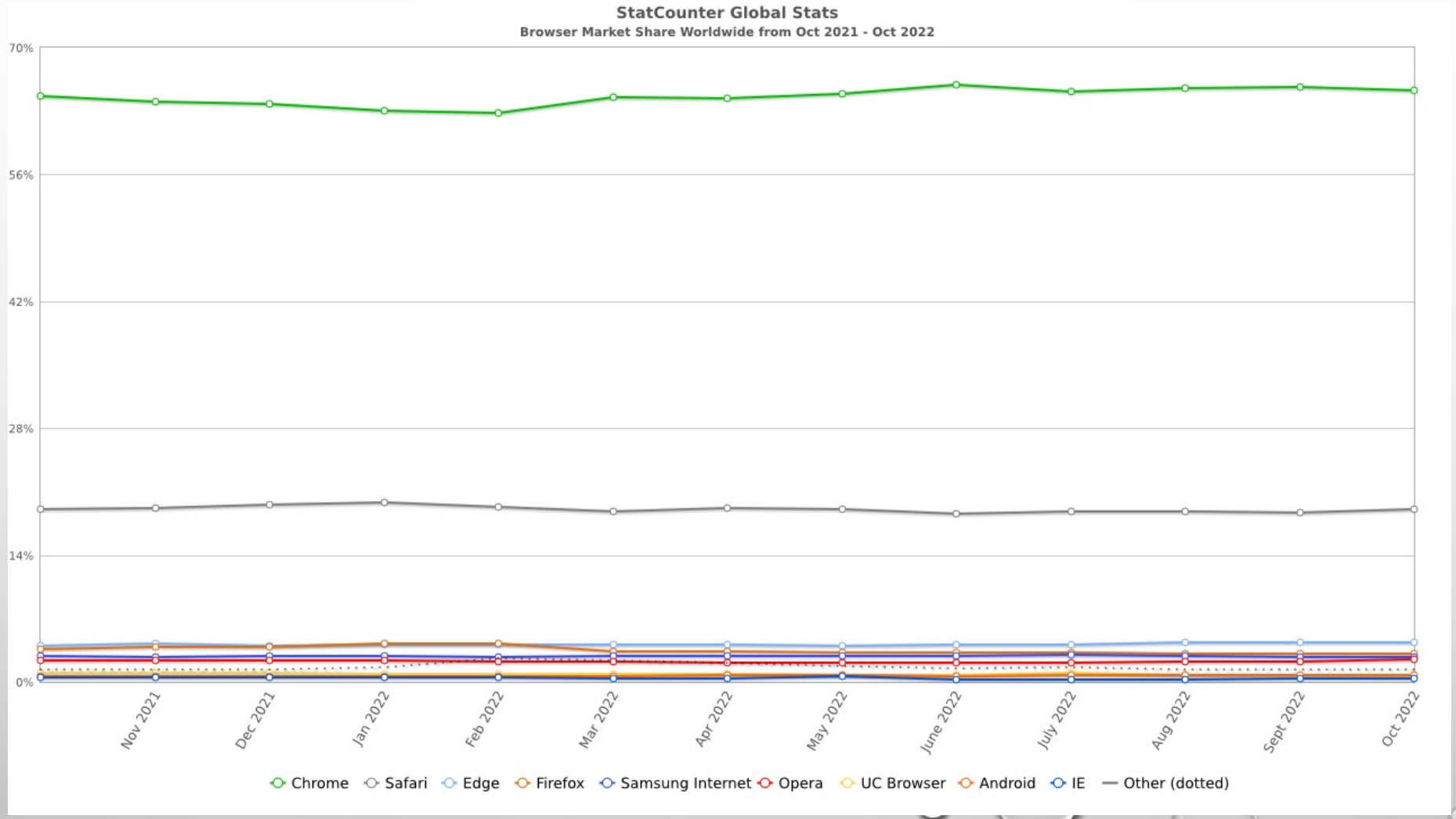
1

Method of Deep Web Collection for Mobile Application Store Based on Category Keyword...

Guosheng Xu, Zhimin Wu, ... L. Wang · SpaCCS · 2019

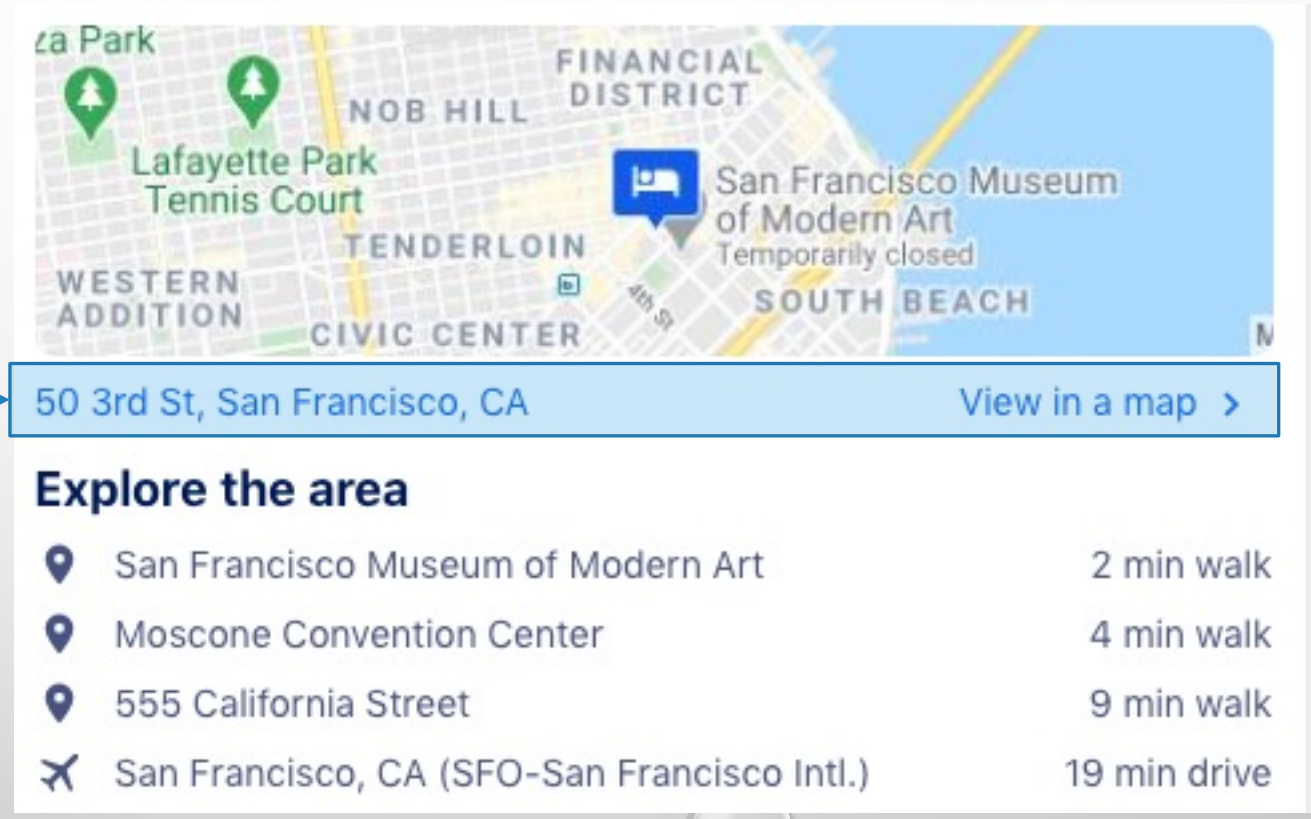
widest

MAINSTREAM WEB BROWSERS



EXAMPLES CAPTIONS

figure caption



Expedia.com

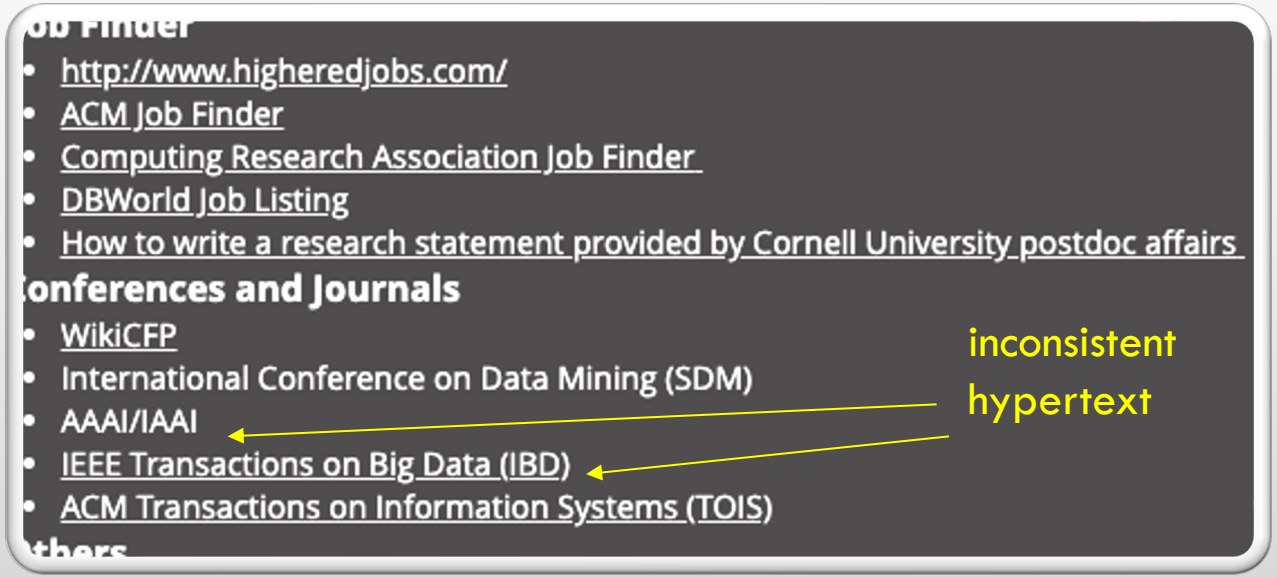
PERFORMANCES

- Ideal website load time for mobile sites is <1 second.
- 53% of mobile site visits are abandoned if pages take longer than 3 seconds to load.
- All graphics serve a purpose
- Images not significantly scaled: a negative example
- Indicate progress
  - Under construction templates: https://colorlib.com/wp/free-under-construction-templates/#html-templates

How Fast Should A Website Load & How To Speed It Up

APPEARANCE

- Consistent/purposeful fonts
- Clearly indicate hyperlinks
- Sort long lists: example citations of papers (next page)
- Highlight action areas/errors: modify URL and see what happens (next next page)



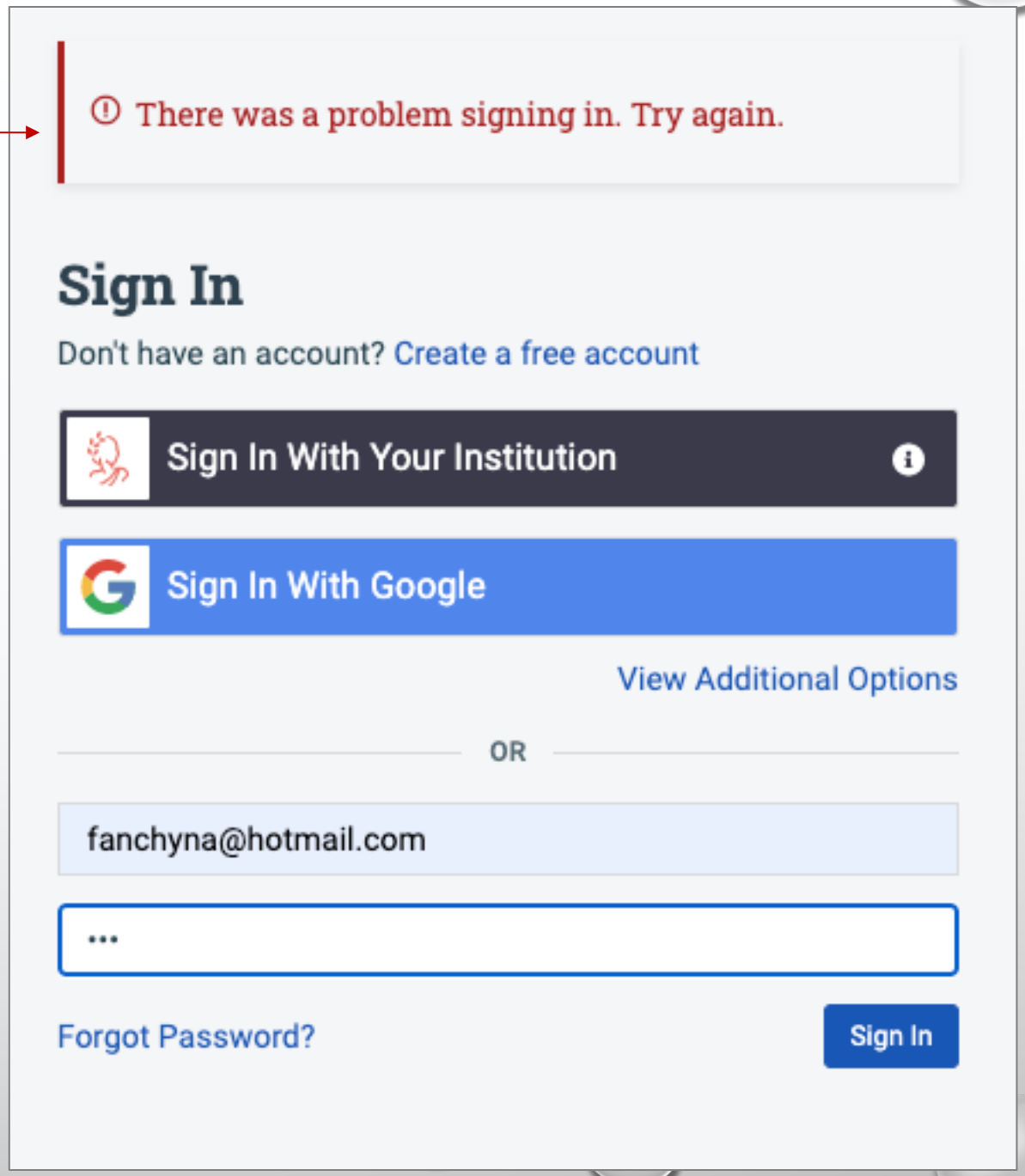
Sort long list

Summary | Citations | Active Bibliography | Co-citation | Clustered Documents | Version History

Citations

4671 The anatomy of a large-scale hypertextual web search engine. Computer Networks - Brin, Page - 1998 (Show Context)
373 The Web as a graph: Measurements, Models, and Methods - Kleinberg, Kumar, et al. - 1999 (Show Context)
203 A technique for measuring the relative size and overlap of public web search engines - Bharat, Broder - 1998 (Show Context)
133 and H.Garcia-Molina, "Parallel crawlers". - Cho - 2002 (Show Context)
128 Introduction to Informetrics: Quantitative methods in library, documentation, and information science. - Egghe, Rousseau - 1990 (Show Cont
128 Challenges in web search engines. - Henzinger, Motwani, et al. - 2002 (Show Context)
96 An adaptive model for optimizing performance of an incremental web crawler. - Edwards, McCurley, et al. - 2001 (Show Context)
93 Engineering a multipurpose test collection for web retrieval experiments. - Bailey, Craswell, et al. - 2003
84 Hyperlink analysis for the web. - Henzinger - 2001
77 The RBSE Spider - balancing effective search against Web load - Eichmann - 1994 (Show Context)
69 Crawling towards eternity: Building an archive of the world wide web. - Burner - 1997 (Show Context)
67 Bibliometrics and beyond: some thoughts on web-based citation analysis. - Cronin - 2001 (Show Context)
65 A comparison of techniques to find mirrored hosts on the WWW - Bharat, Broder, et al. - 1999
62 The connectivity sonar: Detecting site functionality by structural patterns. - Amitay - 2003
52 Performance Web Crawling - Najork, Heydon, et al. - 2001 (Show Context)
45 A web crawler design for data mining, - Thelwall - 2001 (Show Context)
41 Data Collection Methods on the Web for Informetric Purposes – A Review and Analysis', - Bar-Ilan - 2001
30 Focused crawls, tunneling, and digital libraries. - Bergmark, Lagoze, et al. - 2002 (Show Context)
29 Internet search engines - fluctuations in document accessibility. - Mettrop, Nieuwenhuysen - 2001
25 Secondary research: Information sources and methods. - Stewart, Kamins - 1993 (Show Context)
10 Informetric Theories and Methods for Exploring the Internet: An Analytical Survey of Recent Research Literature', - Bar-Ilan, Peritz - 2002
3 Google's Web page ranking applied to different topological Web graph structures - Meghabghab - 2002 (Show Context)
2 Scholarly communication and bibliometrics. In B. Cronin (Ed.) Annual Review of Information Science and Technology - Borgman, Furner - 20
1 Harvest-NG homepage. Retrieved June 18, 2003, from http://webharvest.sourceforge.net/ng - Harvest-NG - 2003 (Show Context)
1 Guidelines for robot writers. Retrieved May 12, 2003, from http://www.robotstxt.org/wc/guidelines.html - Koster - 1993 (Show Context)
1 Research Report 173, Compaq: Systems Research - Center, California - 2001
1 Growth and structure of the World Wide Web: Towards realistic modelling - Tadić - 2002
1 Methodologies for crawler based surveys - Thelwall - 2002

Highlight errors

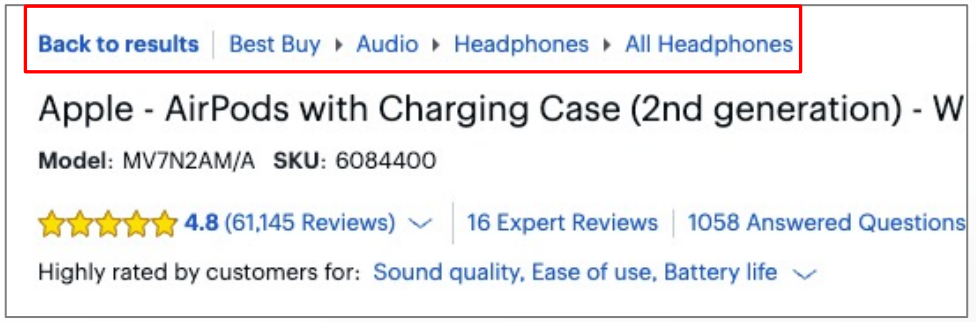


COLORS

- Consider black & white printing
  - https://ColorBrewer2.org
- Consistent color schemes
  - Same functions have same colors
  - Bootstrap may help?
- Bootstrap is a free and open-source CSS framework directed at responsive, mobile-first front-end web development.

Bootstrap contains CSS- and (optionally) JavaScript-based design templates
for typography, forms, buttons, navigation, and other interface components.
See examples.
See search engine templates.
More search engine examples.

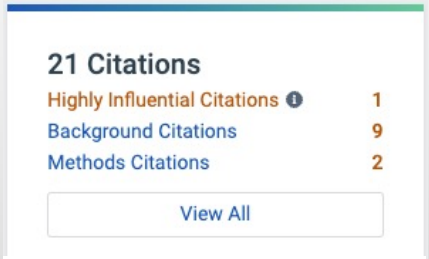
NAVIGATION



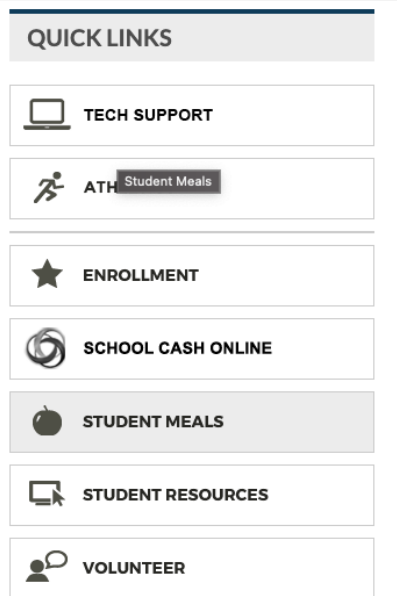
breadcrumbs on Bestbuy.com

- Use breadcrumbs/pagination
  - For large websites and websites that have hierarchically arranged pages, e.g., e-commerce websites, in which a large variety of products is grouped into logical categories.
  - You **shouldn't** use breadcrumbs for single-level websites that have no logical hierarchy or grouping.
- Use fragment IDs
- Minimize number of clicks to reach a page
  - e.g., creating shortcut (Old Donation School website with quick links)
- Navigation menus can be grouped
  - The "View All" Menu
  - Also see the next page
- Cool URLs: use descriptive words, instead of just numbers

  https://time.com/5894969/the-recession-isnt-over/

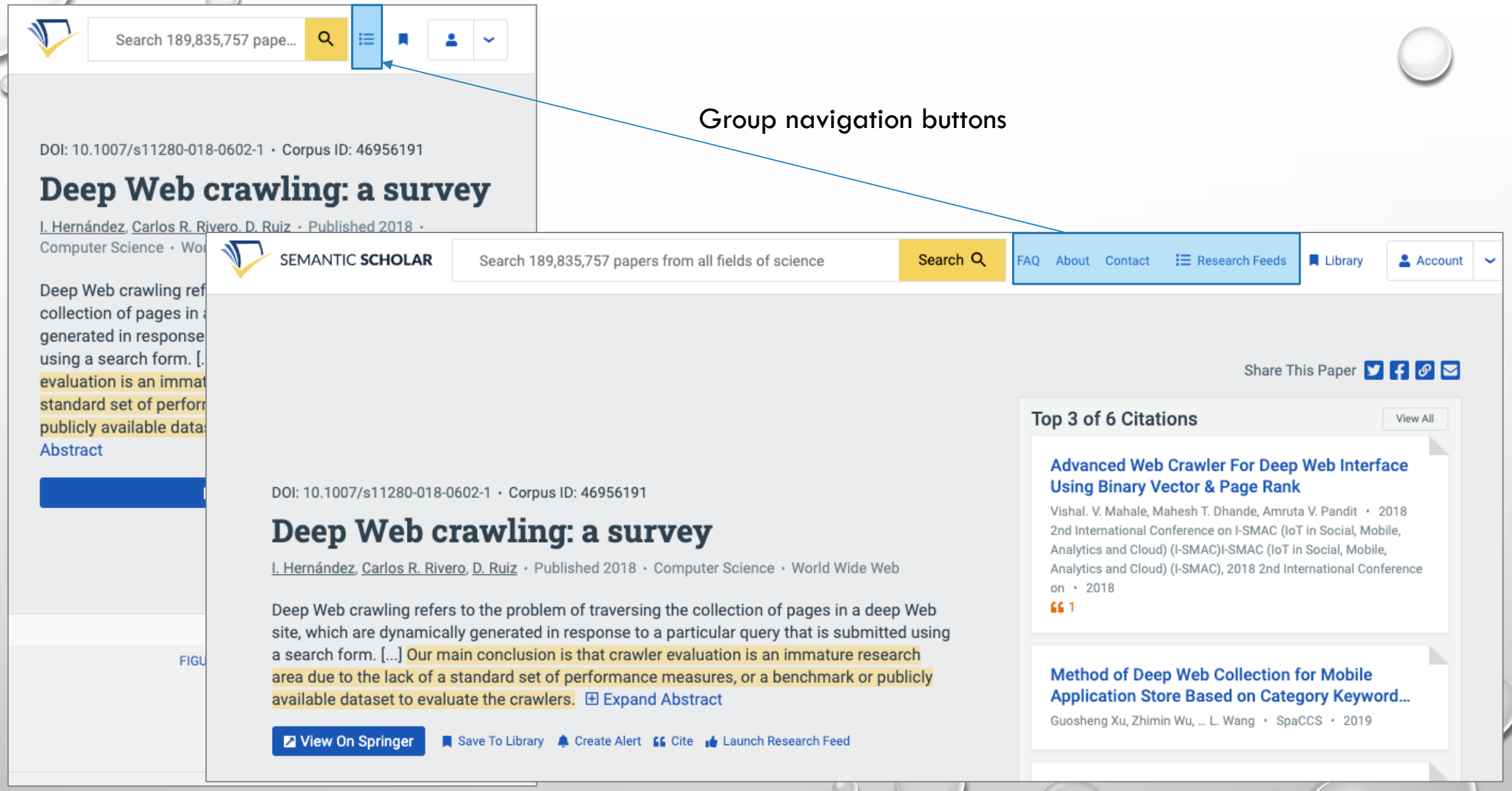


The "View All" menu on semantic scholar website



Quick links on the Old Donation School website

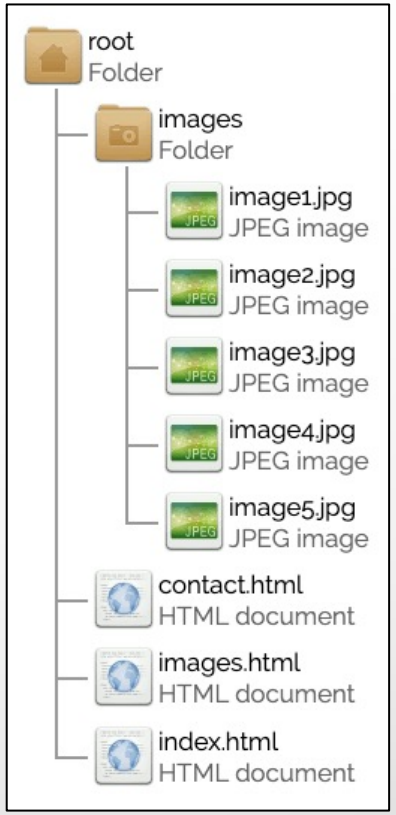
Hansel and Gretel

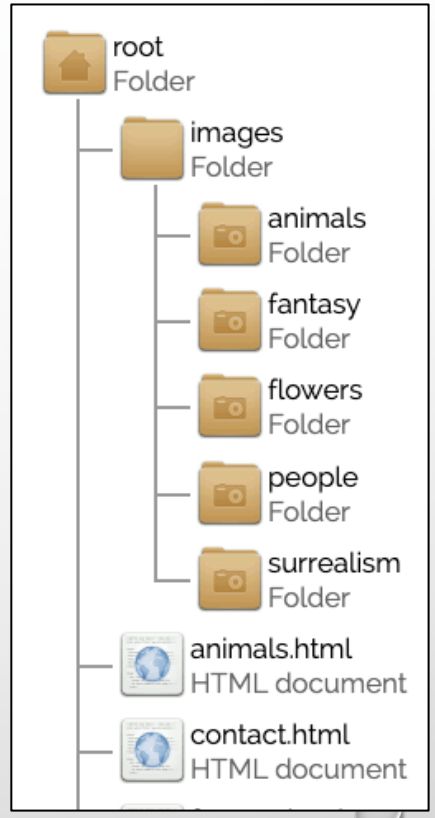

Search 189,835,757 pape...
DOI: 10.1007/s11280-018-0602-1 · Corpus ID: 46956191
Deep Web crawling: a survey
I. Hernández, Carlos R. Rivero, D. Ruiz · Published 2018 ·
Computer Science · Wo
Deep Web crawling ref
collection of pages in
generated in response
using a search form. [.
evaluation is an immat
standard set of perfor
publicly available data
Abstract
FIGU
Group navigation buttons
SEMANTIC SCHOLAR
Search 189,835,757 papers from all fields of science
Search
FAQ
About
Contact
Research Feeds
Library
Account
Share This Paper
DOI: 10.1007/s11280-018-0602-1 · Corpus ID: 46956191
Deep Web crawling: a survey
I. Hernández, Carlos R. Rivero, D. Ruiz · Published 2018 · Computer Science · World Wide Web
Deep Web crawling refers to the problem of traversing the collection of pages in a deep Web site, which are dynamically generated in response to a particular query that is submitted using a search form. [...] Our main conclusion is that crawler evaluation is an immature research area due to the lack of a standard set of performance measures, or a benchmark or publicly available dataset to evaluate the crawlers. Expand Abstract
View On Springer
Save To Library
Create Alert
Cite
Launch Research Feed
Top 3 of 6 Citations
View All
Advanced Web Crawler For Deep Web Interface Using Binary Vector & Page Rank
Vishal. V. Mahale, Mahesh T. Dhande, Amruta V. Pandit · 2018
2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2018 2nd International Conference on · 2018
1
Method of Deep Web Collection for Mobile Application Store Based on Category Keyword...
Guosheng Xu, Zhimin Wu, ... L. Wang · SpaCCS · 2019

FILE ORGANIZATION AT THE BACKEND

Which one do you prefer?

- Directory structure, file names, etc.
  - CiteSeerX code repository
- Logical flow
  - Where would you want it to be?
  - Site diagram
- Sitemap for websites that
  - are large
  - contains isolated links
  - has few external links to it
  - has lots of rich media
  - See the introduction on Google Search Central
  - https://developers.google.com/search/docs/crawling-indexing/sitemaps/overview





CODE

- Use validators:
  - W3C markup validation service
  - Other validators: https://w3c.github.io/developers/tools/
- Use Git!
  - Rollbacks
  - Recovery
  - Issue tracking…
- Comment
  - **Do make comment and documentation!**
  - Just enough, do not abuse it (i.e., things that are confusing)
- Grouping, functions, common naming/camelCase
  - The "Don't repeat yourself" (DRY) principle
- Use MVC to logically group functions, application/logic/infrastructure

What is **Code Validation**? **Code validation** is the process of checking that the **coding** of a web page is in compliance with the standards and recommendations set by the World Wide Web Consortium (W3C) for the web. **Code validation** helps to produce clean **code**.

BACK-END STUFF

- Storage that makes sense
  - Files when necessary: because disk I/O is the slowest
  - SQL when necessary: communication may have a significant overhead
  - Key-value vs. array when appropriate
- Improve markup and metadata for SEO and discoverability
  - https://searchengineland.com/guide/seo/html-code-search-engine-ranking

NO-BRAINERS

- Hyperlinks
- Forms
- JavaScripts
- **ALL SHOULD WORK!**
- NO TYPOS!