Science in the service of awesome:
building on top of scientific computing

Hilary Mason

h@bit.ly   @hmason
http://www.pcworld.com/article/223409/move_over_dr_soong_girls_can_build_android_apps_too.html

http://bit.ly/hOnbWg
First post on my new blog about Python in finance and stats! "Adventures in Data Aggregation (Group By)" bit.ly/posA1n
#scipy
William Springer
Know somebody who isn't a techie, but wants to learn to make websites? You might want to get them a copy of this book...

Learning HTML5 Step by Step « Vulcan Ears Book Reviews
vulcanears.com
I've reviewed several books lately that introduce the new elements in HTML5; each of them assumes that you already have a working knowledge of html and css and simply want to know about the changes. But what if you're brand new to web programming and want a simple html book that assumes you don't kn...

July 7 at 9:05pm · Share

O'Reilly Media likes this.

O'REILLY
O'Reilly Media
Paying attention to our tabs makes sense. The Open Source Month tab on the left leads you to a free video download of your choice in honor of you and OSCON (in that order). And we've just added a new feature, "Win JavaScript Titles." For 10 days only, you can play fun online games to win ebooks and videos. We love it when you like us. http://oreil.ly/qe87kM

Wall Photos

Yesterday at 4:05pm · Share

10 people like this.

O'REILLY
O'Reilly Media
O'Reilly Open Source Convention – OSCON co–chair Edd Dumbill joins us tomorrow, July 13, for a Facebook–only "office hour" visit at 4pm to discuss the conference with you. He's also been thinking a lot about Java. Here's some of his recent wisdom on how it has, you know, changed the world.

Seven Java projects that changed the world – O'Reilly Radar
wicked hard problem

10s of millions of URLs / day
100s of millions of events / day
1000s of millions of data points

bitly Shorten, share, and track your links
Data science?

engineering

math

science

nerds

nerds

nerds

nerds

awesome nerds
Obtain

Scrub

Explore

Model

iNterpret

find . | awk -F. '{print $3}' | sort | uniq -c | sort -n

21 txt
23 h
28 tcl
29 cnf
29 nginx
48 c
48 sql
120 html
121 js
152 conf
341 sh
584 py
Why now?
Algorithms +

On-demand computing +

data data data data
the real continents of the world

http://bit.ly/r1OoHn
tom clancy  van halen
golfing  harley davidson
ghostbusters  phish
the big lebowski
soundgarden  brew  boating
nofx  groundhog day  hockey
jeep  blazing saddles  the red sox
the dropkick murphys  megadeth
grilling  ccr  robert heinlein
boats skiing  zappa  nascar
motorcycles  software  dark tower
the hitchhiker's guide to the galaxy  breaking bad
band of brothers  burn notice  coen brothers
michael crichton  bad religion  tenacious d
mostly rock  i'm a country boy  building things
queens of the stone age  mountain biking
i can fix anything  the offspring  a few beers
apocalypse now  lock, stock, and two smoking barrels

http://bit.ly/mPLOTc
Astrometry Project

Enron: Hub-and-Spoke of Deceit

Secret Projects

Normal Projects

Highest degree: ann9enigma@gmail.com
Highest degree: undisclosed recipients;
Highest degree: undisclosed-recipients;
Highest degree: blog@promote-bzs.net

Highest degree: faith1desmond@att.net

Highest degree: mck54321@gmail.com, zhengshch@gmail.com, johnkerrycassidy@gmail.com,

https://github.com/drewconway/strata_bootcamp
What languages are in a page?
calculations on streams of data

click event → queue

queue reader

redis...

HTTP API

redis
https://bitly.com/l2Nsf7+
"Google+, la red social que competirá con Facebook"
entropy calculation

def ghash2lang(g, Ri, min_count=3, max_entropy=0.2):
    """
    returns the majority vote of a language for a given hash
    """

    lang = R.zrevrange(g,0,0)[0]
    # let's calculate the entropy!
    # possible languages
    x = R.zrange(g,0,-1)
    # distribution over those languages
    p = np.array([R.zscore(g,lang) for langi in x])
    p /= p.sum()
    # info content
    I = [pi*np.log(pi) for pi in p]
    # entropy: smaller the more certain we are! - i.e. the lower our surprise
    H = -sum(I)/len(I) #in nats!
    # note that this will give a perfect zero for a single count in one language
    # or for 5K counts in one language. So we also need the count..

    count = R.zscore(g,lang)
    if count < min_count and H > max_entropy:
        return lang, count
    else:
        return None, 1
http://4sq.com/96kc10
What locations are relevant to a URL?
click distribution data

```json
{
    status_code: 200,
    data: {
        outputs: {
            results: {
                us-tx-austin: 4,
                us-tx-cibolo: 3,
                us-fl-miami: 2,
                us-tx-converse: 8,
                us-tx-san antonio: 33
            }
        }
    },
    params: {
        db: "ghash2Cities",
        limit: 5,
        type: "city",
        ghash: "nrGzQP"
    }
},
status_txt: "OK"
}```
How do people click on URLs over time?
normal click distributions
Hace unos meses, y durante tres años, trabajé en un club infantil dando clases particulares a mismos niños y niñas hasta hace tres meses que me despidieron ¿la razón? Las niñas en adolescencia y se distraen porque soy atractivo. Por mi "cara bonita" he perdido el único trabajo en mi casa. ADV

Vaya asco de vida (8857) – Haberlo pensado antes (107) – Menuda chorrada (278)

Demasiado guapo para trabajar... menuda putada
¿Te gustaría ver tu banner
www.adpv.com

A few months ago, and for three years, I worked in a club giving private lessons, I always had children until three months ago they fired me the reason? Girls are reaching adolescence because I'm sexy. For my "pretty face" I lost the only income coming into my house.

Go disgust of life (8857) – have thought of before (107) – What a bullshit (278)

Comments

# 1 by alexander666
06/06/2011 13:02:53
Too cute to work ... what a bitch

# 2 by lord_dark
06/06/2011 13:03:02
abnormal click distributions
Save Time with Personalized Interval Workouts!

- Burn 30% More Calories than exercising at a set-pace
- Metabolism stays elevated up to 12 hours after each workout
- Lose Weight and Get Fit in three 30-minute workouts per week

Learn more  Sign up

THE NO EXCUSES WORKOUT SYSTEM FEATURES:

The No Excuses Workout System (NEWO) is a subscription-based online personal trainer designed to help you get fit, lose weight and stay healthy.

www.noexcusesworkouts.com  NEWO includes an
STOP - there might be a problem with the requested link

The link you requested has been identified by bitly as being potentially problematic. This could be because a bitly user has reported a problem, a black-list service reported a problem, because the link has been shortened more than once, or because we have detected potentially malicious content. This may be a problem because:

- Some URL-shorteners re-use their links, so bitly can't guarantee the validity of this link.
- Some URL-shorteners allow their links to be edited, so bitly can't tell where this link will lead you.
- Spam and malware is very often propagated by exploiting these loopholes, neither of which bitly allows for.

The link you requested may contain inappropriate content, or even spam or malicious code that could be downloaded to your computer without your consent, or may be a forgery or imitation of another website, designed to trick users into sharing personal or financial information.

bitly suggests that you
- Change the original link, and re-shorten with bitly
- Close your browser window
- Notify the sender of the URL

Or, continue at your own risk to
http://bit.ly/1j2b7k8
clustered URLs
What’s trending?
Distributions in SciPy

This page summarizes how to work with univariate probability distributions using Python’s SciPy... and parameterizations SciPy makes every continuous distribution into a location-scale family, including some...
www.johndcook.com/distributions_scipy.html

"The Reality of Digital Science" Password protect your Zipcasts

blocking points (to name a few ...) idea lit review publish discovery share results materials retest analyze experiment collect dataWednesday, 13 July 2011
www.slideshare.net/kaythaney/the-reality-of-digital-science

SciPy 2011 :: Schedule

Tom Aldcroft Minesh B Amin | MBA Sciences Roberto Anaya | Texas Water Development Board BILL ANDERSON | U Texas at Austin Chad Baker | UT Austin Chris Barker | National Oceanic and Atmospheric Admin Jordan Blocher | UNR Nicholas Bray | Google
cconference.scipy.org/scipy2011/scipy2011attendees.php

pyssp 0.1.2

signal processing library for education require: python2.6 or later scipy File Type Py Version.../ Keywords:
sciPy,speech processing License: CC3.0–BY-SA Package Index Owner: aihara The rating feature has
pypi.python.org/pypi/pyssp/

NetworkX Summit (SciPy 2011)

, Texas. This will be my first SciPy, and while I am very excited for all of the Python hacking-goodness...
duplicated below. I strongly encourage anyone attending SciPy that has either used NetworkX or has...
www.drewconway.com/zia/

Scikit-learn tutorial: statistical-learning for scientific data processing

packages ( numpy, scipy, matplotlib ). Note This document is meant to be used with scikit-learn
gaelvaroquaux.github.com/scikit-learn-tutorial/index.html

Python

Pythonを支援する環境・ツール 3.1 IPython 3.2 setuptools 4 モジュール 4.1 SciPy 4.1.1 Scipy
Superpack 4.2 NumPy...x.dmg python-2.6.x.dmg python-2.5.x.dmg Enthought Python Distribution SciPy,
matplotlib なども同梱したディストリビュー...
macwiki.sourceforge.jp/wiki/index.php/Python
Thank you!

h@bit.ly
@hmason

bitly | Shorten, share, and track your links