

Project Overview

Nikhil Prakash

PANDA A News Data Application

- Open – source platform for making data analysis easy for journalists
- User uploads the information and specifies type of data
 - PANDA will organize it so it is easily searchable
- Allows for journalists who lack technical computer experience to still extract information from the data

Uploading

PANDA
A Newsroom Data Appliance

[Search](#)
[Upload](#)

Find a dataset

- [Politics](#)
- [Education](#)
- [Crime](#)

[Need another category?](#)

Upload

Step 1

Start by selecting a **CSV (.csv)** or **Excel (.xls or .xlsx)** file from your computer:

no file selected

Step 2

Your file is currently being uploaded to PANDA. If you have a very large file this may take a little while.

0% uploaded.

Step 3

Your file has finished uploading. PANDA has automatically created a new dataset for this file and begun to make it searchable. Click Continue to view the progress of the import as well as provide a name and other details for your new dataset:

Example Query

PANDA
A Newsroom Data Appliance
[Search](#)
[Upload](#)
[Find a dataset](#)

- [Politics](#)
- [Education](#)
- [Crime](#)

[Need another category?](#)

Search Texas hospitals

[Search](#) [Advanced](#)

About Texas hospitals (542 rows)

Created by User <[user@pandaproject.net](#)> on January 26, 2012 @ 4:17 pm.
Never modified.

CATEGORIES
[Education](#)

DESCRIPTION
This is a table of Texas hospitals in the state's inpatient discharge database.

RELATED FILES
[dischargesproviders.csv](#) uploaded January 26, 2012 @ 4:17 pm [Download \(28483 bytes\)](#)

SAMPLE DATA
Showing 5 sample rows from 542 total rows in this dataset.

THCIC_ID	hospid	PROVIDER_NAME	count
474000	48474000	Parkland Memorial Hospital	53379
331000	48331000	Baylor University Medical Center	42037
154000	48154000	Methodist Hospital	40732
124000	48124000	Methodist Hospital	38466
347000	48347000	Memorial Hermann Hospital	37649

PANDA
A Newsroom Data Appliance
[Search](#)
[Upload](#)
[Find a dataset](#)

- [Politics](#)
- [Education](#)
- [Crime](#)

[Need another category?](#)

Search Texas hospitals

[Search](#) [Advanced](#) [Back to dataset details](#)

2 matching rows (542 total rows)

THCIC_ID	hospid	PROVIDER_NAME	count
266000	48266000	Sierra Medical Center	12881
865000	48865000	Sierra Providence East Medical Center	7951

Response

- Still in the very early stages
 - Only allows for excel documents
 - Search function did not have too many features
- Idea has a lot of potential
 - Anyone can put up raw data that can be analyzed
 - Allows for news companies to organize their data better by constantly adding to previous postings in a private setting

Document Cloud

- A platform for journalists to share primary source documents
- 3 Major Functions:
 1. Runs uploaded document through OpenCalais – analyzes documents and reports salient points
 2. Allows for further human analysis through annotation tools
 3. Provides the option of sharing the document in the public catalog

Annotation

CURRENTPAGESNOTESTEXT

Zoom

Search

Page 1 of 56

the hotel and a man named [redacted] (phonetic) 'is [redacted] of this individual and it is said that Geddy picked him off the streets. [redacted] said that at the present time the hotel is packed with employees of the Italian Consulate and although he is not interested in them, he feels that we should have information about Geddy and perhaps institute an inquiry as to his activities. I assured him that this information would be brought to the attention of the appropriate individuals.

Aug 1940: The FBI takes notice when J. Paul Getty ("Geddy") buys the run-down Hotel Pierre on 5th Avenue in New York City, fires the staff and replaces them with "employees of the Italian consulate." Months earlier, Italy had joined forces with Nazi Germany. An inquiry is opened.

Mr. Glegg
Mr. Foxworth

U.S. DEPARTMENT OF JUSTICE
RECEIVED
AUG 27 1940

p. 2

AR AI

August 27, 1940

Special Agent in Charge
Los Angeles, California

ALL INFO FOR INFO ITALIAN
157-12-10-10
Date 8-28-40 PM SP/6011/4

Description

The FBI's espionage investigation o Paul Getty.

Hotel Pierre "packed with employ the Italian Consulate"

Getty's possible "inimical activity"

Espionage file opened on Getty

"Married four times..."

"personal friend of Adolph Hitler"

Hoover to AG re: Getty

Getty negotiating oil deal in Berlin

Getty attorneys deny a link

Original Document (PDF) »

Print Notes »

Related Article »

Contributed by: Jason Felch, Chasing Aphrodite

DocumentCloud

Searching Documents

DocumentCloud

SearchHelp

Home / Join / Log In

Better Government Association
68 documents, 0 notes

Borderzine
1 document, 0 notes

The Boston Globe
177 documents, 35 notes

Bozeman Daily Chronicle
202 documents, 59 notes

The Bureau of Investigative Journalism
3 documents, 28 notes

Business Green
5 documents, 0 notes

California Watch
15 documents, 0 notes


Carbon Brief
6 documents, 0 notes


CBC News
475 documents, 41 notes


Center for Constitutional Rights
140 documents, 37 notes


All Documents


OpenSort▼Analyze▼Save▼

**6611 Edna Karr Summer Program Inspection Report**
3 pages – Contributed by Jessica Williams, The Lens – Feb 02, 2012

**Woolsey 3901 Short Form 19870301 D1839Z02[1]**
3 pages – Contributed by Ben Freeman, Project on Government Oversight – Feb 02, 2012

**Minutes from 2011 KTSC board meetings**
4 pages – Contributed by Erin Chapin, Knoxville News Sentinel – Feb 02, 2012
Minutes from 2011 meetings of the Knoxville Tourism and Sports Corporation board.

**Minutes from 2010 KTSC board meetings**
10 pages – Contributed by Erin Chapin, Knoxville News Sentinel – Feb 02, 2012
Minutes from 2010 meetings of the Knoxville Tourism and Sports Corporation board.

**Minutes from 2009 KTSC board meetings**
6 pages – Contributed by Erin Chapin, Knoxville News Sentinel – Feb 02, 2012
Minutes from 2009 meetings of the Knoxville Tourism and Sports Corporation board.

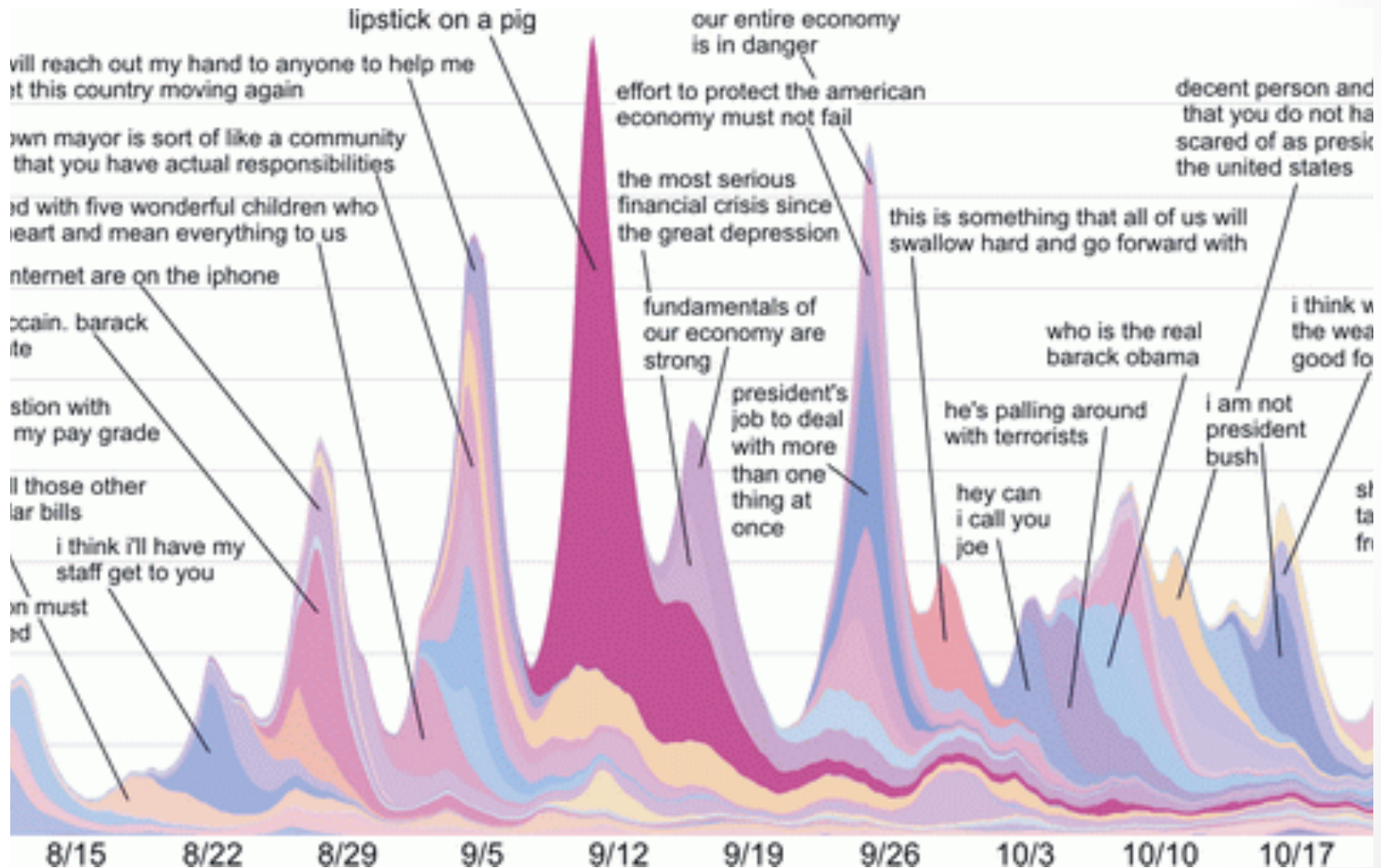
Response

- Already functioning quite well with a large database
- Wide variety of documents
 - Organized by news source contributor
- Annotations allow for greater analysis of texts
- Developing projects to help make contributions easy
 - Ex: Docsplit - “A command-line utility and Ruby library that converts documents into PDFs and breaks them apart into images and text”

Meme Tracker

- “builds maps of the daily news cycle by analyzing around 900,000 news stories and blog posts per day from 1 million online sources, ranging from mass media to personal blogs”
- Tracks most common quotes for particular news topics by using 17 million phrases
- Generates graphs based on a number of criteria: time period, topic, new story, etc.
- Extracts phrases from raw data from Spinn3r

Most Mentioned Phrases: 2008 Presidential Election



Response

- Website already has graphs of commonly used phrases for a quite a few news topics
- Interesting concept that gives people a flavor of a news story
 - Information could be misleading at times or not very substantive
- Helpful for tracking the latest stories and observing trends
- Results are useful because of the large sample size