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

Example Query

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

weakly oil may of German descent who comes from California.
This occurred about a year and a half ago, the hotel being in
rather bad shape and the purchase price being about $2,000,000.
[redacted] said he remembered he knew the head housekeeper of
the hotel who had been head housekeeper at the Plaza hotel
for 25 years. On contacting her he found that she and all
the other old employees of the Hotel Pierre had been dismissed.
This in itself was not so alarming but it was discovered that

Hotel Pierre "packed with employees of the Italian Consulate"   NEXT ▶

the hotel and a man named [redacted] (phonetic) is
[redacted].   Nothing whatsoever is known about the background
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. Clegg
Mr. Foxworth

p. 2

AR AI

August 27, 1940

Special Agent in Charge
Los Angeles, California

ALL INFORMATION CONTAINED

be
b7c

Description

The FBI's espionage investigation of
Paul Getty.

**Hotel Pierre "packed with employe[...]
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

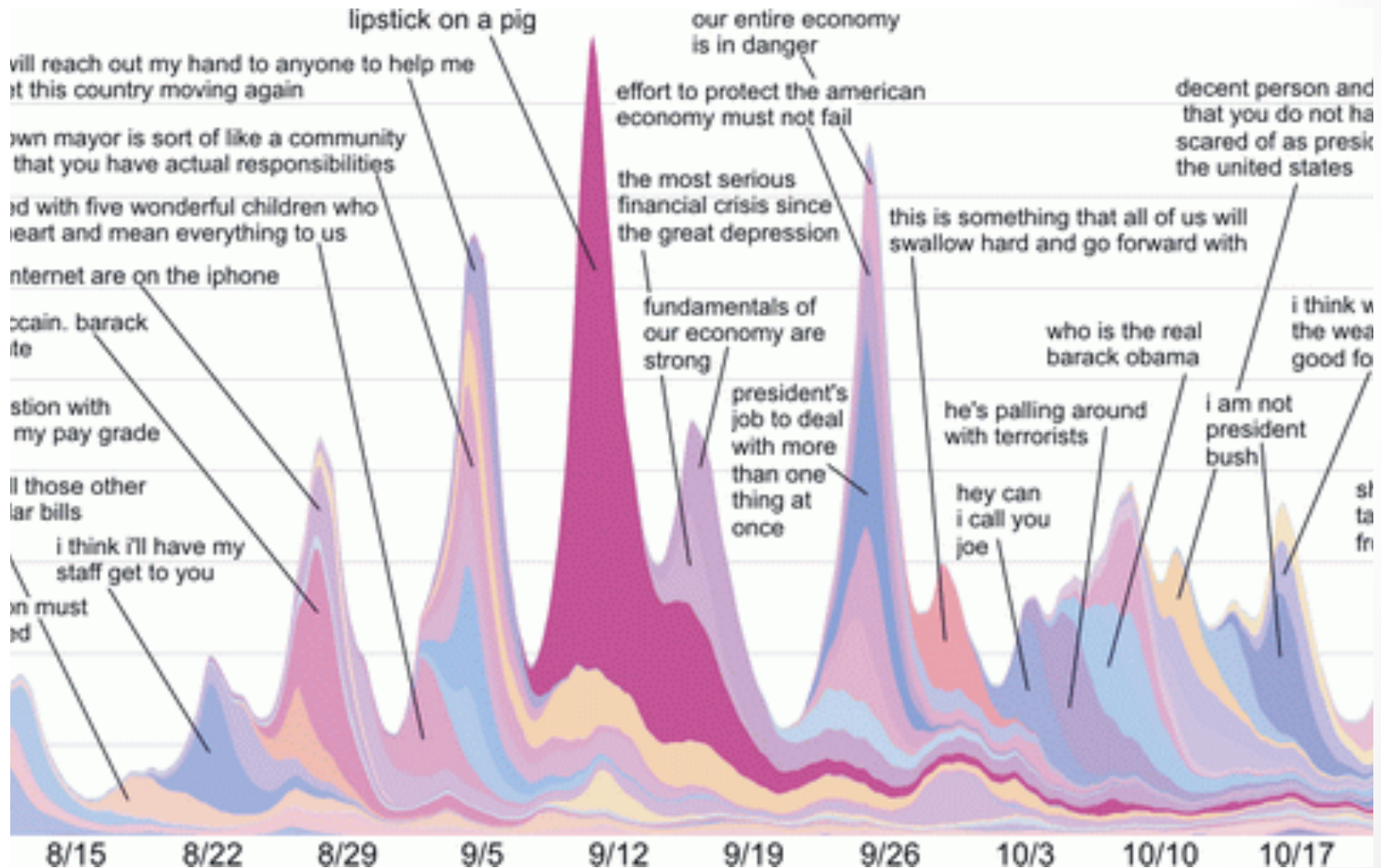# Searching Documents

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