

# News Event Detection Website

*Joe Acanfora, Briana Crabb, Jeff Morris*

CS 4624 Multimedia and Hypertext - a Capstone

May 5, 2015

Virginia Tech College of Engineering Department of Computer  
Science

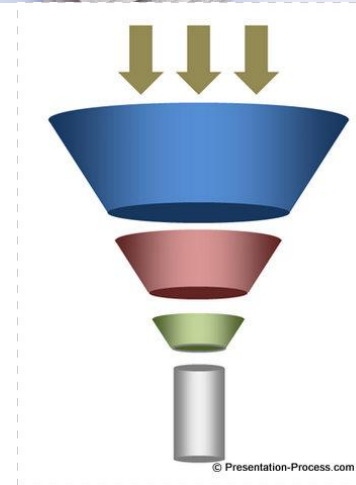
Dr. Edward A. Fox

# Overview

1. Project Purpose and Background
2. Twitter crawling script
3. Web crawling script
4. Reporting Service

# Project Background

The digital archive is a database to collect large amounts of articles about large events and to summarize those events.



Event Summary

# Project Purpose

The overall objective of this project was to build a website front end to help the client automate the web archiving process for their big data doctorate research.

# A Quick Walk Through

# News Submitter

Event Type:

Event Type

Keywords:

Hashtags:

Description:

Date:

Email:

Positives:

Browse...

No file selected.

<http://babs.dlib.vt.edu/twitter/index2.php>

## News Submitter

Event Type:

Shooting

Keywords:

virginia tech shooting

Hashtags:

#vtshooting #hokieshooting #32 #april16

Description:

tweets from the vt shooting

Date:

4/16/2007

Email:

bbcrabb@vt.edu

Positives:

Browse...

seed.txt






Submit

*input data*

Sign in with Twitter






# Your TwapperKeeper

archive your own tweets

Archive ID	Keyword / Hashtag	Description	Tags	Screen Name	Count	Create Time	
1	#WalterScott	Shooting - Walter Scott, SC		ideal_vt	5472	Thu, 09 Apr 2015 13:32:13 -0400	
2	Walter Scott	Shooting - Walter Scott, SC		ideal_vt	5058	Thu, 09 Apr 2015 13:32:28 -0400	
3	VirginiaTech	VT	#VT	newsEventDetect	2520	Fri, 17 Apr 2015 11:14:22 -0400	
12	hillary clinton	hillary running for president	#hillary16 #clinton16		10979	Wed, 22 Apr 2015 19:41:35 -0400	  
11	test	DESCRIPTION			19946	Wed, 22 Apr 2015 18:29:29 -0400	  

*Brings us to yourtwapper*



Archive ID	Keyword / Hashtag	Description	Tags	Screen Name	Count	Create Time	
1	#WalterScott	Shooting - Walter Scott, SC		ideal_vt	5472	Thu, 09 Apr 2015 13:32:13 -0400	
2	Walter Scott	Shooting - Walter Scott, SC		ideal_vt	5058	Thu, 09 Apr 2015 13:32:28 -0400	
3	VirginiaTech	VT	#VT	newsEventDetect	2520	Fri, 17 Apr 2015 11:14:22 -0400	
12	hillary clinton	hillary running for president	#hillary16 #clinton16		10979	Wed, 22 Apr 2015 19:41:35 -0400	 

*some of the previous twitter searches we have done*

```

starting ProjFocusedCrawler...
[(6.579729825986222, 2655), (6.438079308923196, 525), (6.236441962829949, 2807), (6.220355825078324, 205), (6.123963979403259, 602), (6.003946305945459, 3016),
(5.997212273764115, 3556), (5.941642422609304, 3415), (5.859812404361672, 3129), (5.762173934797756, 577), (5.61512051684126, 1414), (5.564348191467836, 1111),
(5.543294782270004, 2664), (5.418840607796598, 1121), (5.394449154672439, 2909)]
[u'sydney', u'sieg', u'cafe', u'police', u'police', u'news', u'news', u'hostag', u'add', u'comment', u'edit', u'said', u'say', u'australian', u'australia', u'peopl', u'video']
1, http://blogs.wsj.com/dispatch/2014/12/14/sydney-police-investigate-siege-situation-live/?mod=e2tw,
1, http://redflag.org.au/article/law-and-order-agenda-big-winner-out-sydney-siege,
1, http://tavernkeepers.com/australias-gun-control-laws-responsible-sydney-seige/,
1, http://tenplay.com.au/news/national/2014/12/16/Sydney%20siege%20ends,
1, http://www.abc.net.au/news/2014-12-16/sydney-siege-survivors-hostages-martin-place/5971188,
1, http://www.aunews.com/2014/12/three-reported-dead-as-sydney-siege.html,
1, http://www.bbc.co.uk/news/world-australia-30584120,
1, http://www.cnn.com/2014/12/15/world/asia/australia-sydney-hostage-situation/index.html,
1, http://www.dailytelegraph.com.au/news/world/islamic-state-praises-sydney-siege-attack-in-monthly-dabiq-magazine/story-fni0xs63-1227169934131,
1, http://www.independent.co.uk/news/world/australasia/sydney-siege-gunman-identified-by-police-as-wellknown-islamist-cleric-sheikh-man-haron-monis-9926122.html,
1, http://www.lifefnews.com/2014/12/16/sydney-siege-she-died-protecting-her-pregnant-friend-and-her-unborn-baby/,
1, http://www.theaustralian.com.au/in-depth/sydney-siege-victim-tori-johnson-loved-flowers-partner-says/story-fnqxbywy-1227161509920,
1, http://www.theGlobeandmail.com/news/world/sydney-cafe-siege/article22082542/,
1, http://www.theguardian.com/australia-news/2014/dec/21/sydney-siege-families-of-katrina-dawson-tori-johnson-direct-donations,
1, https://twitter.com/AJEnglish/status/546051801116856321/photo/1,
0.639687223835, http://www.independent.co.uk/news/world/australasia/sydney-lindt-cafe-siege-hostages-display-islamic-flag-as-up-to-20-people-are-held-by-armed-man-9924641.html, Two hostages and gunman dead in Sydney siege independent co uk news world australasia sydney lindt cafe siege hostages display islamic flag as up to 20 people are held by armed man 9924641
0.572365969652, http://www.independent.co.uk/news/world/australasia/sydney-siege-live-video-shows-the-moment-heavily-armed-police-officers-enter-building-9926304.html, Video:the moment police enter the building independent co uk news world australasia sydney siege live video shows the moment heavily armed police officers enter building 9926304
0.570781829733, http://www.independent.co.uk/voices/comment/sydney-cafe-siege-australia-becomes-the-latest-country-to-have-its-sense-of-invulnerability-shattered-9925230.html, Comment: Australia has its sense of invulnerability shattered independent co uk voices comment sydney cafe siege australia becomes the latest country to have its sense of invulnerability shattered 9925230
0.555176882184, http://www.wsj.com/articles/sydney-cafe-siege-police-negotiate-with-gunman-1418630919, Read more. wsj articles sydney cafe siege police negotiate with gunman 1418630919 mod WSJ_hp_RightTopStories
0.553076181493, http://www.independent.co.uk/news/world/australasia/sydney-siege-timeline-of-events-as-they-have-unfolded-at-lindt-cafe-9924741.html, Timeline of events as they unfolded independent co uk news world australasia sydney siege timeline of events as they have unfolded at lindt cafe 9924741
0.548009323571, http://mobile.abc.net.au/news/2014-12-16/sydney-siege-survivors-hostages-martin-place/5971188, Change to mobile view mobile abc au news 2014 12 16 sydney siege survivors hostages martin place 5971188 pfm sm
0.548009323571, http://www2b.abc.net.au/EAF/View/MailToQuery.aspx, Email www2b abc au EAF View MailToQuery abc au news 2014 12 16 sydney siege survivors hostages martin place 5971188
(<<class 'requests.exceptions.ConnectionError', ConnectionError(MaxRetryError("HTTPConnectionPool(host='www.kwoff.com', port=80): Max retries exceeded with url: /submit.php (Caused by <class 'socket.error': [Errno 104] Connection reset by peer"))),), <traceback object at 0x7f1ad95d8908>)
0.535441907181, http://www.newsvine.com/_wine/save, Add to Newsvine newsvine _wine save popoff 0 u dailytelegraph au news world islamic state praises sydney siege attack in monthly dabiq magazine story fni0xs63 1227169934131 sv e959be0e0f80f3e3cc74bc4686dced9e h IS praises Monis 26 238217 3B Sydney siege attack
0.534473367122, http://www.independent.co.uk/news/world/australasia/sydney-siege-outrage-as-bystanders-pose-for-selfies-outside-lindt-cafe-9924883.html, Bystanders pose for selfies outside Lindt cafe independent co uk news world australasia sydney siege outrage as bystanders pose for selfies outside lindt cafe 9924883
0.532110755749, http://www.wsj.com/video/islamic-flag-in-sydney-hostage-siege-analyzed/94887915-FBEB-49DC-859A-35F7F66800B8.html, Video: Flag in Sydney Hostage Siege

```

*focused crawler output*

Your request has been processed!

You can view your yourTwapperKeeper [HERE](#)

You can view your archived website [HERE](#)

*HTML email sent upon completion*

# Problems Faced

## Cloud9/Bluehost

- Blocking of PHP Functions
- Lack of SSH

Calling PHP Scripts non locally

Calling Python Scripts from Flask

# Acknowledgements

Mohamed Magdy Gharib Farag

- project sponsor
- [mmagdy@vt.edu](mailto:mmagdy@vt.edu)

PhD. Edward A. Fox

- class professor
- [eafox@vt.edu](mailto:eafox@vt.edu)

Sunshin Lee

- extra resource
- [sslee777@vt.edu](mailto:sslee777@vt.edu)