Using Social Media Data in Research

WebDataRA from WSI

Prof Leslie Carr

# Web Data Research Assistant

38 1933234183631015936

39 1933237469574606852

41 (933290443009380352

42 1933306807862226944

43 [93330720803243622

44 1933308253441798144

45 1933308379522584578

46 1933311814577909761

[93327724805199052

91 1.2

96 4.7

797 18

2 Janice G

3 Janice\_G

11 elynmoody

3 fujitsu uk

uncitamor

TechnicallyBo

sturdyAlex

Annual GCSE Computing Update with Awarding Organisat

his name is morrisseypic.twitter.com/0hbG5

X ☆	om				🕁 🔼 🕙 😥 🖌 🔚 🖪 🛆
s 🗯 Apple 🞽 Yahoo! 🄀 Google Maps 🗈 YouTi	ube 🗋 Wikipedia	Web Data Research Assistant	:		
Thome & Moments	ns 🖂 Messag	This table has been copied onto you	r clipboard - paste it ir	nto Excel for further	r analysis.
	What's	<ol> <li>On a Twitter search results pr more results, until the page s</li> <li>Data will also be extracted fr results and the data will be co</li> <li>Please be sure to treat the data acco small print.</li> </ol>	tays static for more the or a Google search pa ollected into a single ta	an 5 seconds. ge. Press the butto ble for convenienc	
lescarr	-		es Retweets Favorite	s Author	Text
@lescarr	BBC Ra Have yo	0E (0222424127E77700601 42	353 599	Harryslaststand	
TweetsFollowingFollowers12.6K8322,778	And hav @BBCiF	94 [933306391/36/4909/]	3	BBCRadio4	Signs, fossils, fissures in the rock, a missing child. This is amazing. http://bbc.in/2hKeFi0 #Trackspic.twitter.c
	Email: ip	93 [933320900568649728] 1	1 13	Rowenaaaaa	Working on playlists for BBC Music and the wonderful see that Celluloid Jukebox returns to @BBCRadio2 for
Trends for you · Change	BI	92 [933355800071942146]	14 24	Harryslaststand	
#Budget2017 It's Budget Day #PMQs		91 [933358690840784896]		sparrowbarley	Yes, and so the synthetic data must be made to be me too!https://twitter.com/lornamcampbell/status/933344
Corbyn tackles Brexit in pre-budget PMQs		90 [933098941607952385] 320	7.3K 27K	TVietor08	2016: @lvankaTrump will be a great moderating force dad that child molesters are bad.
Andrew Neil 2,259 Tweets Philip Hammond		89 [933358718229663744]	4 3	ODIHQ	Hart unit offin brief of the optimized of the optimize
Chancellor cracks jokes in his budget speech #hairyarchives	20	88 [933333317121970177] 14	182 100	HackneyAbbott	The Tories have U-turned on their promise to protect by £500m in real terms by 2020. #Budget2017pic.twit
@theonlywayisup, @Jisc and 1 more are Tweeting about this	2	87 [933319266052472838]	6 19	StandardIssueUK	K "I think the quickest way to subvert the status quo is b being diverse groups of people." Well said @susan_wo and @ladvcariad on this week's podzine. http://bit.lV/2
Top Gear 1,473 Tweets	-			L. I	BuzzFeed News AM to DM Watch
#budgetday2017 @PennyRed is Tweeting about this		States and the second	1.14		THE DIT
#StampDuty First time buyers will benefit from today's budget		And she's in the b	MMM	MAN	© 2017 Twitter About Help Center Terms Privacy policy Cookies Ads info Brand Blog Status Apps Jobs Marketing

- Scrapes Twitter, Facebook and Google data into a spreadsheet
- Uniquely allows free historic data capture
- No programming required
- Browser extension, one-click install •



NEW from @CLEAPSS CLEAPSS Keep it Safe! Health & Safety in practical Science for New TeachersSouthampton 12Dechts/https://twitter.com/Sotonmsic/status/93323418363101593/

Thanks to the General #DataProtection Resulation companies like #Uber will face multi-million \$ sanctions when they go https://twitter.com/JanAlbrecht/status/9332904430093803

This video will empower your staff to take ownership & become problem-solvers if tech fails & impacts customers http://chttps://twitter.com/fujitsu\_uk/status/93330680786222694

Are you looking to attend #uxcitymcr as a group of 3 or more? Get in touch for a group booking discount! http://goo.gl/YC https://twitter.com/uxcitymcr/status/933308253441798144

Just witnessed this elant seal being chased out the door of a fishmoneers shop in Wicklow town.pic.twitter.com/WGIGxiAi https://twitter.com/Derek1052/status/933308379522584578

In the UK, an annual rail season ticket from Horsham to London costs £3,912In Germany you can get a BahnCard 100 for € https://twitter.

ReTweeting yourself is like when you tell a joke and nobody laughs so you wait until there's a gap in the conversation and https://twitter.com

hampton@CompAtSch http://community.computingati https://twitter.com/CRCSouthEast/status/93323746957460

ne/status/93327724805199052

https://twitter.com/XEECEEVEVO/status/9333072080324

Nov-22 1511335342

Nov-22 1511336126

Nov-22 1511348755

23h23 hours age 1511352657

23h23 hours age 1511352753

23h23 hours age 1511353002

23h23 hours are 1511353032

23h23 hours age 151135385

Nov-22 1511345610

### Overview of Use



- The Web Data RA will capture Twitter, Facebook and Google data from a browser and allow you to paste a table of information directly into a spreadsheet. This tutorial focuses on its use with Twitter.
- 1. Visit **bit.ly/WebDataRA** in Chrome, click on the blue "+ Add to Chrome" button. The small green icon will appear in the top right of the browser window, next to the URL bar.
- 2. Go to **twitter.com** and create a Twitter search or display a timeline
- 3. Click on the **WebDataRA** icon to start collecting tweets.
  - Every 5 secs the browser will automatically scroll to the bottom of the page to make Twitter load the next batch of results and add the updates to the clipboard.
- 4. When you have collected enough results, paste the data into an Excel spreadsheet.
- 5. Use Excel to analyse data, or export to other programs such as Gephi or Voyant for other kinds of analysis.

#### WebDataRA Tables



			Count	
1234margs	1	3		
2202Gayle	1	1		
25Score	1	2		
50Folds	1	3		
90_mile_beach	1	1		
949powerfm	1	2		
ABRAMSbook s	1	5		
ADJBlog	1	2		
AFARmedia	1	2	2	6
ASMRLittleMe lon	1	1		
ASTsupportAA li	1	2		
AbhaReach	1	2		
AdamtheBullF AN			1	1
AdeldMeyer			1	1
AdrianChen			1	2
AgenceW	1	2		
AlanSee	1	3		
AlexRWilliams	1	4		
AlishaValerie	1	2		
Aloe_Vera_FL P	1	2		
AmlyBotanical s			1	6
Andrea_MacE	1	2		
AngelaInBold			1	1
Anitapratapsing		2		
AnnaBell_write s			1	5
Anonfighter5			1	1
AppliedIG	1	4		
Argentine58			2	2
AstraLifeID	1	5		
AvonworthEA	1	2		
BBC			1	1
BBCJonSopel	1	1		
BBCWomansH our			1	2

Account Author Count ...Inc RTs Mention ...Inc RTs

Hashtags	Count	Inc RTs
##skincare	2	2
#124;	1	1
#365daysofself care.	1	1
#AI.	1	1
#AageSeRight	1	4
#AbuDhabi	1	2
#Apps	1	2
#Argyll	1	15
#BCNMI	1	1
#BLOGOSPHE RECHAT	1	2
#BabyZakTime	1	2
#Bachelor	1	3
#BeWellDoWe ll	1	5
#BienvenueChe zNous	1	2
#BlogosphereC hat	15	27
#BlogosphereC hat.	1	1
#CBCconf18	1	3
#Cafepress	2	3
#CaribbeanSea!	1	7
#CellPhone	1	7
#Cheshire	2	19
#Children	15	51

Target	Weight
BlogosphereM	
ccfest	
edfringe	
sundarakarma	
tanyagoodin	
Anonfighter5	
iamwellandgoo d	
DanConnors16	
BBC	
OilPaul	
SebGorka	
adidas	
betablogr	
LloydEdwards9	
	ecfest edfringe sundarakarma tanyagoodin Anonfighter5 iamwellandgoo d DanConnors16 BBC OilPaul SebGorka adidas betablogr LloydEdwards9

The tweet data, with author, mentions, hashtags, text and counts of retweets, replies and likes broken out in separate columns. Account occurrence summary, a count of the number of times that each Twitter account appears in the dataset as author or a mention (including retweets). Counts of the appearances of each hashtag.

1

#Classer

A table of edges of the conversational network, i.e. the number of times each pair of accounts communicate with each other.

### Using The Tweet Data Table

- The tweet data (gray) contains the basic data about each tweet: what was said, when, by who and to whom.
- Use this data to form a general overview of the communication over time and identify the most significant tweets.
- Examine specific tweets and their context by referring back to the Twitter site using each tweet's URL.



# Pivot Table Visual Twitter Timeline

- Click on any gray cell in the Tweet Data table
- Choose "Pivot Table" from the Insert ribbon.
- In the Pivot Table builder
  - drag "Author" from the Field Name panel into the "Rows" panel
  - drag "Timestamp" into the "Columns" panel
  - drag "Author" (again) into the "Values" panel (it will automatically turn into "Count of Author").



- Reformat to create a helpful Timeline summary of contributors (vertical axis) by days (horizontal axis).
  - narrow the columns, slant the column headings, change the angle of the text to 60°
  - use the "Row Labels" control to sort by the author count
  - show only the rows where the total author count is greater than a chosen threshold.
  - use conditional formatting to highlight the most extreme values.

### Other Questions to ask of the Data

- All kinds of summaries and analyses are possible using Excel on this data, including:
  - Showing the distribution of the tweet sample through time
  - Identifying the most prolific and/or popular actors, and showing their activity through time
  - Showing the use of individual hashtags (this might be useful in a big conversation, or one that evolves over a longer period)
  - Comparing the relative proportion of contributions from different actors / hashtags

# Using The Account Data Table

- The account table (green) shows
  - the most active tweeters,
  - the most frequent repliers,
  - the most retweeted users.
- This shows the key actors in a conversation, and the main roles that they take.
- Get detailed information by clicking on the account names (linked) to see the account bios and the relevant timelines of these actors in the Twitter website.
- Understand whether they are corporate accounts, private individuals, bots or trolls.

Account	Author Count	Inc RTs	Mention Count	Inc RTs
1234margs	1	3		
2202Gayle	1	1		
25Score	1	2		
50Folds	1	3		
90_mile_beach	1	1		
949powerfm	1	2		
ABRAMSbook				
5	1	5		
ADJBlog	1	2		
AFARmedia			2	6
ALLITAUSUAII	1	2		
a ASMRLittleMe lon	1	1		
ASTsupportAA				
li	1	2		
AbhaReach	1	2		
AdamtheBullF				
AN			1	1
AdeldMeyer			1	1
AdrianChen			1	2
AgenceW	1	2		
AlanSee	1	3		
AlexRWilliams	1	4		
AlishaValerie	1	2		
Aloe_Vera_FL P	1	2		
AmlyBotanical			1	6
Andrea MacE	1	2		
AngelaInBold		-	1	1
Anitapratapsing	1	2		
AnnaBell_write				
5			1	5
Anonfighter5			1	1
AppliedIG	1	4		
Argentine58			2	2
AstraLifeID	1	5		
AvonworthEA	1	2		
BBC			1	1
BBCJonSopel	1	1		
BBCWomansH				
our			1	2

### Inspecting Twitter Accounts

Account	#	Bio
ItsTimeToLogOff	30	Time To Log Off is the home of digital detox. We're spearheading the movement to disconnect regularly from digital devices and reconnect with the world offline. We do this through collecting facts on the need for digital detox, running campaigns to get everyone off their screens and hosting retreats, events and workshops.
DinnerTableMBA	9	A commercial organisation working together to help families become more confident, successful, and self-empowered
SpareFoot	8	A storage company. We make it easy to move and store your stuff. Reserve storage for free and get your mind out of the clutter.
CultureEffect	5	Author of Digitox: How to Find a Healthy Balance for your Family's Digital Diet

The account names in the account "author and mentions" (green) table are clickable, and open the page of the account profile in your default web browser.

Following the account hyperlinks for the most prolific authors in the green table, we see that they are all commercial or institutional actors to one extent or another.

### Using The Hashtag Data Table

• The hashtag table (blue) shows you the most frequently used hashtags. This can help you extend your data gathering to look for more tweets relevant to your research question.

Hashtags	Count	Inc RTs
##skincare	2	2
#124;	1	1
#365daysofself	1	1
care.	1	1
#AI.	1	1
#AageSeRight	1	4
#AbuDhabi	1	2
#Apps	1	2
#Argyll	1	15
#BCNMI	1	1
#BLOGOSPHE	1	2
RECHAT	1	2
#BabyZakTime	1	2
#Bachelor	1	3
#BeWellDoWe	1	5
11	1	5
#BienvenueChe	1	2
zNous		-
#BlogosphereC hat	15	27
#BlogosphereC		
hat.	1	1
#CBCconf18	1	3
#Cafepress	2	3
#CaribbeanSea!	1	7
#CellPhone	1	7
#Cheshire	2	19
#Children	15	51
#ClaraSC	1	1

# Using The Edge Data Table

- The edge table (yellow) will help you to see the interactions between actors, and help you to understand groupings of actors, and the pattern of their interaction.
  - Is a key account dominating a conversation and talking to many others?
  - Are they responding or just being passive recipients of marketing messages?
  - Is there a group of equals having a balanced conversation with equal participation?

Source	Target	Weight	
1234margs	YouTube		1
2202Gayle	BlogosphereM		1
2202Gayle	cefest		1
2202Gayle	edfringe		1
90_mile_beach	sundarakarma		1
ABRAMSbook s	tanyagoodin		1
ASMRLittleMe lon	Anonfighter5		1
Andrea_MacE	iamwellandgoo d		1
AppliedIG	DanConnors16		1
BBCJonSopel	BBC		1
BBCJonSopel	OilPaul		1
BBCJonSopel	SebGorka		1
BVG_Kampag ne	adidas		1
BVG_Kampag ne	betablogr		1
BarryEdwardsJ r	LloydEdwards9 1		1
BellyDanceRav			

#### Inspecting the Conversation Network

- Copy and paste the yellow table *into a separate spreadsheet* and save it as a CSV file (call it *edgetable.csv* or similar).
- Load up the network visualisation program "Gephi", and start a new project.
- In the "Data Laboratory", choose "Import Spreadsheet" and load up the CSV data as an *edge table*.



You can then apply a variety of network layout algorithms in the "Overview" pane.

### Understanding the Conversation Network

- Many summaries and analyses are possible using Gephi's network visualisations.
- Showing the interaction of the network actors
- Identifying the communities and active participant subgroups within the larger sample
- Identifying the roles of different actors in the communications network



## Textual Analyses of the Social Conversation

- In the gray table, copy the "Sanitised Text" column.
  - This contains the text of all the texts, but with all the Twitter features (@names, #hashtags, URLs) removed to leave only the English text.
- Go to the Voyant-Tools.org website
  - Voyant Tools is a textual corpus analyser. It considers a Twitter conversation as a single document & individual tweets as individual sentences.
- Paste the text into the textbox
- Press the "Reveal" button.
  - You will see a screen with several panels that help you explore the text of the tweets in different ways.

# **Textual Analyses**

- Voyant includes a variety of textual analysis components
  - Word cloud
  - Trend analyser
  - Concordance
  - Summary
  - Vocabulary cluster analysis
  - Dimensional Reductions
  - Co-occurrence Network



### Sentiment Analysis

- Sentiment analysis can help you identify positive or negative comments in your sample.
  - This is a popular method in industry, especially with brand management companies. However it is academically contested, and does not have a high degree of transparency in the lexical processing.
- Paste the "Sanitised Text" column into **sentigem.com**.
- Consider to what extent the results seem accurate to you? How well does it identify positive and negative 'sentiment' in a tweet?
  - What kinds of inaccuracies can you see?
- Does it help you to identify any points of interest in your data for more thorough investigation?

entigem	Sign in / Register
ositive sentiment egative sentiment eutral sentiment	Overall sentiment – positive
now all	This Is What It's Like To Not Own A Smartphone In 2018 Digital detox weekend starting now
nalyze again	Have a good one everyone
	I'm off to put my head in the offline world
	Hahaha
	Digital detox time
	Being off 'social media' for a bit makes you realise what gossip you've missed out on
	What a bunch of bitchy gossipers we all really are
	not quite the digital detox I had in mind I love social media and the benefits it offers, yet I'm aware it sometimes takes over too much of my life
	This year I've pledged to change that; it's working so far
	Some fab tactics for digital detox by here too: Make sure you take the time you need
	Social media keeps us connected, maybe not always a good thing though, I'm considering a digital detox
	sounds like a 100 days suggestion right there
	Try It: Your 2018 'Digital Detox' Guide How to Do a Digital Detox with Your : [And why it's super healthy for your to do] Woooohoooo
	It's the weekend – the perfect time to get your digital detox on: "taps wineglass" folks, i have gathered you here today to something something digital detox gwenyth paltrow going retrograde embracing obsolescence etcetera
	i'll be online & on email if you need me, still posting, just no messenger or whatsapp
	I'll probably only last a week
	As a result, I feel better rested and that I'm enriching my 'soul'
	The 'digital detox' has left me freer and more engaged
	It also helps that I am 4 months free of coffee so generally carrying a little less nervous energy
	So, all in all, feeling pretty
	Dependency On The Internet/A Digital Detox Fysiek detoxen en nu ook digitaal detoxen: Digital-detox, zondag 18 februari
	Wie durft het aan
	I am struggling to read books lately, very unlike me, and looking at my statsmaybe a digital detox is in order
	But, y'know not a total detox because it's not like I have a problem?
	Is it time that you divorced the digital
	(not forever don't panic) We have a package coming right up that will give you the digital detox you need ;-) Digital detox is hard
	Christina Wodtke decided to take a "digital sabbath" from social media for one day a week