The archived web for research in the humanities and social sciences

Jane Winters, Professor of Digital Humanities, School of Advanced Study, University of London
TPDL 2018 workshop, Porto, 13 September 2018
Live web v archived web

• We’re all familiar with the live web, and we know how to search and browse effectively
• BUT, contrary to the narrative that nothing ever disappears from the web, it’s extraordinarily ephemeral
• 2014 study of URLs cited in the *Harvard Law Review* – 70% no longer linked to the originally cited information (Zittrain, Albert & Lessig, ‘Perma: scoping and addressing the problem of link and reference rot in legal citations’)

sas.ac.uk
Where Are They Now?
Not Really OK

Percentage of OK or MOVED URLs

SIMILAR = 0
SIMILAR > 0
SIMILAR > 20
SIMILAR > 30
SIMILAR > 40
SIMILAR > 50
SIMILAR > 60
SIMILAR > 70
SIMILAR > 80
SIMILAR > 90
IDENTICAL

www.bl.uk
Web archives are an increasingly important resource for arts and humanities researchers, yet we have neither the expertise nor the tools to use them effectively. Both the data itself and the process of collection are poorly understood, and it is possible only to draw the broadest of conclusions from current methods of analysis. The project will work with the dataset derived from the UK domain crawl from 1996 to 2013 (that is, when legal deposit legislation was extended to cover digital materials), totalling approximately 65 terabytes and constituting many billions of words. For the arts and humanities, this is very big data indeed.
Aims of the project

• To highlight the value of web archives for research
• To develop a theoretical and methodological framework for the analysis of web archives
• To explore the ethical implications of big data research
• To inform collections development and access arrangements at the British Library
• To train researchers in the use of big data
Project outputs

• An enhanced interface providing access to the data at the British Library
• An open-access monograph: *The Web as History: Using Web Archives to Understand the Past and the Present* (UCL Press)
• Two short animations explaining web archives to the general public
• A series of case studies presenting research using the archive
Researcher case studies I

• The UK Web Archive and Beat literature
• Online reactions to institutional crises: BBC Online and the aftermath of Jimmy Savile
• A history of UK companies on the web
• Digital barriers and the accessible web
• Searching for home in the historic web – the blogs of the London-French
Researcher case studies II

- Revealing British Euro-scepticism in the UK Web Archive
- Looking for public archaeology in the UK Web Archive
- Do online networks exist for the poetry community?
- Capture, commemoration and the citizen historian
- The online development of the Ministry of Defence and armed forces
Welcome

Welcome to SHINE, our prototype historical search engine.

About SHINE

This tool has been developed as part of the Big UK Data Arts and Humanities project funded by the AHRC. Read more about the project on our blog.

The data was acquired by JISC from the Internet Archive (IA) and includes all .uk websites in the IA web collection crawled between around 1996 until April 2013 (you can read more details about this dataset here). The collection comprises over 3.5 billion items (urls, images and other documents) and has been full-text indexed by the UK Web Archive. Every word of every website in the collection can be searched for and analysed.

What can I do?

Trend analysis

Use ‘trends’ to analyse the number of pages a word or phrase appears in the collection over a given period (within 1996-2013). Comparisons can be drawn by adding several words or phrases separated by a comma. E.g. cat, dog, goldfish
Click on a point on the graph to show a random sample of the matching records from that year.
Found 100 samples matching "credit crunch" from 2007.

<table>
<thead>
<tr>
<th>Link</th>
<th>Matching Text</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>solihullobserver.co.uk</td>
<td>CREDIT crunch, high interest rates</td>
<td>realbusiness.co.uk</td>
</tr>
<tr>
<td>estateangels.co.uk</td>
<td>and you’re left with an unpalatable set of...</td>
<td>lowapcredit.co.uk</td>
</tr>
<tr>
<td>residentialandlord.co.uk</td>
<td>CREDIT crunch, and you’re left with an unpalatable set of...</td>
<td>lutterworthobserver.co.uk</td>
</tr>
<tr>
<td>estateangels.co.uk</td>
<td>credit crunch has made the adverse credit mortgage marke...</td>
<td></td>
</tr>
</tbody>
</table>
Figure 9: Maps of the UK universities under study for three years: 2000, 2005, and 2010. The connections are the hyperlinks, and color corresponds to the normalized strength of each link ($\sigma_v$). The reddest links correspond to the strongest connections.

Hale, Yasseri, Cowls, Meyer, Schroeder and Margetts, ‘Mapping the UK webspace: fifteen years of British universities on the web’, arXiv:1405.2856v1
Grenfell Tower Fire

The Grenfell Tower fire broke out on 14 June 2017 in the 24-storey Grenfell Tower block of flats in North Kensington, West London leading to the death of over 70 people. This collection brings together some of the websites, news articles and social media accounts related to the fire.

Search “Grenfell Tower Fire” special collection

69 sites in collection
London Terrorist Attack 7th July 2005

Collection of Internet sites with content about the terrorist attacks in London on 7th July 2005 begun immediately after the event. The collection comprises news and commentary, public information and support for victims and weblogs written by those who had first hand experience of the bombs.

Sites (65)

- 7th July Family Assistance Centre
- Abu Basir al-Tartusi
- al-Aqsa
- Al Arab Online
- Al Haqaeq
- al-Khila\fah
- Al-Mushahid Assayasi Magazine
- Al Quds - Al Arabi
- Asharq Al-Awsat
- Association for British Muslims
- Association of Chief Police Officers - London Terrorist Attack 7th July 2005
- Azzaman Newspaper Online
- BBC News: BNP campaign uses bus bomb photo
- BBC News: London Attacks
- Brazilian News
- British Humanist Association (BHA)
- British National Party (BNP)
Click here for London terror attack coverage

THURSDAY, JULY 07, 2005

London tube explosions

So, what’s this “bang” on the tube all about then? Anyone got any clearer idea than “either a bomb or a big crash”? And does this count as Sod’s Law, coming the day after the Olympics announcement?

Update: BBC now reports “reports of an explosion”, and says Transport for London are saying it’s thanks to a power failure. Hmm... “walking wounded” caused by a power failure?

Fuck: Explosions at Aldgate, Edgware Road, King’s Cross, Old Street and Russell Square.

Unconfirmed rumours of 30 dead.

Piss - Russell Square may have been on a bus.

10:05am - Right - I’m calling this - terrorist attack. Sky News seems to be the only one sitting. I’m trawling for eyewitnesses.

“I saw lots of people running up a road and then saw the top of a bus destroyed,” one eyewitness told Sky News.

LBC News radio is still reporting “power surges”, but also advising “get away from the City, but don’t use public transport”.

10:15 - Eyewitness reports of “half the bus in the air” at Russell Square. BBC news site down. Radio still confused.

10:20 - FTSE dropping 3%, Pound dropping vs. the dollar. Reports of “at least one explosive device on the Underground”. If you have TV access, ITV have rolling coverage, apparently.
July 07, 2005

Fat Accompli

Surviving a Terrorist Attack

Fate is a strange thing. On this particular day a series of events transpired such that I ended up on a Tube train that was destroyed by terrorists. Fortunately it was only the carriage in front of me, but tragically it resulted in a serious amount of injuries. This is my story.

I boarded the train at King's Cross after a series of line closures forced me onto a Circle Line train; little did I know at the time that this was probably the worst thing I could have done.

Travelling just past Edgware Road Station the train entered a tunnel. We shook like any usual tube train as it rattled down the tracks. It was then I heard a loud bang.

The train left the tracks and started to rumble down the tunnel. It was incapable of stopping and just rolled on. A series of explosions followed as if tube electric motor after motor was exploding. Each explosion shook the train in the air and seems to make it land at a lower point.

I fell to the ground like most people, scrunch up in a ball in minimize injury. At this point I wondered if the train would ever stop, I thought 'please make it stop', but it kept going. In the end I just wished that it didn't hit something and crash. It didn't.

When the train came to a standstill people were screaming, but mainly due to panic as the carriage was rapidly filling with smoke and the smell of burning motors was giving clear clues of fire.

As little as 5 seconds later we were unable to see and had all hit the ground for the precious air that remaining. We were all literally choking to death.

The carriage however was pretty sealed; no window could open, no door would slide and no hammers seemed to exist to grant exit. If there were instructions on how to act then they were impossible to see in the thick acrid black smoke.

In the end I opted to do something about the problem and began shouting to find out in which direction the fires were emanating from. I then tested with the inter-carriage door to see if venting the smoke caused fire to spread. It didn't so I held the door open trying to clear the carriage and look for escape routes.
July 08, 2005

Fat Accompli

Surviving a Terrorist Attack: What a difference a day makes.

The world seems hungry for information. My last blog entry has totally mushroomed over the world and where I’d normally get a few hits per day from friends and family I’m now being flooded. Even various news agencies have been trying to get me to talk, some being met with silence. I feel I need to clarify and tell you what happened next.

Before I do however, I have a basic and clear warning to all the journalists that are reading this: I know you need to verify your stories, and I will help you verify that I exist as much as I can, but if you’re unsure on my authenticity then simply don’t print and quote me.

Look, I’ve travelled in a train that was out of control and dealt with a situation. I’ve not contacted you, and when you did contact me I’ve not asked for money, I’ve been very clear that I only want to report facts and I’ve denied every TV and radio station access for this very reason. If you’re interested in over-inflating a situation then you might as well close your browser now and walk away; I view such journalism with contempt. That’s blunt, but I want to be clear.

Theories, Speculation and Rumour

First: Explosions. It is impossible for me to verify whether I was on the particular train that was blown up or one of the secondary trains. I heard no loud explosions, merely loud but muffled bangs. This could merely be us riding debris from an original explosion being derailed and then hitting further debris as we travelled down the tunnel out of control. I feel this is important to point out as news agencies are desperate for me to say that MY train was the one that exploded.

Know what? I don’t care. I cared about surviving and living. I cared about escaping and not choking. I cared about not walking into a King’s Cross ticket station style fire. I had no inkling of terrorism at the time, but then again I also had no inkling on where I could escape to, whether I could breathe again or whether the train would actually stop. My priorities were very survival related I’m afraid.

Second: Medical Attention. It appears that medical attention was relatively rapid in coming, but it also appears that access was from the other end of the tunnel. So my original statements stand: I do not deny that the fire services and medics weren’t fantastic, merely that at OUR END of the tunnel.
08/07/2005

Brendan Murphy

Hello MCB. Just to say the mails you are getting saying "it's war on muslims in Britain" are written by morons. They belong to the same stone age as the people who did the bombing. They're not representative - they just have the loudest voices.

There may be a rise in hate crimes against muslims, but long term these animals will not succeed. There'll always be ignorant, racist bigots in the UK but most of us do not feel this way. I would like to state categorically that my experience of Muslim people (here in the UK and elsewhere) is that they are hardworking, peaceable, sociable and extremely hospitable. In fact - it's an ill-wind that blows no good - but this type of extremist violence may have long term positive results. If Islamic organizations like the MCB start to have a higher profile in community life in the UK, through awareness raising about the Islam, what the life of ordinary muslims is like in the UK, and closer co-operation with other community agencies, then some good will come of these tragedies.

A small anecdote from my own experience. I have friends in Morocco (Muslims of course) whose first reaction after September 11 was to send me a letter begging me to write back to tell them I was OK. They're poor and uneducated, and were perhaps confused about where the attacks had happened. (UNITED states/UNITED kingdom). It was a touching reminder that ordinary people everywhere find violence against innocent people repulsive and distressing.

Peace to you all

08/07/2005

Gary

Hello,

I understand that in the aftermath of yesterday's attacks in London, you will have received many abusive emails. I have just heard that a mosque has also been attacked in Leeds. I just wanted to point out that for every 'hater' there are many many thousands of ordinary decent people who are the exact opposite and wish only for peace for all. I'm one of those.

Cheers

08/07/2005

Chris Thomson

Dear Muslim Council,

I have just heard on the TV that you have received derogatory emails following yesterday's atrocities. I am sure you have also received messages of support and camaraderie - we are Britons together. I only wish to add my own humble support and apologise for those dim-wits who don't engage a single brain cell in distinguishing between right and wrong, good and evil, fair and unfair.... the list could go on.

Of course you condemn acts of terror. I am amazed that such a thing needs to be spelled out in the UK.

I hope you are safe and well.

Yours sincerely
London explosions - 7 July 2005

Around 200 firefighters were called to explosions at Aldgate, Edgware Road and Kings Cross London underground stations and an explosion on a bus at Tavistock Square on Thursday 7 July.

Twelve fire appliances with 60 firefighters attended the incident at Edgware Road, twelve fire appliances with 60 firefighters attended the incident at Kings Cross, ten fire appliances with 50 firefighters attended the Aldgate incident and four fire appliances with 20 firefighters were called to Tavistock Square.

Throughout the morning several of our new specialist fire rescue units were deployed to work with the other emergency services to evacuate casualties and make the incidents safe.

The Mayor of London, Ken Livingstone, issued the following statement on 7 July after the incidents in London.

Book of condolence and relief fund

The Mayor opened a book of condolence to the public on Monday 11 July.
All programmes on this week
Programmes with a web site are highlighted

A
Afternoon Shift, The
All in the Mind
Analysis
Any Answers?
Any Questions?
Archer, The

B
Ballad of the Sad Cafes
Belle on Sunday
Book at Bedtime: A Month in the Country
Brain of Britain
Breakaway

C
Call Edward Stourton
Children's BBC Radio 4: Over Sea, under Stone
Classic Serial: Mansfield Park
Classics with Kay
Comparing Notes with Brian Kay

D
Dear Diary
Desert Island Discs
Diamonds

E
Education Matters
Ambridge

For a virtual tour of Ambridge click on the place names below. If you just want to brush up on your 'Who's Who in Ambridge go to The Cast.

KEY:
- The Village
- Lakey Hill
- River An
This week's summaries

Nice Mr Chris Harrison (of University Challenge fame) has enjoyed himself so much as your Ambridge correspondent this summer, that he has volunteered to carry on for the foreseeable future in providing you with his plot summaries. This means that I get a break after about a year, for which I am most grateful!

There's not a lot on the Internet about The Archers, but here's a (not very definitive) list of what there is. If you know of any more, please let me know. Please remember that this is a best efforts service, so please don't take it too badly if it's not up to your highest standards, but feel free to make constructive comments.

- The UMRA newsgroup
- Chris Harrison's Archers page
- Daniel Gordon's map of Ambridge
- The Archers Family Tree
- "Archers Addicts" fan club

Past weeks beginning:

1997:

- January 6th, 13th, 20th, 27th
- February 3rd, 10th, 17th, 24th
- March 3rd, 10th, 17th, 24th, 31st
- April 7th, 14th, 21st, 28th
- May 5th

1996:

- December 2nd, 9th, 16th, 23rd, 30th
- November 4th, 11th, 18th, 25th
- October 7th, 14th, 21st, 28th
- September 2nd, 9th, 16th, 23rd, 30th
- August 5th, 12th, 19th, 26th
- July 1st, 8th, 15th, 22nd, 29th
- June 3rd, 10th, 17th, 24th
- May 6th, 13th, 20th, 27th
- April 1st, 8th, 15th, 22nd, 29th
- March 4th, 11th, 18th, 25th
I am reliably informed that this page can be printed for friends who are not blessed with an internet connection in the following manner: (X-)Windows users can highlight the text of the tree, copy it, and paste it into a word processor. Changing the font to a non-kerning one such as Courier at 6pt should display the tree within one page A4 landscape. Whether anyone will be able to read it once printed is another matter. Non (X-)Windows users (aka The Chosen Few) can find their own alternatives; they might like to start by saving the page as HTML verbatim.

Notes

- Names in CAPITALS were born Archers.
- Characters in brackets ( ) are not (as far as I know) speaking characters.
- Characters in square brackets [ ] are dead.
- Numbers in curly brackets {} refer to the footnotes below.
Ambridge anguish boosts Archers audience

Julia Day
Thursday January 30, 2003

BBC Radio 4’s The Archers - a usually mild-mannered tale of country folk - gripped a record number of 5 million listeners with a racy plot about sex and lies.

The story of Brian Aldridge’s adultery had 4.7 million people a week hooked, desperate to hear him coming clean to his wife about fathering Siobhan Hathaway’s love child.

The storyline also helped the programme to its second largest ever audience for its Sunday omnibus edition - 2.1 million listeners - according to figures for the last three months of 2002 from audience measurement company Rajar.

However, Radio 4 lost 114,000 listeners for the period - with its total number of listeners down to 9.798 million - although for the second quarter running it was London’s most listened to station, closely followed by BBC Radio 2.

Radio 2 piled on 775,000 listeners, with Steve Wright gaining his biggest ever audience of 6.4 million.

But Radio 2’s gain is BBC Radio 1’s loss. Radio 1 lost 374,000
Conservative party deletes archive of speeches from internet

Decade's worth of records is erased, including PM's speech praising internet for making more information available
BBC Food website is closing down - here's how to make sure you don't lose all your favourite recipes

The BBC may be archiving thousands of its beloved recipes, but fear not! There is still a way to get hold of them online.
Thank You

It's because of everyone involved, all across the country, that we achieved this magnificent result. This is your victory.
The £350 million we give to the EU every week could pay for...

- a fully staffed NHS hospital/
- 10 brand new schools/
- 8,000 new police officers

GET THE FACTS