

UK Team Resources:

(Screenshot souvenirs of work produced in Summer Research Camp)

Working UK publications database (categorized by region, political affiliation, etc.):

Source title	Hits in Nexis Uni	Priority Level	Years We Have	Query Terms
i - Independent Print Ltd	159 Humanities, 3093 The Arts	Already Queried	2012-2017 (H); 2012-2017 (A)	"humanities"; "arts"
Metro (UK)	77 Humanities, 841 The Arts	Already Queried	2012-2017 (H); 2012-2017 (A)	"humanities"; "arts"
mirror.co.uk		Already Queried	2017 (H); 2014-2016 (H); 2013-2017 (A)	"humanities"; "arts"
Morning Star	38 Humanities, 763 The Arts	Already Queried	2012-2013 (H); 2012-2013 (A)	"humanities"; "arts"
The Daily Mail and Mail on Sunday		Already Queried	2017 (H); 2012-2017 (A); 2000-2016 (H)	"humanities"; "arts"
The Daily Star		Already Queried	2017 (H); 2013, 2015-16 (H); 2013-2017 (A)	"humanities"; "arts"
The Daily Star Sunday		Already Queried	2017 (H); 2012-2017 (A)	"humanities"; "arts"
The Daily Telegraph		Already Queried	2017(H), 2000-2016(H), 2012-2017 (A)	"humanities"; "arts"
The Express	61 Humanities, 1893 The Arts	Already Queried	2016-2017 (H); 2012-2017 (A)	"humanities"; "arts"
The Guardian (London)		Already Queried	2012-2017 (arts); 2000-2016 (humanities)	"humanities"; "the arts"
The Independent (United Kingdom)		Already Queried	2017 (H), 2000-2016 (H), 2012-2017 (A)	"humanities"; "arts"
The Mirror, The Daily Mirror and The Sunday Mirror		Already Queried	2017 (H); 2012-2017 (A); 2000-2011; 2012-2017 (H)	"humanities"; "arts"
The New European	11 Humanities, 71 The Arts	Already Queried	2012-2017 (H); 2012-17 (A)	"humanities"; "arts"
The News of the World	8 Humanities	Already Queried		
The Observer (London)		Already Queried	2017 (H); 2012-2017 (A); 2000-16 (H)	"humanities"; "arts"
The People	4 Humanities	Already Queried	2012-2017 (A); 2003, 2017 (H)	"arts," "humanities"
The Sunday Express		Already Queried	2012-2014 (H); 2012-2017 (A)	"humanities"; "arts"
The Sunday Telegraph (London)		Already Queried	2017 (H); 2012-2017 (A); 2001-2016 (H)	"humanities"; "arts"

RegionalNationalSunday NewspapersPolitical AffiliationSpecialty/NicheOnlineTabloidsHigh Impact/Topic

Source title (by UK county)	Hits in Nexis Uni	Priority Level	Years We Have	Query Terms
Durham (North)				
The Northern Echo	93 Humanities, 1365 The Arts	Already Queried	2012-2017 (A); 2007, 2009-10, 2012-17 (H)	"humanities", "arts"
Sunderland Echo	27 Humanities, 410 The Arts	Already Queried	2009-2010, 2014-2016 (H); 2012-2017 (A)	"humanities", "arts"
Derby and Yorkshire (North)				
Derby Telegraph	42 Humanities, 287 The Arts			
The Star (Sheffield)	160 Humanities, 661 The Arts			
Hull Daily Mail	76 Humanities, 1804 The Arts	Already Queried	2012-17(A), 2000-1 (H), 2003-17 (H)	"arts", "humanities"
Yorkshire Evening Post	54 Humanities, 946 The Arts	Already Queried	2012-17(A), 2002 (H), 2005-8 (H), 2010-12 (H), 2013-17 (H)	"arts", "humanities"
York Press	51 Humanities, 661 The Arts			
Bradford Telegraph and Argus	34 Humanities, 437 The Arts			
Ashbourne News Telegraph	3 Humanities, 77 The Arts			
Malton & Pickering Mercury	2 Humanities, 29 The Arts			
Matlock Mercury	4 Humanities, 159 The Arts			
Yorkshire Post		Already Queried	2012-17 (A), 2002 (H), 2005-08 (H), 2010-12 (H), 2013-17 (H)	"arts", "humanities"
Lancashire (North)				
Liverpool Echo	135 Humanities, 1622 The Arts	Already Queried	2012-2017 (A), 2000-2004 (H), 2006-2017 (H)	"arts", "humanities"
Manchester Evening News	282 Humanities, 4851 The Arts	Already Queried	2012-17 (A), 2000-08 (H), 2010-17 (H)	"arts", "humanities"
Lancashire Evening Post	41 Humanities, 276 The Arts			

RegionalNationalSunday NewspapersPolitical AffiliationSpecialty/NicheOnlineTabloidsHigh Impact/Topic

UK List of workable topic models created:

Topic Model Number	Grouping Name	Names Of News Sources Used	Year Range of Data	Number of Topics	Number of Clean Articles
2	National Papers	The Eveningstandard (London), The Guardian (London), The Daily Mail and Mail on Sunday, The Observer (London), The Mirror: the Daily Mirror and the Sunday Mirror, Daily Star, The Sunday Telegraph (London), Daily Star Sunday, Mirror.co.uk, Daily Telegraph (London), The Independent (United Kingdom), Metro (United Kingdom), City AM, Independentprintitd, The New European, The Express	2017	75	
3	National Papers	The Eveningstandard (London), The Guardian (London), The Daily Mail and Mail on Sunday, The Observer (London), The Mirror: the Daily Mirror and the Sunday Mirror, Daily Star, The Sunday Telegraph (London), Daily Star Sunday, Mirror.co.uk, Daily Telegraph (London), The Independent (United Kingdom),	2017	50	
4	National Papers	The Eveningstandard (London), The Guardian (London), The Daily Mail and Mail on Sunday, The Observer (London), The Mirror: the Daily Mirror and the Sunday Mirror, Daily Star, The Sunday Telegraph (London), Daily Star Sunday, Mirror.co.uk, Daily Telegraph (London), The Independent (United Kingdom),	2017	200	
5	Scotland Regional Papers	The Herald (Glasgow), The Sunday Herald, The National, Scotsman, dailyrecord.co.uk	2017	50	
6	Conservative Papers	The Express, The Daily Mail and Mail on Sunday, The Daily Telegraph, The Sunday Telegraph, City AM, The Evening Standard	2017	50	
7	Conservative Papers	The Express, The Daily Mail and Mail on Sunday, The Daily Telegraph, The Sunday Telegraph, City AM, The Evening Standard	2017	50	
8	Labor Papers	The Mirror: the Daily Mirror and the Sunday Mirror, The Guardian, Mirror.co.uk	2017	50	
9	Guardian	The Guardian (London)	2012-17	50	
10	Guardian	The Guardian (London)	2012-17	50	

AM Team 2 Log

The purpose of this log is to maintain a record for the benefit of future project researchers (e.g., next year's RAs) about what a team did during summer 2018. Please quickly jot down main tasks worked on, and include any failed or dead-end lines of research (so that future researchers will know you tried that!)

7/5/18

Working from Alan's list of proposed newspapers, found which ones were in LexisNexis and prioritized those for queue over the weekend. Tried to identify Sunday papers where appropriate, though this may need a more thorough check. Four tiers of priority, based on the relative number of hits for key terms. Queries for "humanities," "the arts," and "liberal arts" in each source for the date range 1/1/12-12/31/17.

7/9/18

Discussions of possible topics to run - by political orientation of paper, by search term, by national/local categories. Some sources seem to be unavailable through LexisNexis API (like the London Times).

7/10/18

Attempted UK topic model with January 2017 “humanities” data that was returned from LexisNexis. UCSB team had trouble getting through Jupyter notebooks in full - got stuck in notebook 3 making topic model. Sam able to get a topic model at CSUN.

These sources unable to be retrieved from LexisNexis: The Daily Record and Sunday Mail; The Sun (England); Financial Times (London); The Sunday Times; The Times (London).

7/11/18

UCSB team got topic model to work for Humanities Jan 2017.

Spent time labeling topics in both CSUN and UCSB models and taking notes (in Topic Model Interpretation doc).

Possible tasks for tomorrow: 1) UCSB should finish labeling last few topics 2) CSUN and UCSB could compare 3) decide whether to query more sources over the weekend - should we, for instance, query all UK sources in Lexis Nexis? (tons of these) 4) make a plan for next week while Abigail away.

7/12/18

Sam generated new topic models:

- 100 topics for Humanities 2017 in the Mirror
- 100 topics for Humanities 2017 in Guardian
- 100 topics for Humanities (without Humanity) 2017 in Guardian

Jennifer and Abigail discovered that the search features on Lexis Nexis may not be helpful beyond large categories - for instance, “Entertainment and Arts” doesn’t seem like a meaningful category for us. Made spreadsheet for compiling new possible sources, starting with categorizing list of UK newspapers from LexisNexis. Scott provided code for limiting search to “humanities” only in Jupyter Notebooks. While away next week, Abigail will work on drafting the Focus Report for UK.

7/16/18

Jennifer added more high-circulation publications to the sources spreadsheet, including a few popular National publications that we might want to add to our queries, Sunday newspapers, and some of the most popular regional newspapers. This task is ongoing. Samantha began

working on topic models of the UK national newspapers. Samantha also created a spreadsheet to keep track of the topic models that were created and the browsers that were generated.

7/17/18

Jennifer continued to work through/categorize the list UK newspapers so that we can better-understand our source corpus and prepare informed queries/analysis. Samantha continued working on topic models of National UK papers. Samantha also began experimenting with the number of optimal topics based on the number of articles being topic modeled. Preliminary findings suggest that: if the number of articles is between 10 and 50, then 50 topics is optimal; if the number of articles is between 51 and 250, then 75 topics is optimal; if the number of articles is higher than 251, then 100 topics is optimal. Further experimentation is required to validate if these number of topics are actually optimal. Samantha will look into the optimal number of topics based on words in the corpus.

7/18/18

Jennifer continued to work through/categorize the list UK newspapers so that we can better-understand our source corpus and prepare informed queries/analysis. Samantha continued to work on creating topic models and experimenting with the optimal number of topics. Samantha refined her preliminary findings on the optimal number of topics. **The refined findings are: if the number of articles is between 30 and 100, then 50 topics is optimal; if the number of articles is between 101 and 250, then 75 topics is optimal; if the number of articles is higher than 251, then 100 topics is optimal.**

7/19/18

Samantha continued working on creating topic models. Samantha began experimenting with combining publications to create topic models. Samantha continued to test the number of optimal topics based on the number of articles in the corpus. Samantha used excel to find the exponential trendline, using the number of articles in the corpus as the x-axis and the number of optimal topics as the y-axis, to create a preliminary equation that will calculate the optimal number of topics. The equation is: $y=62.146e^{0.0006x}$. Further experimentation is required to further refine the equation and ensure its efficacy for determining the optimal number of topics. Jennifer began to interpret the topic models that have been generated this week, continued categorizing sources, and submitted additional source queries.

7/23/18

Sam began rerunning topic models with over 150 articles with the fixed DFR browser. In Lab Notebook, Abigail compiled file names from Jupyter directory that we could combine for large-scale topic models of “humanities” and “the arts” in 2017 UK print sources. Jennifer tracked hits for “the arts” alongside the “humanities” data in the source spreadsheet to guide our selection of future queries. Jennifer also found a wikipedia article about the political endorsements of different newspapers. Questions came up for us about how best to categorize our topic models - for instance, the difficulty of pegging certain newspapers as national or regional.

7/24/18

Decision that we should not focus effort on interpreting previous topic models before today. Sam made a new topic model for 75 topics, humanities in national papers. Abigail worked on categorizing the files we have for possible future topic models. Jennifer continued to classify spreadsheet for political parties, regions, etc. Abigail and Jennifer worked on interpreting Sam's topic model. **A first goal is to compare Arts results to Humanities results to see if the arts is worthwhile pursuing.**

(Working) Topic Model

#2 National Papers, Humanities, 2017

Date run: 7/24/18

Topics: 75

Notes:

-Many topic models are difficult to interpret, but others are productively specific

On scaled view:

- There are a few groupings of educational topics - seems like University education, Primary/Secondary education, and then an accreditation cluster? (about funding and scores, as compared with China, eg, application process, job salaries, etc)
- Trump topics are outliers
- British politics topics tend to cluster together
- Social justice issues as a cluster? (Climate, race politics)

(Working) Topic Model

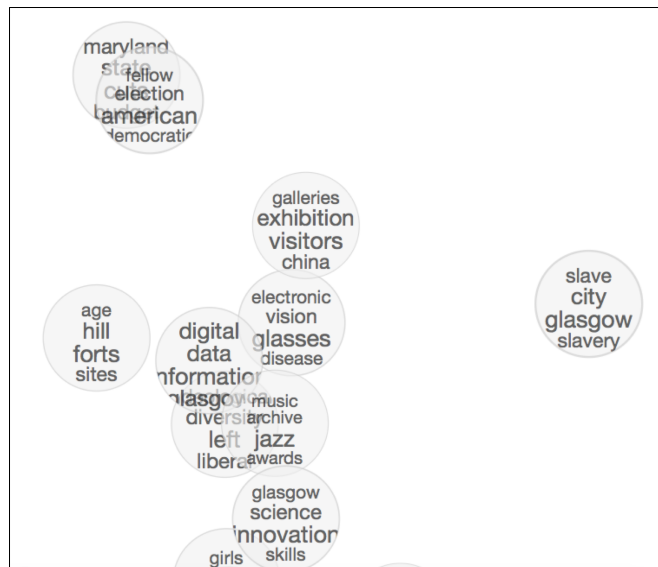
#5 Scotland Regional Papers - Humanities - 2017

Date run: July 25, 2018

Topics: 50

Notes:

- Small number of articles - does the topic modelling still have the same effect on this size corpus (in the 40s)
- Very interesting scaled view - everything clusters together except two of the most prevalent topics
- Possible that this model includes data from the Daily Record from Baltimore, MD alongside the scottish version? (Yes, rerunning)



7/26/18

Sam made topic model for Labour Party papers. Jennifer began comparing Labour to Conservative party topic models. We also decided which projects to delete (not migrate) over to Harbor (everything on and before 7/23). Decision to prioritize “the humanities” queries over “the arts” just so that it’s more manageable. Plenary meeting with whole group.

(Working) Topic Model

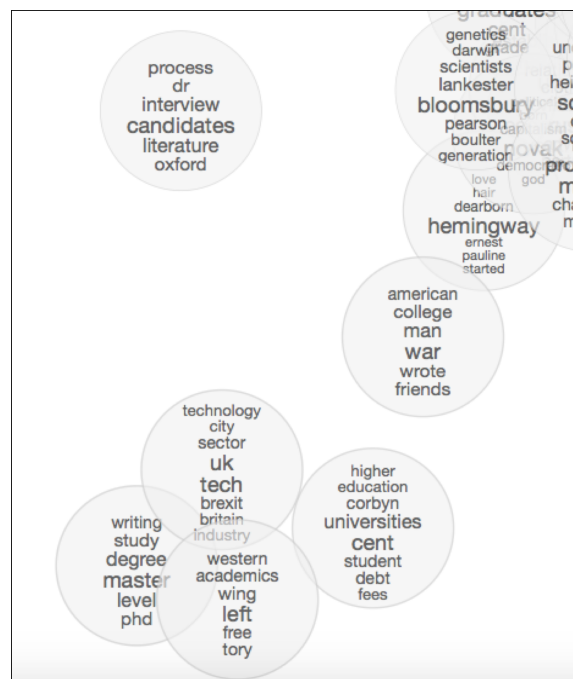
#6 Conservative Papers - Humanities - 2017

Date run: July 26, 2018

Topics: 50

Notes:

- A lot of the topics are undecipherable
- A lot of the topics that are decipherable are about school or university life
 - In fact, topic 1, which is very focused on school/university, is 16% of the corpus
- On the scaled view, the topics that are outliers are about school (perhaps topics that are more squarely on the topic of “humanities”)
- The model produces topics with the words “liberal” or “conservative” (which seems somewhat unique)
- Also a lot of topics that mention gender terms



7/30/18

Updated sources spreadsheet to reflect new data collected over the weekend. Worked on editing and expanding the focus report. Reported number of publications per region to Lindsay. Began new topic modeling on Harbor - a priority is to get some humanities/arts comparisons.

7/31/18

Samantha generated topic models to compare arts and humanities in The Guardian. Published UK Team Focus Report.

8/1/18

Samantha re-ran outdated topic models. Samantha also continued to refine the optimal number of topics equation. Samantha found that a logarithmic equation was a more accurate “best fit trendline” to predict the optimal number of topics than an exponential equation. Samantha will continue to update the exponential equation, just in case, and will begin to refine a logarithmic equation as we collect more qualitative data. Samantha’s preliminary logarithmic equation is: $y=23.389\ln(x)-41.638$.

8/2/18

Samantha made new topic models for both Wales and Northern Ireland. Jennifer and Abigail interpreted models from National newspapers (humanities) 2017, and Scotland (humanities) 2000-2017. Looking good! Getting discrete topics in productive directions.