

Tasks for 27-07-2020 :

- stemming + Lemmatization : done
- In my effort to make sense of the articles' maximum occurrence number (7 in our case) to be taken in the analysis, I came across new findings that I need to discuss with you.
- I managed to find some articles with considerably high number of occurrences (my theory is that they are pop-ups but still we should include them)

For example :

APPLAUSE was heard across the East End today as Celtic fans bid an emotional farewell to another of their greatest heroes.\n\nStevie Chalmers' funeral cortege passed by Celtic Park for the last time on Wednesday afternoon.\n\nSupporters gathered at the ground in their numbers five days after thousands turned out to pay their last respects to Chalmers' former captain, Billy McNeill.\n\nStevie Chalmers funeral as it happened: Glasgow's farewell to second Lisbon Lion in a week\n\nThe forward scored the most important goal in Celtic's history, the 1967 European Cup final winner against Inter Milan, one of 231 he netted for the club.\n\nA funeral mass took place in Calton, near Parkhead, for Chalmers, who died aged 83 last week after suffering from dementia in recent years.\n\nThe service took place at 11.30am at St Mary's Church, where Celtic Football Club was first established during a meeting in 1887.\n\nNeil Lennon, Sir Alex Ferguson, the Celtic squad and the remaining Lisbon Lions were all in attendance.\n\nArchbishop Philip Tartaglia told the congregation: "I offer our sympathies to the Celtic family. He was, of course, one of the incomparable Lisbon Lions who won the European Cup. Stevie had the tremendous distinction of scoring the winning goal.\n\nREAD MORE: Celtic submit plans to parade through Glasgow following treble treble win\n\n"Stevie's family say his priorities in life were his faith, family and football. That is surely a fitting legacy.\n\n"Stevie and Billy's names have become synonymous over the past two days. This has been a distressing time for their families.\n\n"You could not hope to meet two different men. And yet, as we all know, there was something that joined them together - and it was how their lives came to an end.\n\n"To score the winning goal earned him the right to say he scored the most important goal in the history of Scottish football. He was proud to score that goal, but to him it was more important that the team won.\n\nFor all your breaking Glasgow news, click here, or head to our Facebook, Twitter and Instagram pages\n\n"What Stevie did that day in Lisbon had an affect on the whole nation. Nobody was really convinced a team from Scotland could be kings. It gave the country a whole new set of aspirations."n\nAfter the service, the cortege then passed through the Celtic Way just after 1pm to rapturous applause before going onto a private family cremation.\n\nCeltic Football Club described Chalmers as "the man who scored the most important goal in the club's history."

Outlook for Friday to Sunday\n\nFurther showers or outbreaks of rain will continue to move in from the west on Friday, and it will be rather windy. The winds and showers will ease a touch on Saturday, although more widespread rain may move in from the south-west overnight as an area of low pressure moves in. This will lead to a rather wet and unsettled day on Sunday, with some heavy and prolonged rainfall.	27
RAPTUROUS applause was heard across Glasgow on Friday as the funeral cortege of Billy McNeill made its way from the city's Garnethill to Celtic Park.\n\nFans and mourners earlier gathered in the City Centre to pay their respects to the former Celtic captain and manager.	27
Heavy showers are likely across south Wales and southern England on Tuesday with a small chance of disruption from flooding.\n\nSpray and flooding could lead to difficult driving conditions and some road closures\n\nWhere flooding occurs, there is a slight chance of delays or cancellations to train and bus services\n\nThere is a small chance that homes and businesses could be flooded, causing damage to some buildings\n\nThere is a small chance of fast flowing or deep floodwater causing danger to life\n\nThere is a small chance that some communities become cut off by flooded roads\n\nThere is a slight chance of power cuts and loss of other services to some homes and businesses	27
1 Permit to do something (6)\n\n4 Potential oaks? (6)\n\n8 Put down on paper (5)\n\n9 Dominance (7)\n\n10 Vehicle seen on the roads in wintry conditions (7)\n\n11 Synthetic silk-like fabric (5)\n\n12 Bogus trap (anag) — somewhere good to eat and drink? (9)\n\n17 Ghastly (5)\n\n19 Convey (7)\n\n21 Out of order — erratic (7)\n\n22 Cause of computer failure? (5)\n\n23 One suffering for their beliefs (6)	27
News... but not as you know it	27

CELTIC fans have been blasted for leaving a section of Glasgow in a "disgusting" state following treble treble celebrations.\n\nDespite an open-top parade to celebrate the Hoops' victory being cancelled on Saturday amid safety fears, hundreds still poured out onto Saltmarket, High Street, Trongate and Gallowgate to party.\n\nNow, Evening Times readers and environmental charity Keep Scotland Beautiful have hit out at the piles of bottles, cans, plastic bags and other debris left strewn across the streets.\n\nOne reader wrote on our social media: "Look at all the mess they made all the rubbish shocking." \n\nAnother added: "The mess they left they streets in - never heard of a bin?!??" \n\n"Mess of streets and I say that about any team fans the mess is terrible," was another comment.\n\n"The mess left behind is disgusting," added another.\n\nOne reader wrote: "Disgusting mess they have left to clean up." Another reader replied: "You beat me to it. I was in town on Saturday night to pick my daughter up from work and the mess in Saltmarket was awful." \n\nPaul Wallace, operations manager at Keep Scotland Beautiful, told the Evening Times the mess was "not acceptable." \n\nIN PICTURES: Celtic fans party on Gallowgate in Treble Treble celebrations\n\nHe said: "Everyone wants to have a good time and celebrate but, sadly, this is all too typical across Scotland after public events – and it is not acceptable." \n\n"We all have a responsibility to act differently, not drop litter, and to use the bins when they are available." \n\n"We are already bringing local authorities, businesses and communities together to tackle litter levels across Scotland. However, they alone cannot solve this problem." \n\n"We all need to work together to deliver the level of behaviour change required so we don't drown in a sea of disregarded rubbish." \n\nREAD MORE: Number of Glasgow children given £80 fines for dropping litter in last year revealed\n\nIain Gulland, chief executive of Zero Waste Scotland, added: "Dealing with litter and fly-tipping costs over a million pounds a week in Scotland - and that's money that can be better spent on other priority services." \n\n"Litter is national problem and one that can be easily solved. It's everyone's responsibility to put waste in the right place and help keep our streets clean. We would encourage everyone to do their part and make use of recycling and general waste bins." \n\nGlasgow City Council says on its website that it "delivers a zero-tolerance approach to littering," adding that an £80 penalty is applied "in line with the Scottish Government's national guidance." \n\nThe local authority also says it provides a 24-hour street cleansing and litter service, seven days a week, 365 days a year. Allowing the public to report litter, the local authority adds: "Litter is unsightly and potentially dangerous. It creates unnecessary work for our teams who are responsible for litter clearance throughout the city." \n\nA spokesman for the council told us "there is no excuse for dropping litter." \n\nFor all your breaking Glasgow news, click here, or head to our Facebook, Twitter and Instagram pages\n\nHe added: "Everyone has the responsibility to dispose of their rubbish in the right way at all times." \n\nPeople are required to put their rubbish in a bin and should keep their waste with them until they can find a bin they can use." \n\nIt is a constant source of disappointment and frustration that there are people in Glasgow who use the streets as their own personal litter bin."

21

These are also considered articles even though they have high occurrences. Then you find these advertisement texts that also have high occurrence and are irrelevant to the analysis:

	text
Series 1: 1. Nice Face\n\nprotect the only witness.\n\nWhen a politician is murdered, an MI5 security officer must protect the only witness. ... \n\nWhen a politician is murdered, an MI5 security officer must protect the only witness.	24
How to use a discount code for online shopping\n\nRedeeming online deals is as quick as 1, 2, 3, so hop on these voucher codes and promotional offers while they're still around! The first step to easy savings? Preview our chosen favourites in the Top 20 or click through some of the most popular shops featured above.\n\n\n\nThe second step to easy savings: Visiting our pages where we list every available valid code and deal. Once you've found the voucher that piques your interest, click on the green button to open the webshop in a new tab. Clicking on a code? Copy the promo code and continue to shop. Before you finalise your order, just paste the code where needed. Apply to save! Clicking on a deal? You'll be led directly to the sale or product, or your discount will be applied automatically to your basket. How simple was that?\n\n\n\nWe're always on the lookout for a bargain and we continuously add fresh deals and vouchers, so bookmark our page and never pay full price again!	24
Login\n\nEnter your details below to login\n\nIf you are an existing member of The Times and The Sunday Times enjoying the full benefits of thetimes.co.uk, then simply enter your Times+ login details below and press 'Enter'\n\nEnter your details to login\n\nEmail address\n\nPassword	24
Happy Birthday !\n\nWe hope you like your gift!\n\nPlanning a birthday binge? Have your cake and eat it with our box sets on All 4.	24

!!!! I went a step back and also combined the text count with the irrelevant subdomains to see if those are omitted or not from the data when we used the domain paths given by Ina. I realised that the number of occurrences alone is irrelevant for the cleaning and that I need to find another way.

Irrelevant subdomains :

https://docs.google.com/document/d/1ZumFo_s8Urrl0H9sAx245fDHfmyA_tyaWQInoqE_nqI/edit#heading=h.192ux9de45up

Here is an example to clarify using the subdomain “discountcode” even though they don’t appear a lot. The title and the text are clearly not articles

In [78]: `media_portals[media_portals['subdomain']=="discountcode"]`

Out[78]:

	title	text	authors	imgs	movies	meta_lang	meta_keywords	htmlfile	
257	Warehouse discount code	How do I redeem a Warehouse code?	Click on ...	None	[https://d3tsytm1wtjqo2.cloudfront.net/images/...	None	en	GER_2019-03-15_19465_web_page.html	discountcode.c
258	Warehouse discount code - June	jumpers\njumpsuits\ncoats\nndenim\nnskir...		None	[https://d17fy28x8ov4h8.cloudfront.net/images/...	None	en	GER_2019-03-15_19468_web_page.html	discountoo
569	Beefeater voucher → Save on your next meal in ...	Savour a mouthwatering meal with these Beefeater...		None	[https://d3tsytm1wtjqo2.cloudfront.net/images/...	None	en	GER_2019-03-15_45270_web_page.html	discountcode
905	FREE SAMPLES: check the Molton Brown discount ...	Relax with exotic fragrances and classic Londo...		None	[https://d3tsytm1wtjqo2.cloudfront.net/images/...	None	en	GER_2019-03-15_73794_web_page.html	discountcode.dail
1191	Exclusive Kwik Fit discount codes → 10% OFF	Why spend with Kwik Fit?	If you're looking ...	None	[http://fox13now.com/assets/images/build/metro...	None	en	GER_2019-03-16_8319_web_page.html	discou
...
95661	Domino's voucher code	Save more at Domino's!	Domino's deals The ...	None	[https://d3tsytm1wtjqo2.cloudfront.net/images/...	None	en	GER_2019-06-15_77348_web_page.html	discountcode.daily
95729	Argos discount codes	Find it, get it, Argos discount code it!	Re...	None	[https://d3tsytm1wtjqo2.cloudfront.net/images/...	None	en	GER_2019-06-15_80337_web_page.html	discountc
95730	Argos discount codes	Find it, get it, Argos discount code it!	Re...	None	[https://d3tsytm1wtjqo2.cloudfront.net/images/...	None	en	GER_2019-06-15_80338_web_page.html	discountcode.daily
95731	Argos discount codes	Find it, get it, Argos discount code it!	Re...	None	[https://d3tsytm1wtjqo2.cloudfront.net/images/...	None	en	GER_2019-06-15_80339_web_page.html	discountcode.daily
95732	Free Voucher Code Finder	We and our partners use cookies and similar te...		None	[https://d3tsytm1wtjqo2.cloudfront.net/images/...	None	en	GER_2019-06-15_80340_web_page.html	discountco

ftjqo2.cloudfront.net/images/...	None	en	GER_2019-03-15_45270_web_page.html	discountcode.dailyemail.co.uk/beefeater	NaN	uk	dailyemail.co.uk	discountcode	1	1
ftjqo2.cloudfront.net/images/...	None	en	GER_2019-03-15_73794_web_page.html	discountcode.dailyemail.co.uk/molton-brown	NaN	uk	dailyemail.co.uk	discountcode	1	1
m/assets/images/build/metro...	None	en	GER_2019-03-16_8319_web_page.html	discountcode.metro.co.uk/mwik-fit	NaN	uk	metro.co.uk	discountcode	1	1
...
ftjqo2.cloudfront.net/images/...	None	en	GER_2019-06-15_77348_web_page.html	discountcode.dailyemail.co.uk/dominos#voucher-3...	NaN	uk	dailyemail.co.uk	discountcode	3	3
ftjqo2.cloudfront.net/images/...	None	en	GER_2019-06-15_80337_web_page.html	discountcode.dailyemail.co.uk/argos	NaN	uk	dailyemail.co.uk	discountcode	3	8
ftjqo2.cloudfront.net/images/...	None	en	GER_2019-06-15_80338_web_page.html	discountcode.dailyemail.co.uk/redirect-to?url=x...	NaN	uk	dailyemail.co.uk	discountcode	3	8
ftjqo2.cloudfront.net/images/...	None	en	GER_2019-06-15_80339_web_page.html	discountcode.dailyemail.co.uk/argos#voucher-334207	NaN	uk	dailyemail.co.uk	discountcode	3	8
ftjqo2.cloudfront.net/images/...	None	en	GER_2019-06-15_80340_web_page.html	discountcode.dailyemail.co.uk/extras?utm_source...	NaN	uk	dailyemail.co.uk	discountcode	14	10

- Evaluation of different lda models with tf-idf :

The table below is the result of LDA model trained over Tf-Idf matrix : (the discussion is below) for multiple numbers of topics.

Number of topics	Topic keywords	coherence	Perplexity
20	[(0, '0.036**"lbc" + 0.024**"vaccin" + 0.024**"tattoo" + 0.023**"promo" + 0.022**"harper" + 0.020**"cheshir" + 0.017**"hiv" + 0.017**"salah" + 0.014**"cheryl" + 0.013**"measl"', (1, '0.060**"badg" + 0.030**"golf" + 0.017**"hudson" + 0.017**"huey" + 0.015**"mcilroy" + 0.014**"neill" + 0.012**"jd" + 0.012**"murphi" + 0.012**"ni" + 0.012**"chicago"', (2, '0.019**"pier" + 0.015**"spice" + 0.014**"bangladesh" + 0.014**"colin" + 0.013**"defin" + 0.012**"thailand" + 0.012**"friendship" + 0.011**"homophob" + 0.011**"hitler" + 0.010**"mileston"', (3, '0.036**"arsen" + 0.023**"leagu" + 0.020**"liverpool" + 0.018**"tottenham" + 0.016**"furi" + 0.015**"europa" + 0.014**"juventu" + 0.012**"klopp" + 0.012**"solskjaer" + 0.012**"ford"', (4, '0.006**"love" + 0.005**"star" + 0.004**"girl" + 0.004**"credit" + 0.004**"fan" + 0.004**"ann" + 0.004**"itv" + 0.004**"murder" + 0.003**"mum" + 0.003**"life"', (5, '0.051**"mep" + 0.038**"jen" + 0.030**"sharon" + 0.022**"bowi" + 0.017**"librari" + 0.016**"mason" + 0.016**"dijk" + 0.016**"olli" + 0.013**"sullivan" + 0.012**"bloc"', (6, '0.045**"vettel" + 0.031**"hamilton" + 0.024**"sec" + 0.024**"isi" + 0.023**"ferrari" + 0.021**"caravan" + 0.016**"opera" + 0.016**"rylan" + 0.016**"davidson" + 0.016**"merced"', (7, '0.049**"iranian" + 0.032**"holli" + 0.028**"netanyahu" + 0.020**"solar" + 0.019**"bruce" + 0.018**"mueller" + 0.018**"israel" + 0.014**"fletcher" + 0.014**"gaza" + 0.013**"stanley"', (8, '0.013**"hancock" + 0.012**"acid" + 0.011**"lincolnshir" + 0.009**"nurs" + 0.009**"diet" + 0.009**"chef" + 0.008**"hampshir" + 0.008**"belgium" + 0.008**"recip" + 0.008**"jane"', (9, '0.007**"trump" + 0.006**"johnson" + 0.004**"bbc" + 0.004**"presid" + 0.004**"bori" + 0.003**"music" + 0.003**"women" + 0.003**"radio" + 0.003**"queen" + 0.003**"polit"', (10, '0.032**"tariff" + 0.027**"korean" + 0.025**"cliff" + 0.019**"bolton" + 0.018**"halifax" + 0.017**"coal" + 0.016**"croatia" + 0.016**"jong" + 0.016**"valeri" + 0.014**"guitar"', (11, '0.028**"norri" + 0.022**"kri" + 0.018**"xbox" + 0.016**"trailer" + 0.012**"nuditi" + 0.012**"shah" + 0.011**"margat" + 0.011**"endgam" + 0.011**"trevor" + 0.010**"chemistri"', (12, '0.029**"lgbt" + 0.029**"stacey" + 0.023**"emmerdal" + 0.018**"arcadia" + 0.017**"maya" + 0.015**"erupt" + 0.014**"chloe" + 0.014**"treasur" + 0.014**"gogglebox" + 0.013**"kerri"', (13, '0.083**"bst" + 0.032**"broadband" + 0.031**"dwp" + 0.027**"thame" + 0.020**"newport" + 0.015**"clare" + 0.014**"florenc" + 0.013**"durham" + 0.010**"indec" + 0.010**"jan"', (14, '0.024**"password" + 0.018**"normandi" + 0.014**"newborn" + 0.013**"nicol" + 0.013**"ant" + 0.012**"gulff" + 0.012**"terrac" + 0.011**"jennif" + 0.011**"riot" + 0.011**"tactic"', (15, '0.004**"imag" + 0.003**"polic" + 0.003**"uk" + 0.002**"peopl" + 0.002**"england" + 0.002**"subscrib" + 0.002**"email" + 0.002**"cup" + 0.002**"london" + 0.002**"club"', (16, '0.038**"walsal" + 0.032**"backgrid" + 0.012**"kirsti" + 0.012**"johnston" + 0.011**"snooker" + 0.011**"maxwel" + 0.011**"darren" + 0.010**"morton" + 0.010**"erdogan" + 0.010**"clooney"', (17, '0.038**"brexit" + 0.031**"parti" + 0.024**"vote" + 0.022**"mp" + 0.020**"tori" + 0.020**"eu" + 0.019**"elect" + 0.018**"labour" + 0.017**"conserv" + 0.016**"minist"', (18, '0.030**"saudi" + 0.022**"jenni" + 0.019**"vegan" + 0.019**"embassi" + 0.018**"arabia" + 0.017**"zara" + 0.014**"gaga" + 0.013**"wolf" + 0.013**"drake" + 0.013**"priest"', (19, '0.010**"destin" + 0.010**"appl" + 0.010**"closur" + 0.010**"cannabi" + 0.010**"sarri" + 0.010**"aircraft" + 0.009**"swim" + 0.008**"warner" + 0.008**"clinic" + 0.007**"tini"])]	0.44535757102 14375	-11.64527649533 0531
19	[(0, '0.026**"nevil" + 0.018**"chloe" + 0.015**"asylum" + 0.012**"dani" + 0.012**"moscow" + 0.011**"low" + 0.011**"countdown" + 0.010**"xbox" + 0.010**"gigg" + 0.010**"hollyoak"', (1, '0.019**"arcadia" + 0.015**"fifa" + 0.013**"jessi" + 0.013**"backgrid" + 0.010**"trailer" + 0.010**"carriag" + 0.010**"spider" + 0.009**"leopard" + 0.009**"tenant" + 0.009**"cva"', (2, '0.016**"polic" + 0.009**"court" + 0.009**"attack" + 0.008**"offic" + 0.008**"car" + 0.008**"man" + 0.008**"road" + 0.007**"hospit" + 0.007**"investig" + 0.006**"murder"', (3, '0.025**"stacey" + 0.022**"nadal" +	0.47918022215 59691	-11.36331243997 2303

	<p>0.015**solskjaer" + 0.015**coventri" + 0.015**sean" + 0.014**ronaldo" + 0.013**anderson" + 0.013**neymar" + 0.013**homophob" + 0.012**johan"), (4, '0.014**hull" + 0.010**lincolnshir" + 0.009**corbyn" + 0.009**wild" + 0.009**midfield" + 0.009**exam" + 0.009**hamilton" + 0.009**subscript" + 0.008**laura" + 0.008**vettel"), (5, '0.011**walker" + 0.008**cocain" + 0.007**fat" + 0.007**strictli" + 0.007**cook" + 0.007**freddi" + 0.006**growth" + 0.006**meat" + 0.006**athlet" + 0.006**sugar"), (6, '0.017**sandwich" + 0.016**netanyahu" + 0.016**consent" + 0.015**vaccin" + 0.015**pd" + 0.014**israel" + 0.012**leigh" + 0.012**burk" + 0.011**mueller" + 0.011**nurseri"), (7, '0.055**brexit" + 0.045**farag" + 0.032**mcvey" + 0.030**gove" + 0.029**tori" + 0.026**parti" + 0.023**vote" + 0.021**eu" + 0.021**raab" + 0.020**mp"), (8, '0.003**email" + 0.003**trump" + 0.003**life" + 0.003**peopl" + 0.003**video" + 0.003**women" + 0.003**pictur" + 0.003**tell" + 0.003**children" + 0.003**ve"), (9, '0.004**imag" + 0.003**england" + 0.003**win" + 0.003**cup" + 0.003**club" + 0.003**leagu" + 0.003**uk" + 0.003**season" + 0.003**june" + 0.003**citi"), (10, '0.036**bst" + 0.030**inning" + 0.029**wicket" + 0.019**troop" + 0.015**thatcher" + 0.014**vessel" + 0.011**mileston" + 0.010**over" + 0.009**valeri" + 0.008**bundl"), (11, '0.031**subtitl" + 0.029**iplay" + 0.026**peni" + 0.018**librari" + 0.018**christi" + 0.013**insp" + 0.012**past" + 0.012**clare" + 0.011**viagra" + 0.011**basket"), (12, '0.061**iranian" + 0.027**reset" + 0.027**turnbul" + 0.024**an" + 0.020**davidson" + 0.015**lesbian" + 0.014**walford" + 0.013**homepag" + 0.010**ira" + 0.009**tracey"), (13, '0.069**arsen" + 0.044**sarri" + 0.032**chelsea" + 0.030**emerl" + 0.030**europa" + 0.026**gunner" + 0.018**maurizio" + 0.016**unai" + 0.015**bournemouth" + 0.014**hudson"), (14, '0.024**acid" + 0.013**erupt" + 0.011**muscl" + 0.011**tag" + 0.011**norman" + 0.011**script" + 0.011**dementia" + 0.010**riley" + 0.009**borough" + 0.009**gut"), (15, '0.009**credit" + 0.006**island" + 0.005**sun" + 0.005**holiday" + 0.005**rain" + 0.005**villa" + 0.005**shop" + 0.005**store" + 0.004**price" + 0.004**star"), (16, '0.022**samuel" + 0.017**klopp" + 0.016**pt" + 0.015**juventu" + 0.015**liverpool" + 0.015**midfield" + 0.014**tottenham" + 0.012**ligt" + 0.012**kane" + 0.012**var"), (17, '0.019**ballot" + 0.019**hancock" + 0.014**viral" + 0.013**chef" + 0.012**belgium" + 0.012**tank" + 0.012**sander" + 0.011**cameron" + 0.010**gulf" + 0.009**eric"), (18, '0.078**sri" + 0.078**lanka" + 0.029**fibr" + 0.026**archi" + 0.013**wolf" + 0.012**sullivan" + 0.011**slater" + 0.011**guitar" + 0.010**vacuum" + 0.010**tesla")]</p>		
18	<p>[(0, '0.020**sander" + 0.018**broadband" + 0.017**nicol" + 0.016**gogglebox" + 0.014**elton" + 0.014**everton" + 0.014**frontrunn" + 0.014**jenni" + 0.013**protein" + 0.012**valeri"), (1, '0.025**fibr" + 0.024**maya" + 0.018**bake" + 0.016**cha" + 0.015**leicestershir" + 0.015**dryer" + 0.013**xbox" + 0.013**louisa" + 0.012**flour" + 0.012**microsoft"), (2, '0.007**girl" + 0.007**death" + 0.006**murder" + 0.006**woman" + 0.006**hospit" + 0.006**daughter" + 0.006**health" + 0.006**die" + 0.005**mother" + 0.005**life"), (3, '0.020**korean" + 0.019**jen" + 0.018**harper" + 0.014**monthli" + 0.014**cheshir" + 0.014**theft" + 0.013**luther" + 0.012**hurrican" + 0.012**lavish" + 0.012**jong"), (4, '0.068**hull" + 0.056**password" + 0.042**exam" + 0.026**warrington" + 0.020**communist" + 0.016**breezi" + 0.015**sutton" + 0.011**iceland" + 0.010**workout" + 0.010**ofst"), (5, '0.016**mel" + 0.016**echo" + 0.013**navi" + 0.012**saudi" + 0.012**kyl" + 0.011**bruce" + 0.011**hitler" + 0.011**keith" + 0.010**radiat" + 0.010**isi"), (6, '0.012**ranger" + 0.012**normandi" + 0.011**joshua" + 0.011**holli" + 0.009**cliff" + 0.009**sandwich" + 0.008**saint" + 0.008**moli" + 0.008**japanes" + 0.008**middlesbrough"), (7, '0.031**vettel" + 0.024**emerl" + 0.021**hamilton" + 0.019**ligt" + 0.018**sharon" + 0.017**ferrari" + 0.017**thame" + 0.017**sec" + 0.016**jessi" + 0.014**pochettino"), (8, '0.068**stacey" + 0.043**lbc" + 0.020**paedophil" + 0.017**toast" + 0.012**hunger" + 0.011**bethani" + 0.010**boob" + 0.010**robby" + 0.009**morton" + 0.009**efl"), (9, '0.004**imag" + 0.003**trump" + 0.003**peopl" + 0.003**uk" + 0.003**polic" + 0.003**parti" + 0.003**brexit" + 0.002**news" + 0.002**work" + 0.002**servic"), (10, '0.075**bst" + 0.034**klopp" + 0.021**geri" + 0.020**salah" + 0.018**satellit" + 0.017**robertson" + 0.013**kohli" + 0.010**chicago" + 0.010**horner" + 0.009**ballet"), (11, '0.034**acid" + 0.027**freddi" + 0.017**jimmi" + 0.017**dylan" + 0.017**blackburn" + 0.014**librari" + 0.014**cruel" + 0.013**dom" + 0.011**fletcher" + 0.011**sitcom"), (12, '0.037**mcvey" + 0.018**leadsom" + 0.018**brexit" + 0.017**dem" + 0.016**lib" + 0.016**thatcher" + 0.015**backstop" + 0.013**knowledg" + 0.013**membership" + 0.012**shine"), (13, '0.010**music" + 0.010**film" + 0.007**episod" + 0.007**china" + 0.006**cook" + 0.005**song" + 0.005**guest" + 0.005**india" + 0.005**food" + 0.005**plant"), (14, '0.041**neymar" + 0.028**archi" + 0.017**inmat" + 0.015**pastri" + 0.015**rochdal" + 0.012**dioxid" + 0.010**skull" + 0.010**esa" + 0.007**hydrogen" + 0.007**claudia"), (15, '0.010**babi" + 0.009**gove" + 0.008**dress" + 0.008**hair" + 0.008**lister" + 0.008**instagram" + 0.007**credit" + 0.007**danc" + 0.007**snap" + 0.006**pakistan"), (16, '0.031**badg" + 0.021**vaccin" + 0.020**succeed" + 0.015**lincoln" + 0.014**successor" + 0.013**plymouth" + 0.013**autism" + 0.012**backgrid" + 0.012**harbour" +</p>	0.43398322317 08188	-11.27003727941 7332

	0.012**raf"), (17, '0.004**win" + 0.004**play" + 0.004**england" + 0.004**leagu" + 0.003**club" + 0.003**game" + 0.003**cup" + 0.003**final" + 0.003**season" + 0.003**johnson")]		
17	[(0, '0.042**hull" + 0.019**sunni" + 0.017**kyli" + 0.017**dwp" + 0.016**promo" + 0.015**maya" + 0.014**sainsburi" + 0.014**currenc" + 0.014**turnbul" + 0.014**christin"), (1, '0.035**iranian" + 0.031**samuel" + 0.017**gregg" + 0.014**extradit" + 0.013**leicestershir" + 0.012**bean" + 0.011**cemeteri" + 0.011**pearson" + 0.011**poland" + 0.010**huey"), (2, '0.005**trump" + 0.004**polic" + 0.004**england" + 0.003**peopl" + 0.003**court" + 0.003**support" + 0.003**presid" + 0.003**leagu" + 0.003**women" + 0.003**offic"), (3, '0.011**inning" + 0.011**wicket" + 0.010**sri" + 0.010**lanka" + 0.009**wood" + 0.009**stacey" + 0.008**password" + 0.008**cocain" + 0.008**russian" + 0.008**sweet"), (4, '0.032**holli" + 0.023**walsal" + 0.023**sharon" + 0.023**warrington" + 0.020**fractur" + 0.014**riot" + 0.013**sutton" + 0.013**hungari" + 0.012**skull" + 0.011**unpaid"), (5, '0.032**netanyahu" + 0.021**tehran" + 0.021**israel" + 0.020**fifa" + 0.017**dom" + 0.016**gaza" + 0.015**playstat" + 0.013**palestinian" + 0.013**isra" + 0.012**arabia"), (6, '0.029**nadal" + 0.021**klopp" + 0.020**solskjaer" + 0.018**albion" + 0.018**inspector" + 0.017**roland" + 0.015**pd" + 0.014**thiem" + 0.014**jenni" + 0.013**librari"), (7, '0.054**badg" + 0.037**nevil" + 0.030**warner" + 0.025**iplay" + 0.023**eurovis" + 0.014**fletcher" + 0.014**lgbtq" + 0.012**barbara" + 0.011**dragon" + 0.011**gaga"), (8, '0.029**vaccin" + 0.025**subtitl" + 0.022**eamonn" + 0.019**holm" + 0.016**ree" + 0.016**vicki" + 0.015**inmat" + 0.015**dani" + 0.014**countdown" + 0.014**nurseni"), (9, '0.003**imag" + 0.003**play" + 0.003**day" + 0.003**bbc" + 0.003**credit" + 0.003**subscrib" + 0.003**win" + 0.002**news" + 0.002**email" + 0.002**live"), (10, '0.020**loan" + 0.012**coffe" + 0.011**cinema" + 0.011**eastend" + 0.010**tank" + 0.010**sander" + 0.010**wealth" + 0.009**adrian" + 0.009**cruis" + 0.008**nicol"), (11, '0.013**polic" + 0.013**iran" + 0.012**stab" + 0.010**pizza" + 0.010**arrest" + 0.008**attack" + 0.007**nazi" + 0.006**pier" + 0.006**spice" + 0.006**murder"), (12, '0.045**brexit" + 0.030**parti" + 0.027**vote" + 0.024**tori" + 0.024**mp" + 0.024**eu" + 0.022**johnson" + 0.020**elect" + 0.020**bori" + 0.019**labour"), (13, '0.037**arsen" + 0.031**bst" + 0.026**midfield" + 0.021**leagu" + 0.021**brazil" + 0.021**chelsea" + 0.018**striker" + 0.017**sarri" + 0.016**goalkeep" + 0.014**europa"), (14, '0.028**mel" + 0.023**bikini" + 0.019**shepherd" + 0.016**geri" + 0.015**demi" + 0.013**tin" + 0.011**chaser" + 0.011**carriag" + 0.011**ofcom" + 0.010**debenham"), (15, '0.005**pay" + 0.005**compani" + 0.005**busi" + 0.004**offer" + 0.004**servic" + 0.004**deal" + 0.004**plan" + 0.004**cost" + 0.004**pension" + 0.004**bank"), (16, '0.031**raab" + 0.030**melania" + 0.029**acid" + 0.023**normandi" + 0.018**soviet" + 0.013**protein" + 0.013**liquid" + 0.012**craft" + 0.011**tokyo" + 0.010**reev")]	0.48364511157 262946	-11.10791272249 7504
16	[(0, '0.007**women" + 0.005**health" + 0.004**patient" + 0.003**nh" + 0.003**doctor" + 0.003**dr" + 0.003**cancer" + 0.003**medic" + 0.003**diseas" + 0.003**blood"), (1, '0.029**mep" + 0.016**peni" + 0.016**fianc" + 0.015**elton" + 0.013**glow" + 0.013**chicago" + 0.011**bloc" + 0.011**buzzfe" + 0.011**salah" + 0.010**ax"), (2, '0.031**iran" + 0.011**normandi" + 0.011**vaccin" + 0.009**banquet" + 0.008**gulf" + 0.008**licens" + 0.007**kerri" + 0.007**burden" + 0.007**natali" + 0.007**radiat"), (3, '0.004**play" + 0.004**win" + 0.003**credit" + 0.003**cup" + 0.003**day" + 0.003**love" + 0.003**star" + 0.003**game" + 0.003**final" + 0.003**bbc"), (4, '0.035**indi" + 0.033**raab" + 0.033**melania" + 0.030**spice" + 0.026**mel" + 0.020**mm" + 0.020**pt" + 0.017**grenfel" + 0.016**breezi" + 0.016**blackburn"), (5, '0.037**stacey" + 0.019**kyli" + 0.017**walsal" + 0.017**pd" + 0.014**christin" + 0.014**gossip" + 0.014**pyramid" + 0.013**burk" + 0.011**scotsman" + 0.011**dryer"), (6, '0.025**inning" + 0.024**wicket" + 0.021**sri" + 0.021**lanka" + 0.020**badg" + 0.012**sander" + 0.011**soviet" + 0.011**bowler" + 0.010**consent" + 0.010**neymar"), (7, '0.010**polic" + 0.006**imag" + 0.005**court" + 0.005**offic" + 0.005**car" + 0.005**road" + 0.005**london" + 0.004**man" + 0.004**west" + 0.004**south"), (8, '0.031**arsen" + 0.025**chelsea" + 0.024**leagu" + 0.015**madrid" + 0.015**acid" + 0.014**striker" + 0.013**sarri" + 0.013**tottenham" + 0.013**goalkeep" + 0.011**europa"), (9, '0.036**nadal" + 0.030**vettel" + 0.026**nevil" + 0.021**hamilton" + 0.020**djokov" + 0.018**thiem" + 0.016**sec" + 0.016**ferrari" + 0.015**archi" + 0.014**rafael"), (10, '0.035**maya" + 0.023**warrington" + 0.016**leicestershir" + 0.016**jacob" + 0.015**cha" + 0.014**huey" + 0.013**playstat" + 0.012**louisa" + 0.011**gallagh" + 0.011**northbound"), (11, '0.014**cricket" + 0.011**javid" + 0.009**nuclear" + 0.008**buckingham" + 0.008**rachel" + 0.008**root" + 0.008**venu" + 0.008**stab" + 0.007**duke" + 0.007**coffe"), (12, '0.036**brexit" + 0.030**parti" + 0.024**johnson" + 0.023**vote" + 0.021**mp" + 0.019**tori" + 0.019**elect" + 0.019**eu" + 0.017**minist" + 0.017**labour"), (13, '0.005**trump" + 0.005**uk" + 0.003**pay" + 0.003**presid" + 0.003**govern" + 0.003**plan" +	0.49126252111 29416	-10.92552323381 1269

	0.003**"countri" + 0.003**"offer" + 0.003**"free" + 0.003**"deal"), (14, '0.026**"bgt" + 0.026**"jen" + 0.025**"netanyahu" + 0.018**"promo" + 0.016**"israel" + 0.015**"reev" + 0.013**"ryan" + 0.013**"fletcher" + 0.012**"holder" + 0.012**"gaza"), (15, '0.036**"exam" + 0.036**"hancock" + 0.027**"korean" + 0.026**"dwp" + 0.022**"subtitl" + 0.022**"sharon" + 0.020**"reset" + 0.013**"tristan" + 0.012**"cheryl" + 0.010**"uber"))]		
15	[(0, '0.010**"password" + 0.008**"print" + 0.008**"expir" + 0.008**"coffe" + 0.008**"virgin" + 0.008**"cod" + 0.008**"tariff" + 0.007**"zone" + 0.007**"bat" + 0.007**"belgium"), (1, '0.028**"iranian" + 0.022**"outbreak" + 0.017**"carbon" + 0.015**"fibr" + 0.014**"pd" + 0.014**"radiat" + 0.013**"cheshir" + 0.010**"gibson" + 0.010**"boob" + 0.009**"gigg"), (2, '0.024**"hancock" + 0.020**"samuel" + 0.018**"cardiff" + 0.014**"welsh" + 0.013**"collis" + 0.012**"fractur" + 0.011**"middlesbrough" + 0.011**"spi" + 0.011**"madam" + 0.010**"halifax"), (3, '0.025**"vettel" + 0.019**"anita" + 0.018**"neymar" + 0.017**"roland" + 0.016**"hamilton" + 0.014**"backgrid" + 0.013**"sec" + 0.013**"ferrari" + 0.013**"kri" + 0.011**"dani"), (4, '0.051**"brexit" + 0.030**"tori" + 0.030**"farag" + 0.024**"vote" + 0.024**"parti" + 0.021**"mcvey" + 0.021**"eu" + 0.021**"mp" + 0.018**"labour" + 0.016**"conserv"), (5, '0.011**"voucher" + 0.010**"vaccin" + 0.010**"warner" + 0.008**"scientif" + 0.008**"iplay" + 0.008**"discoveri" + 0.008**"mi" + 0.008**"kyli" + 0.007**"bruce" + 0.007**"plung"), (6, '0.005**"pay" + 0.005**"school" + 0.005**"servic" + 0.005**"govern" + 0.005**"compani" + 0.005**"licenc" + 0.005**"busi" + 0.004**"council" + 0.004**"cost" + 0.004**"pension"), (7, '0.021**"symptom" + 0.018**"sarri" + 0.014**"maya" + 0.014**"inquest" + 0.012**"harper" + 0.011**"molli" + 0.011**"thailand" + 0.011**"chloe" + 0.010**"kerri" + 0.010**"kennedi"), (8, '0.004**"imag" + 0.003**"polic" + 0.003**"bbc" + 0.003**"trump" + 0.003**"peopl" + 0.002**"women" + 0.002**"time" + 0.002**"parti" + 0.002**"london" + 0.002**"johnson"), (9, '0.039**"nadal" + 0.029**"troop" + 0.029**"nevil" + 0.023**"tran" + 0.021**"djokov" + 0.020**"sharon" + 0.019**"thiem" + 0.015**"rafael" + 0.015**"bercow" + 0.015**"reopen"), (10, '0.015**"leagu" + 0.014**"club" + 0.014**"cup" + 0.014**"season" + 0.014**"game" + 0.012**"player" + 0.011**"final" + 0.011**"win" + 0.010**"match" + 0.010**"england"), (11, '0.030**"inning" + 0.025**"lanka" + 0.024**"indi" + 0.020**"sri" + 0.014**"klopp" + 0.014**"windi" + 0.014**"solskjaer" + 0.013**"ronaldo" + 0.013**"wicket" + 0.011**"rey"), (12, '0.005**"subscrib" + 0.005**"credit" + 0.004**"love" + 0.004**"email" + 0.004**"star" + 0.003**"babi" + 0.003**"hospit" + 0.003**"itv" + 0.003**"mum" + 0.003**"ve"), (13, '0.033**"mel" + 0.025**"meghan" + 0.021**"walsal" + 0.016**"jenni" + 0.015**"tyson" + 0.015**"geri" + 0.013**"biden" + 0.013**"dijk" + 0.012**"tristan" + 0.012**"georgia"), (14, '0.028**"archi" + 0.028**"porn" + 0.013**"spider" + 0.013**"insp" + 0.013**"gregori" + 0.012**"mcilroy" + 0.012**"cheryl" + 0.012**"ch" + 0.010**"gaga" + 0.010**"lopez"))]	0.47798775854 41481	-10.79903708220 2519
14	[(0, '0.038**"cannabi" + 0.024**"broadband" + 0.021**"nicol" + 0.018**"halifax" + 0.015**"opera" + 0.014**"inmat" + 0.012**"tristan" + 0.012**"norfolk" + 0.010**"heroin" + 0.010**"northbound"), (1, '0.016**"ghost" + 0.016**"inspector" + 0.012**"invas" + 0.011**"georgia" + 0.011**"copa" + 0.010**"belgian" + 0.010**"heavyweight" + 0.010**"poland" + 0.010**"bale" + 0.009**"cocktail"), (2, '0.008**"stab" + 0.006**"leed" + 0.006**"sam" + 0.006**"session" + 0.006**"nuclear" + 0.006**"ami" + 0.006**"road" + 0.005**"mari" + 0.005**"jami" + 0.005**"unexpect"), (3, '0.013**"badg" + 0.012**"acid" + 0.010**"strictli" + 0.009**"vettel" + 0.008**"emmerdal" + 0.008**"stacey" + 0.008**"sugar" + 0.007**"recip" + 0.007**"scan" + 0.007**"ferrari"), (4, '0.042**"brexit" + 0.029**"parti" + 0.026**"vote" + 0.023**"mp" + 0.023**"tori" + 0.022**"eu" + 0.021**"elect" + 0.019**"labour" + 0.018**"leadership" + 0.017**"conserv"), (5, '0.006**"win" + 0.005**"england" + 0.005**"play" + 0.005**"leagu" + 0.005**"club" + 0.005**"game" + 0.005**"cup" + 0.005**"season" + 0.005**"final" + 0.005**"team"), (6, '0.003**"imag" + 0.002**"polic" + 0.002**"peopl" + 0.002**"bbc" + 0.002**"uk" + 0.002**"trump" + 0.002**"work" + 0.002**"day" + 0.002**"credit" + 0.002**"news"), (7, '0.020**"bradford" + 0.020**"meghan" + 0.014**"usa" + 0.013**"duchess" + 0.011**"exam" + 0.010**"vaccin" + 0.009**"princ" + 0.009**"duke" + 0.009**"harass" + 0.009**"arcadia"), (8, '0.030**"lbc" + 0.020**"jenner" + 0.020**"jessi" + 0.019**"carol" + 0.018**"communist" + 0.016**"sasha" + 0.015**"gossip" + 0.014**"paedophii" + 0.014**"romeo" + 0.012**"countdown"), (9, '0.020**"dwp" + 0.018**"gulfi" + 0.016**"bournemouth" + 0.014**"chicago" + 0.014**"newsbeat" + 0.013**"xtra" + 0.012**"ebay" + 0.011**"xbbox" + 0.010**"solid" + 0.010**"grime"), (10, '0.023**"joshua" + 0.019**"archi" + 0.016**"warrington" + 0.015**"walsal" + 0.012**"valeri" + 0.012**"guitar" + 0.012**"ree" + 0.012**"sutton" + 0.011**"bean" + 0.011**"clearanc"), (11, '0.034**"anita" + 0.025**"sandwich" + 0.024**"turnbul" + 0.016**"radiat" + 0.015**"dani" + 0.012**"worcest" + 0.012**"librari" + 0.011**"pastri" + 0.009**"pompeo" + 0.007**"zuckerberg"), (12, '0.015**"bst" + 0.009**"hancock" + 0.008**"helicopt" + 0.008**"podcast" + 0.007**"museum" + 0.006**"aircraft" + 0.006**"tension" + 0.006**"compens" + 0.006**"wildlif" + 0.006**"hampshir"), (13, '0.031**"maya" + 0.021**"carbon" + 0.020**"pd" + 0.020**"fibr" + 0.018**"ant" + 0.016**"emerii" + 0.013**"cha" + 0.012**"ozil" + 0.012**"arsen" + 0.012**"unai"))]	0.45724783472 835273	-10.65087174552 4869

13	<p>[(0, '0.003**peopl" + 0.003**trump" + 0.003**imag" + 0.002**uk" + 0.002**women" + 0.002**england" + 0.002**bbc" + 0.002**work" + 0.002**support" + 0.002**getti"), (1, '0.006**hell" + 0.006**nh" + 0.006**coffe" + 0.006**virgin" + 0.005**netflix" + 0.005**clinic" + 0.005**sander" + 0.005**mm" + 0.005**pollut" + 0.005**trend"), (2, '0.039**inning" + 0.036**wicket" + 0.033**sri" + 0.033**lanka" + 0.031**badg" + 0.018**windi" + 0.014**breezi" + 0.013**over" + 0.012**everton" + 0.011**gu"), (3, '0.029**klopp" + 0.022**gogglebox" + 0.022**ligt" + 0.019**ant" + 0.017**pochettino" + 0.014**paedophil" + 0.012**robber" + 0.011**griezmann" + 0.011**sutton" + 0.010**lukaku"), (4, '0.034**leagu" + 0.024**arsen" + 0.019**chelsea" + 0.016**midfield" + 0.015**madrid" + 0.013**tottenham" + 0.012**liverpool" + 0.011**password" + 0.010**goal" + 0.010**sarri"), (5, '0.017**abort" + 0.013**acid" + 0.012**discount" + 0.012**huawei" + 0.011**voucher" + 0.010**outbreak" + 0.010**vaccin" + 0.009**salari" + 0.009**runner" + 0.009**bristol"), (6, '0.005**polic" + 0.004**subscrib" + 0.003**email" + 0.003**man" + 0.003**play" + 0.003**credit" + 0.003**news" + 0.003**imag" + 0.003**car" + 0.003**road"), (7, '0.027**brexit" + 0.025**parti" + 0.020**johnson" + 0.017**vote" + 0.016**mp" + 0.015**minist" + 0.014**tori" + 0.014**elect" + 0.014**eu" + 0.014**deal"), (8, '0.023**hancock" + 0.023**melania" + 0.014**golf" + 0.013**robot" + 0.011**fifa" + 0.010**dementia" + 0.009**hurrican" + 0.009**valeri" + 0.009**loyalti" + 0.008**ree"), (9, '0.022**iranian" + 0.015**tank" + 0.012**tehran" + 0.012**netanyahu" + 0.012**communist" + 0.011**lesbian" + 0.011**middlesbrough" + 0.011**pd" + 0.010**asylum" + 0.010**israel"), (10, '0.024**lincolns" + 0.017**jen" + 0.015**warrington" + 0.015**gulf" + 0.014**mark" + 0.014**meghan" + 0.013**gregg" + 0.013**eurovis" + 0.012**denis" + 0.012**boe"), (11, '0.018**rape" + 0.012**inspector" + 0.011**archi" + 0.011**anita" + 0.010**saudi" + 0.010**dwp" + 0.010**alami" + 0.008**backgrid" + 0.008**sacrific" + 0.008**borough"), (12, '0.009**star" + 0.008**love" + 0.007**music" + 0.006**itv" + 0.006**episod" + 0.005**fan" + 0.004**dress" + 0.004**instagram" + 0.004**film" + 0.004**actor")]</p>	0.46285955437 55662	-10.42095184291 4306
12	<p>[(0, '0.021**brexit" + 0.019**parti" + 0.018**trump" + 0.013**vote" + 0.012**mp" + 0.012**deal" + 0.012**minist" + 0.011**tori" + 0.011**elect" + 0.011**eu"), (1, '0.017**leagu" + 0.016**madrid" + 0.015**liverpool" + 0.014**acid" + 0.013**chelsea" + 0.012**tottenham" + 0.011**barcelona" + 0.010**juventu" + 0.010**aston" + 0.010**midfield"), (2, '0.022**archi" + 0.021**kyli" + 0.018**fifa" + 0.016**eamonn" + 0.016**gu" + 0.015**christin" + 0.011**playstat" + 0.011**miley" + 0.010**ill" + 0.010**loyalti"), (3, '0.020**password" + 0.018**nadal" + 0.015**vettel" + 0.010**gulf" + 0.010**garro" + 0.010**sharon" + 0.010**djokov" + 0.009**keith" + 0.009**norri" + 0.009**norman"), (4, '0.024**inning" + 0.021**abort" + 0.020**sri" + 0.020**lanka" + 0.018**hancock" + 0.016**wicket" + 0.014**expir" + 0.014**normandi" + 0.011**banquet" + 0.008**wage"), (5, '0.024**arsen" + 0.020**bst" + 0.011**sarri" + 0.009**swansea" + 0.009**ham" + 0.008**lbc" + 0.008**belgium" + 0.008**holder" + 0.007**emer" + 0.007**erupt"), (6, '0.003**imag" + 0.002**polic" + 0.002**bbc" + 0.002**peopl" + 0.002**day" + 0.002**win" + 0.002**uk" + 0.002**time" + 0.002**play" + 0.002**england"), (7, '0.008**lgbt" + 0.007**kong" + 0.007**phone" + 0.007**hong" + 0.006**paint" + 0.006**swim" + 0.006**freddi" + 0.006**nurs" + 0.006**metro" + 0.006**joshua"), (8, '0.019**stacey" + 0.011**bgt" + 0.010**chloe" + 0.010**dwp" + 0.009**jenner" + 0.009**madam" + 0.009**subtitl" + 0.009**bolton" + 0.008**jamaica" + 0.008**haunt"), (9, '0.006**china" + 0.006**present" + 0.005**august" + 0.005**nh" + 0.005**sentenc" + 0.005**chelsea" + 0.005**tournament" + 0.005**plastic" + 0.004**guest" + 0.004**india"), (10, '0.062**iran" + 0.026**meghan" + 0.016**sandwich" + 0.016**netanyahu" + 0.013**elton" + 0.013**backgrid" + 0.011**valeri" + 0.011**leicestershir" + 0.011**fisher" + 0.010**guitar"), (11, '0.017**commemor" + 0.012**inquest" + 0.011**inspector" + 0.010**cctv" + 0.009**chemic" + 0.008**rapper" + 0.008**weird" + 0.008**jessi" + 0.007**det" + 0.007**riley")]</p>	0.44447014450 05274	-10.20695438130 0992
11	<p>[(0, '0.004**trump" + 0.003**uk" + 0.003**peopl" + 0.003**imag" + 0.003**support" + 0.003**england" + 0.003**deal" + 0.003**govern" + 0.003**pay" + 0.003**presid"), (1, '0.027**arcadia" + 0.018**jenner" + 0.017**backgrid" + 0.013**khloe" + 0.013**cva" + 0.011**climber" + 0.011**fletcher" + 0.009**northbound" + 0.009**creditor" + 0.008**debenham"), (2, '0.014**polic" + 0.008**attack" + 0.008**man" + 0.007**court" + 0.007**offic" + 0.007**car" + 0.007**road" + 0.006**hospit" + 0.006**west" + 0.006**murder"), (3, '0.003**water" + 0.003**food" + 0.003**health" + 0.003**royal" + 0.002**minut" + 0.002**iran" + 0.002**day" + 0.002**natur" + 0.002**babi" + 0.002**product"), (4, '0.044**brexit" + 0.029**parti" + 0.025**vote" + 0.024**tori" + 0.022**mp" + 0.022**eu" + 0.019**elect" + 0.018**labour"</p>	0.53198989248 07671	-9.847306801658 492

	+ 0.016**"leadership" + 0.016**"conserv"), (5, '0.025**"leagu" + 0.019**"player" + 0.018**"club" + 0.016**"season" + 0.015**"liverpool" + 0.014**"goal" + 0.013**"chelsea" + 0.012**"cup" + 0.012**"champion" + 0.011**"score"), (6, '0.028**"inning" + 0.025**"wicket" + 0.024**"sri" + 0.023**"lanka" + 0.018**"lincolnshir" + 0.012**"warner" + 0.012**"bowler" + 0.012**"anita" + 0.012**"bgt" + 0.010**"sharon"), (7, '0.015**"stacey" + 0.013**"emmerdal" + 0.012**"freddi" + 0.011**"vaccin" + 0.011**"maya" + 0.009**"stuart" + 0.008**"nicol" + 0.008**"homophob" + 0.007**"turnbul" + 0.007**"satellit"), (8, '0.027**"bst" + 0.012**"swansea" + 0.011**"hampshir" + 0.010**"konta" + 0.009**"johan" + 0.008**"blackburn" + 0.008**"suffolk" + 0.008**"ligt" + 0.008**"var" + 0.008**"kane"), (9, '0.004**"play" + 0.004**"love" + 0.004**"star" + 0.003**"subscrib" + 0.003**"win" + 0.003**"fan" + 0.003**"pictur" + 0.003**"bbc" + 0.003**"stori" + 0.003**"life"), (10, '0.021**"badg" + 0.015**"normandi" + 0.015**"vettel" + 0.010**"hamilton" + 0.009**"norri" + 0.008**"ferrari" + 0.008**"raf" + 0.008**"sec" + 0.008**"wan" + 0.007**"bissaka")]		
10	[(0, '0.026**"vettel" + 0.019**"hamilton" + 0.017**"breezi" + 0.015**"ferrari" + 0.014**"sec" + 0.011**"merced" + 0.008**"shower" + 0.008**"bpm" + 0.008**"lap" + 0.007**"prix"), (1, '0.003**"bbc" + 0.002**"day" + 0.002**"time" + 0.002**"women" + 0.002**"peopl" + 0.002**"credit" + 0.002**"play" + 0.002**"live" + 0.002**"work" + 0.002**"love"), (2, '0.012**"brexit" + 0.011**"parti" + 0.011**"trump" + 0.009**"johnson" + 0.009**"deal" + 0.008**"vote" + 0.007**"mp" + 0.007**"govern" + 0.007**"minist" + 0.007**"presid"), (3, '0.007**"polic" + 0.005**"imag" + 0.004**"man" + 0.004**"servic" + 0.004**"offic" + 0.004**"attack" + 0.004**"london" + 0.004**"car" + 0.003**"report" + 0.003**"south"), (4, '0.012**"pension" + 0.010**"bank" + 0.009**"iran" + 0.008**"tax" + 0.008**"gove" + 0.007**"inning" + 0.006**"invest" + 0.006**"wicket" + 0.006**"payment" + 0.006**"financi"), (5, '0.029**"froom" + 0.012**"dryer" + 0.012**"pebbl" + 0.011**"mcilroy" + 0.010**"waitros" + 0.009**"koepka" + 0.009**"ineo" + 0.009**"ps" + 0.009**"putt" + 0.008**"birdi"), (6, '0.019**"leagu" + 0.017**"cup" + 0.017**"club" + 0.016**"player" + 0.014**"season" + 0.012**"liverpool" + 0.011**"goal" + 0.010**"champion" + 0.010**"game" + 0.010**"chelsea"), (7, '0.011**"store" + 0.010**"plastic" + 0.008**"amazon" + 0.008**"parkinson" + 0.007**"code" + 0.006**"retail" + 0.006**"infect" + 0.006**"symptom" + 0.006**"acid" + 0.005**"cocain"), (8, '0.012**"court" + 0.008**"abus" + 0.008**"judg" + 0.007**"sexual" + 0.007**"sentenc" + 0.007**"prison" + 0.006**"jail" + 0.006**"alleg" + 0.006**"trial" + 0.005**"kati"), (9, '0.030**"meghan" + 0.014**"duchess" + 0.013**"pd" + 0.013**"princ" + 0.011**"opera" + 0.011**"harri" + 0.010**"markl" + 0.009**"childlin" + 0.009**"royal" + 0.009**"duke")]	0.49810749794 853215	-9.455347063936 431
9	[(0, '0.004**"imag" + 0.003**"peopl" + 0.002**"women" + 0.002**"news" + 0.002**"day" + 0.002**"email" + 0.002**"subscrib" + 0.002**"tell" + 0.002**"live" + 0.002**"famili"), (1, '0.019**"pd" + 0.019**"gogglebox" + 0.016**"sec" + 0.007**"endgam" + 0.006**"cen" + 0.006**"infin" + 0.005**"ita" + 0.005**"aveng" + 0.005**"bingley" + 0.005**"movistar"), (2, '0.026**"brexit" + 0.024**"parti" + 0.018**"johnson" + 0.017**"vote" + 0.016**"mp" + 0.015**"minist" + 0.014**"tori" + 0.014**"elect" + 0.014**"eu" + 0.013**"labour"), (3, '0.007**"star" + 0.006**"love" + 0.005**"music" + 0.005**"babi" + 0.004**"film" + 0.004**"instagram" + 0.004**"itv" + 0.004**"credit" + 0.004**"queen" + 0.004**"royal"), (4, '0.034**"nadal" + 0.019**"garro" + 0.018**"djokov" + 0.017**"thiem" + 0.010**"rafael" + 0.009**"mcilroy" + 0.008**"koepka" + 0.008**"vondrousova" + 0.008**"putt" + 0.007**"birdi"), (5, '0.008**"leagu" + 0.008**"win" + 0.007**"club" + 0.007**"england" + 0.007**"cup" + 0.007**"game" + 0.007**"season" + 0.007**"final" + 0.006**"play" + 0.006**"player"), (6, '0.003**"bbc" + 0.002**"open" + 0.002**"london" + 0.002**"build" + 0.002**"plan" + 0.002**"seri" + 0.002**"busi" + 0.002**"pension" + 0.002**"compani" + 0.002**"air"), (7, '0.013**"emmerdal" + 0.011**"maya" + 0.010**"windi" + 0.010**"pc" + 0.009**"ant" + 0.009**"breezi" + 0.007**"christin" + 0.007**"raf" + 0.007**"gu" + 0.006**"clare"), (8, '0.019**"meri" + 0.018**"bgt" + 0.014**"subtitl" + 0.014**"jessi" + 0.014**"keanu" + 0.012**"munro" + 0.010**"magician" + 0.010**"unai" + 0.009**"rylan" + 0.009**"hollyoak")]	0.48743052120 495683	-9.205309739953 213
8	[(0, '0.013**"khan" + 0.008**"robinson" + 0.008**"custodi" + 0.008**"normandi" + 0.007**"bailey" + 0.006**"carbon" + 0.006**"magistr" + 0.005**"broadband" + 0.005**"sentenc" + 0.005**"sadiq"), (1, '0.008**"royal" + 0.007**"harri" + 0.006**"princ" + 0.006**"babi" + 0.005**"palac" + 0.005**"queen" + 0.005**"meghan" + 0.004**"birth" + 0.004**"pakistan" + 0.004**"brazil"), (2, '0.015**"afghanistan" + 0.013**"brigad" + 0.008**"isi" + 0.008**"bpm" + 0.008**"margat" + 0.007**"wwe" + 0.006**"arson" + 0.006**"infirmari" + 0.006**"superintend" + 0.006**"chelmsford"), (3, '0.005**"food" + 0.004**"nh" + 0.004**"cook" + 0.004**"patient" + 0.004**"eat" + 0.003**"cancer" + 0.003**"diseas" + 0.003**"water" + 0.003**"plastic" + 0.003**"blood"), (4, '0.044**"brexit" + 0.028**"parti" + 0.026**"tori" + 0.024**"mp" + 0.022**"vote" + 0.021**"eu" + 0.019**"elect" + 0.019**"farag" + 0.018**"johnson" + 0.018**"bori"), (5, '0.028**"leagu" + 0.015**"liverpool" + 0.015**"arsen" + 0.013**"chelsea" + 0.012**"goal" + 0.012**"player" + 0.011**"club" + 0.011**"madrid" + 0.010**"midfield" + 0.010**"champion"), (6, '0.002**"imag" + 0.002**"polic" + 0.002**"bbc" +	0.55543759060 99775	-9.141746589925 553

	0.002**"peopl" + 0.002**"uk" + 0.002**"trump" + 0.002**"day" + 0.001**"time" + 0.001**"win" + 0.001**"play"), (7, '0.011**"lister" + 0.010**"kati" + 0.009**"sherif" + 0.008**"stacey" + 0.007**"emmerdal" + 0.007**"violin" + 0.007**"lorrain" + 0.006**"soap" + 0.006**"holli" + 0.006**"maya"]]		
--	--	--	--

Evaluation of different Ida models without tf-idf incorporated :

Number of topics	Topic keywords	coherence	Perplexity
20	[(0, '0.013**"pay" + 0.011**"compani" + 0.009**"uk" + 0.008**"cost" + 0.008**"busi" + 0.008**"free" + 0.007**"custom" + 0.007**"servic" + 0.006**"pension" + 0.006**"month"), (1, '0.016**"health" + 0.012**"peopl" + 0.009**"patient" + 0.007**"medic" + 0.007**"nh" + 0.007**"doctor" + 0.007**"hospit" + 0.007**"cancer" + 0.007**"diseas" + 0.006**"studi"), (2, '0.010**"south" + 0.009**"day" + 0.008**"west" + 0.008**"citi" + 0.007**"water" + 0.007**"area" + 0.007**"north" + 0.007**"imag" + 0.007**"travel" + 0.007**"east"), (3, '0.020**"peopl" + 0.014**"women" + 0.011**"group" + 0.009**"student" + 0.008**"twitter" + 0.008**"social" + 0.008**"protest" + 0.007**"countri" + 0.006**"univers" + 0.006**"facebook"), (4, '0.026**"win" + 0.013**"england" + 0.013**"open" + 0.012**"time" + 0.012**"final" + 0.011**"cup" + 0.011**"team" + 0.011**"match" + 0.010**"play" + 0.010**"race"), (5, '0.022**"leagu" + 0.015**"player" + 0.014**"club" + 0.012**"win" + 0.012**"season" + 0.012**"liverpool" + 0.012**"game" + 0.011**"unit" + 0.010**"play" + 0.010**"final"), (6, '0.011**"work" + 0.011**"servic" + 0.008**"imag" + 0.008**"council" + 0.008**"peopl" + 0.008**"bbc" + 0.007**"govern" + 0.007**"public" + 0.007**"decis" + 0.007**"plan"), (7, '0.025**"credit" + 0.014**"love" + 0.011**"itv" + 0.011**"star" + 0.010**"pictur" + 0.009**"sun" + 0.009**"reveal" + 0.008**"tv" + 0.008**"fan" + 0.007**"island"), (8, '0.014**"wear" + 0.012**"black" + 0.012**"hair" + 0.012**"dress" + 0.010**"star" + 0.010**"white" + 0.008**"pair" + 0.007**"look" + 0.007**"model" + 0.006**"red"), (9, '0.018**"bbc" + 0.012**"imag" + 0.011**"music" + 0.010**"royal" + 0.010**"film" + 0.010**"queen" + 0.008**"princ" + 0.008**"star" + 0.008**"perform" + 0.008**"song"), (10, '0.015**"time" + 0.014**"peopl" + 0.012**"children" + 0.011**"don" + 0.011**"day" + 0.011**"feel" + 0.010**"famili" + 0.010**"work" + 0.010**"life" + 0.010**"ve"), (11, '0.060**"polic" + 0.029**"attack" + 0.020**"man" + 0.019**"offic" + 0.013**"arrest" + 0.012**"incid" + 0.011**"investig" + 0.010**"crime" + 0.010**"victim" + 0.009**"london"), (12, '0.017**"food" + 0.010**"cook" + 0.008**"eat" + 0.007**"water" + 0.007**"restaur" + 0.007**"drink" + 0.006**"buy" + 0.006**"good" + 0.006**"garden" + 0.005**"meat"), (13, '0.010**"season" + 0.009**"episod" + 0.009**"seri" + 0.008**"play" + 0.007**"charact" + 0.006**"time" + 0.006**"game" + 0.005**"stori" + 0.005**"film" + 0.005**"book"), (14, '0.016**"court" + 0.015**"die" + 0.014**"death" + 0.013**"famili" + 0.012**"tell" + 0.010**"hear" + 0.008**"murder" + 0.008**"prison" + 0.007**"sentenc" + 0.007**"abus"), (15, '0.019**"climat" + 0.015**"imag" + 0.013**"hong" + 0.013**"kong" + 0.013**"plastic" + 0.012**"protes" + 0.012**"plant" + 0.012**"china" + 0.012**"chang" + 0.010**"india"), (16, '0.062**"getti" + 0.035**"pa" + 0.033**"imag" + 0.027**"septemb" + 0.026**"august" + 0.022**"afp" + 0.022**"reuter" + 0.022**"london" + 0.015**"day" + 0.010**"protest"), (17, '0.028**"brexit" + 0.026**"parti" + 0.017**"vote" + 0.015**"deal" + 0.015**"mp" + 0.013**"johnson" + 0.013**"minist" + 0.013**"elect" + 0.013**"eu" + 0.013**"ton"), (18, '0.042**"trump" + 0.030**"presid" + 0.010**"donald" + 0.009**"iran" + 0.008**"report" + 0.007**"american" + 0.007**"forc" + 0.007**"war" + 0.006**"countri" + 0.006**"offici"), (19, '0.044**"email" + 0.041**"imag" + 0.030**"subscrib" + 0.021**"news" + 0.020**"car" + 0.018**"newslett" + 0.015**"video" + 0.014**"stori" + 0.013**"privaci" + 0.013**"road")]	0.56328912491 38561	-8.86195183489 259
19	[(0, '0.011**"pay" + 0.010**"compani" + 0.009**"uk" + 0.008**"cost" + 0.008**"servic" + 0.007**"licenc" + 0.007**"busi" + 0.007**"free" + 0.007**"custom" + 0.006**"pension"), (1, '0.011**"peopl" + 0.009**"health" + 0.008**"time" + 0.006**"patient" + 0.005**"feel" + 0.005**"work" +	0.56327857587 41872	-8.81910996763 446

	<p>0.005**"bodi" + 0.005**"life" + 0.005**"test" + 0.005**"medic"), (2, '0.014**"polic" + 0.013**"road" + 0.013**"car" + 0.011**"imag" + 0.010**"park" + 0.010**"servic" + 0.009**"street" + 0.008**"offic" + 0.008**"area" + 0.008**"south"), (3, '0.028**"polic" + 0.023**"court" + 0.012**"murder" + 0.011**"case" + 0.010**"victim" + 0.010**"hear" + 0.010**"arrest" + 0.009**"tell" + 0.009**"attack" + 0.009**"investig"), (4, '0.024**"win" + 0.014**"open" + 0.014**"england" + 0.013**"time" + 0.013**"cup" + 0.012**"race" + 0.011**"final" + 0.011**"match" + 0.010**"team" + 0.010**"sport"), (5, '0.025**"leagu" + 0.017**"player" + 0.017**"club" + 0.014**"season" + 0.014**"liverpool" + 0.013**"unit" + 0.012**"final" + 0.012**"game" + 0.011**"win" + 0.011**"play"), (6, '0.019**"peopl" + 0.018**"imag" + 0.014**"work" + 0.010**"school" + 0.008**"servic" + 0.008**"copyright" + 0.007**"children" + 0.007**"commun" + 0.006**"bbc" + 0.006**"support"), (7, '0.015**"credit" + 0.012**"email" + 0.011**"famili" + 0.011**"tell" + 0.010**"stori" + 0.009**"subscrib" + 0.009**"news" + 0.009**"day" + 0.009**"time" + 0.008**"babi"), (8, '0.011**"wear" + 0.010**"star" + 0.010**"black" + 0.009**"dress" + 0.008**"hair" + 0.008**"white" + 0.007**"love" + 0.007**"instagram" + 0.007**"girl" + 0.006**"pair"), (9, '0.013**"royal" + 0.013**"music" + 0.012**"queen" + 0.011**"imag" + 0.010**"princ" + 0.009**"song" + 0.009**"harri" + 0.007**"palac" + 0.007**"day" + 0.007**"twitter"), (10, '0.015**"win" + 0.010**"lose" + 0.008**"time" + 0.008**"game" + 0.008**"ball" + 0.007**"point" + 0.007**"minut" + 0.007**"play" + 0.007**"team" + 0.007**"score"), (11, '0.024**"attack" + 0.017**"peopl" + 0.011**"kill" + 0.011**"forc" + 0.011**"war" + 0.010**"group" + 0.009**"shoot" + 0.007**"militari" + 0.007**"countri" + 0.007**"die"), (12, '0.025**"food" + 0.014**"cook" + 0.011**"eat" + 0.010**"water" + 0.010**"restaur" + 0.008**"meat" + 0.008**"drink" + 0.007**"day" + 0.007**"pizza" + 0.007**"fruit"), (13, '0.017**"film" + 0.013**"play" + 0.011**"seri" + 0.010**"episod" + 0.010**"season" + 0.008**"charact" + 0.007**"star" + 0.006**"stori" + 0.006**"watch" + 0.006**"game"), (14, '0.017**"love" + 0.013**"bbc" + 0.012**"itv" + 0.012**"tv" + 0.012**"star" + 0.007**"island" + 0.007**"leav" + 0.007**"time" + 0.006**"reveal" + 0.005**"appear"), (15, '0.008**"island" + 0.007**"water" + 0.007**"beach" + 0.006**"build" + 0.006**"citi" + 0.006**"climat" + 0.006**"rome" + 0.006**"sea" + 0.006**"plastic" + 0.005**"garden"), (16, '0.057**"getti" + 0.034**"pa" + 0.029**"imag" + 0.025**"septemb" + 0.022**"august" + 0.022**"london" + 0.022**"afp" + 0.020**"reuter" + 0.017**"protest" + 0.014**"day"), (17, '0.026**"brexit" + 0.025**"parti" + 0.017**"vote" + 0.014**"deal" + 0.014**"mp" + 0.013**"minist" + 0.013**"johnson" + 0.012**"elect" + 0.012**"eu" + 0.012**"tori"), (18, '0.038**"trump" + 0.027**"presid" + 0.009**"donald" + 0.008**"report" + 0.008**"iran" + 0.007**"china" + 0.006**"american" + 0.006**"countri" + 0.006**"media" + 0.006**"politi")]</p>		
18	<p>[(0, '0.015**"servic" + 0.009**"pay" + 0.009**"compani" + 0.008**"govern" + 0.007**"work" + 0.007**"licenc" + 0.007**"uk" + 0.007**"bbc" + 0.006**"cost" + 0.006**"fund"), (1, '0.015**"health" + 0.012**"peopl" + 0.008**"patient" + 0.007**"medic" + 0.007**"doctor" + 0.007**"bodi" + 0.007**"hospit" + 0.006**"cancer" + 0.006**"diseas" + 0.006**"nh"), (2, '0.014**"uk" + 0.012**"sell" + 0.012**"store" + 0.012**"buy" + 0.012**"price" + 0.012**"custom" + 0.011**"offer" + 0.011**"shop" + 0.010**"bank" + 0.009**"save"), (3, '0.021**"peopl" + 0.015**"school" + 0.013**"work" + 0.012**"children" + 0.008**"time" + 0.008**"life" + 0.008**"student" + 0.007**"live" + 0.007**"young" + 0.006**"univers"), (4, '0.021**"win" + 0.013**"open" + 0.012**"england" + 0.011**"car" + 0.011**"race" + 0.011**"team" + 0.011**"time" + 0.010**"match" + 0.010**"sport" + 0.009**"cup"), (5, '0.023**"leagu" + 0.017**"player" + 0.015**"club" + 0.014**"win" + 0.014**"game" + 0.013**"season" + 0.013**"liverpool" + 0.012**"unit" + 0.012**"final" + 0.011**"team"), (6, '0.036**"imag" + 0.017**"email" + 0.015**"subscrib" + 0.009**"hous" + 0.009**"newslett" + 0.009**"copyright" + 0.009**"build" + 0.008**"park" + 0.008**"council" + 0.008**"citi"), (7, '0.016**"love" + 0.014**"credit" + 0.012**"star" + 0.010**"pictur" + 0.009**"share" + 0.008**"itv" + 0.008**"instagram" + 0.008**"reveal" + 0.007**"fan" + 0.007**"video"), (8, '0.013**"women" + 0.011**"imag" + 0.010**"media" + 0.009**"peopl" + 0.009**"claim" + 0.008**"court" + 0.008**"case" + 0.008**"tell" + 0.007**"bbc" + 0.007**"comment"), (9, '0.028**"trump" + 0.020**"presid" + 0.009**"countri" + 0.008**"war" + 0.007**"china" + 0.007**"donald" + 0.006**"nation" + 0.006**"forc" + 0.006**"iran" + 0.005**"american"), (10, '0.016**"time" + 0.012**"ve" + 0.011**"good" + 0.010**"don" + 0.008**"start" + 0.008**"didn" + 0.008**"day" + 0.007**"feel" + 0.006**"play" + 0.006**"lot"), (11, '0.046**"polic" + 0.021**"attack" + 0.016**"offic" + 0.015**"man" + 0.011**"peopl" + 0.011**"incid" + 0.009**"london" + 0.009**"arrest" + 0.008**"investig" + 0.008**"report"), (12, '0.017**"food" + 0.010**"water" + 0.009**"cook" + 0.009**"eat" + 0.008**"plastic" + 0.008**"hair" + 0.007**"drink" + 0.007**"product" + 0.006**"restaur" + 0.006**"meat"), (13, '0.013**"film" + 0.009**"play" + 0.009**"star" + 0.009**"bbc" + 0.009**"seri" + 0.009**"music" + 0.008**"episod" + 0.007**"season" + 0.007**"tv" + 0.006**"perform"), (14, '0.018**"famili" + 0.013**"die" + 0.012**"death" + 0.012**"tell" + 0.009**"life" + 0.008**"leav" + 0.008**"court" +</p>	0.57639823961 63805	-8.78202762692 907

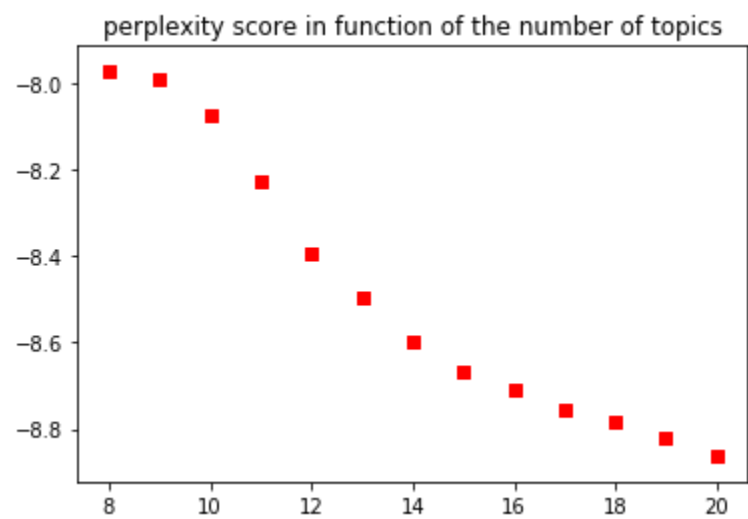
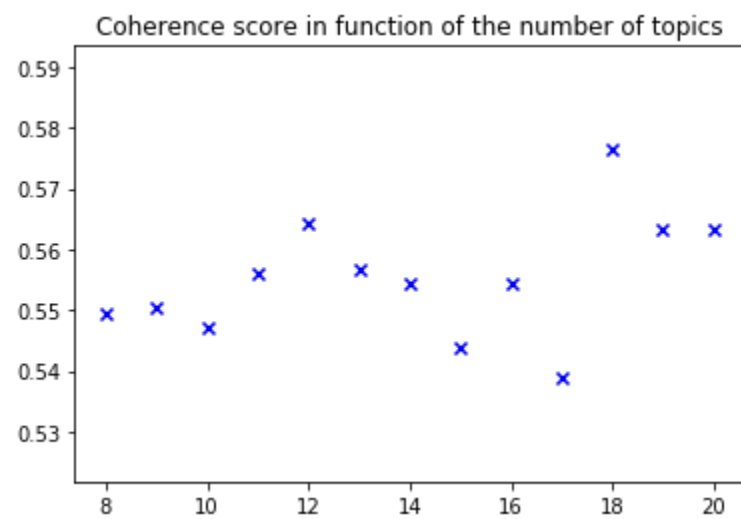
	0.008**"mother" + 0.008**"murder" + 0.008**"children"), (15, '0.020**royal" + 0.018**"queen" + 0.015**"princ" + 0.014**"wear" + 0.014**"dress" + 0.012**"harri" + 0.011**"babi" + 0.011**"black" + 0.011**"palac" + 0.011**"day"), (16, '0.050**"getti" + 0.030**"pa" + 0.024**"septemb" + 0.022**"august" + 0.020**"london" + 0.020**"imag" + 0.019**"afp" + 0.018**"reuter" + 0.015**"protest" + 0.013**"day"), (17, '0.027**"brexit" + 0.020**"parti" + 0.017**"vote" + 0.015**"deal" + 0.014**"mp" + 0.013**"minist" + 0.013**"elect" + 0.013**"johnson" + 0.012**"labour" + 0.012**"tori"]]		
17	[(0, '0.014**"pay" + 0.010**"compani" + 0.009**"cost" + 0.009**"uk" + 0.008**"free" + 0.008**"busi" + 0.008**"custom" + 0.007**"pension" + 0.007**"price" + 0.007**"buy"), (1, '0.009**"food" + 0.006**"peopl" + 0.005**"eat" + 0.005**"time" + 0.005**"day" + 0.005**"health" + 0.005**"diseas" + 0.004**"good" + 0.004**"water" + 0.004**"don"), (2, '0.013**"countri" + 0.011**"war" + 0.010**"forc" + 0.009**"iran" + 0.008**"china" + 0.007**"govern" + 0.007**"attack" + 0.007**"militari" + 0.007**"nation" + 0.006**"offici"), (3, '0.027**"trump" + 0.014**"presid" + 0.013**"peopl" + 0.013**"women" + 0.012**"school" + 0.007**"student" + 0.006**"donald" + 0.005**"time" + 0.004**"polit" + 0.004**"gay"), (4, '0.100**"imag" + 0.072**"getti" + 0.017**"copyright" + 0.015**"race" + 0.015**"win" + 0.014**"afp" + 0.014**"reuter" + 0.012**"time" + 0.010**"open" + 0.007**"team"), (5, '0.018**"leagu" + 0.015**"win" + 0.013**"player" + 0.012**"game" + 0.012**"play" + 0.012**"club" + 0.012**"final" + 0.011**"season" + 0.010**"team" + 0.010**"liverpool"), (6, '0.014**"imag" + 0.013**"peopl" + 0.012**"servic" + 0.011**"work" + 0.010**"bbc" + 0.006**"copyright" + 0.006**"public" + 0.005**"govern" + 0.005**"report" + 0.005**"health"), (7, '0.019**"credit" + 0.014**"email" + 0.012**"love" + 0.012**"news" + 0.012**"subscrib" + 0.011**"video" + 0.011**"stori" + 0.009**"ve" + 0.008**"time" + 0.008**"itv"), (8, '0.012**"star" + 0.009**"wear" + 0.009**"love" + 0.008**"dress" + 0.007**"black" + 0.007**"hair" + 0.007**"girl" + 0.007**"pair" + 0.006**"white" + 0.005**"share"), (9, '0.012**"royal" + 0.012**"queen" + 0.012**"music" + 0.009**"princ" + 0.008**"harri" + 0.008**"song" + 0.007**"visit" + 0.007**"day" + 0.007**"palac" + 0.006**"includ"), (10, '0.009**"hous" + 0.009**"park" + 0.008**"citi" + 0.008**"build" + 0.008**"street" + 0.008**"road" + 0.007**"local" + 0.007**"west" + 0.007**"area" + 0.007**"south"), (11, '0.055**"polic" + 0.019**"attack" + 0.017**"man" + 0.017**"offic" + 0.013**"incid" + 0.012**"investig" + 0.011**"arrest" + 0.011**"car" + 0.009**"report" + 0.009**"victim"), (12, '0.027**"brexit" + 0.026**"parti" + 0.017**"vote" + 0.014**"deal" + 0.014**"mp" + 0.013**"minist" + 0.013**"johnson" + 0.013**"elect" + 0.012**"labour" + 0.012**"eu"), (13, '0.012**"film" + 0.011**"play" + 0.010**"bbc" + 0.009**"seri" + 0.008**"episod" + 0.008**"star" + 0.007**"season" + 0.007**"tv" + 0.006**"time" + 0.006**"charact"), (14, '0.016**"famili" + 0.012**"tell" + 0.012**"children" + 0.010**"die" + 0.010**"life" + 0.010**"court" + 0.009**"death" + 0.008**"leav" + 0.008**"mother" + 0.007**"day"), (15, '0.038**"protest" + 0.021**"getti" + 0.017**"afp" + 0.016**"climat" + 0.013**"rome" + 0.012**"reuter" + 0.012**"hong" + 0.012**"kong" + 0.011**"franc" + 0.011**"india"), (16, '0.036**"pa" + 0.029**"septemb" + 0.027**"august" + 0.026**"london" + 0.015**"england" + 0.014**"day" + 0.010**"getti" + 0.009**"flight" + 0.009**"edinburgh" + 0.008**"test")]	0.53882513374 00836	-8.75447731246 2907
16	[(0, '0.010**"servic" + 0.010**"pay" + 0.008**"compani" + 0.007**"uk" + 0.006**"cost" + 0.006**"licenc" + 0.006**"work" + 0.006**"govern" + 0.006**"busi" + 0.005**"free"), (1, '0.013**"food" + 0.006**"water" + 0.006**"eat" + 0.006**"cook" + 0.006**"product" + 0.006**"buy" + 0.006**"good" + 0.006**"plastic" + 0.005**"day" + 0.005**"drink"), (2, '0.009**"imag" + 0.009**"day" + 0.008**"countri" + 0.007**"south" + 0.007**"peopl" + 0.007**"forc" + 0.006**"air" + 0.005**"war" + 0.005**"land" + 0.005**"franc"), (3, '0.030**"school" + 0.022**"children" + 0.019**"peopl" + 0.015**"student" + 0.013**"univers" + 0.010**"london" + 0.010**"commun" + 0.009**"group" + 0.009**"support" + 0.009**"parent"), (4, '0.059**"getti" + 0.032**"pa" + 0.028**"imag" + 0.026**"septemb" + 0.024**"august" + 0.021**"afp" + 0.020**"reuter" + 0.014**"london" + 0.014**"protes" + 0.013**"day"), (5, '0.017**"leagu" + 0.016**"win" + 0.013**"player" + 0.012**"game" + 0.012**"play" + 0.011**"final" + 0.011**"club" + 0.011**"team" + 0.010**"season" + 0.010**"time"), (6, '0.013**"imag" + 0.010**"hous" + 0.010**"street" + 0.009**"build" + 0.009**"london" + 0.009**"road" + 0.009**"park" + 0.008**"council" + 0.008**"citi" + 0.008**"servic"), (7, '0.021**"credit" + 0.016**"email" + 0.013**"news" + 0.013**"video" + 0.012**"subscrib" + 0.012**"stori" + 0.011**"love" + 0.010**"pictur" + 0.009**"sun" + 0.008**"post"), (8, '0.013**"wear" + 0.013**"star" + 0.011**"black" + 0.011**"dress" + 0.009**"white" + 0.009**"hair" + 0.008**"pair" + 0.006**"look" + 0.006**"model" + 0.006**"girl"), (9, '0.024**"trump" + 0.017**"presid" + 0.006**"media" + 0.006**"report" + 0.006**"donald" + 0.005**"polit" + 0.004**"american" + 0.004**"law" + 0.004**"nation" + 0.004**"iran"), (10, '0.018**"peopl" + 0.011**"time" + 0.010**"women" + 0.010**"health" + 0.009**"life" + 0.009**"feel" + 0.009**"work" + 0.007**"day" + 0.006**"don" + 0.006**"live"), (11, '0.034**"polic" + 0.014**"court" + 0.013**"attack" + 0.012**"man" + 0.011**"offic" + 0.009**"murder" + 0.009**"victim" + 0.009**"investig" + 0.009**"imag" + 0.009**"arrest"), (12, '0.026**"brexit" + 0.025**"parti" +	0.55447708919 65027	-8.70951449406 1082

	0.017**"vote" + 0.014**"mp" + 0.014**"deal" + 0.013**"minist" + 0.013**"johnson" + 0.012**"elect" + 0.012**"labour" + 0.012**"tori")", (13, '0.010**"film" + 0.009**"play" + 0.008**"music" + 0.007**"time" + 0.007**"season" + 0.006**"imag" + 0.006**"episod" + 0.006**"seri" + 0.006**"song" + 0.005**"perform")", (14, '0.015**"bbc" + 0.010**"famili" + 0.008**"love" + 0.008**"leav" + 0.007**"itv" + 0.007**"tv" + 0.007**"tell" + 0.006**"time" + 0.005**"wife" + 0.005**"son")", (15, '0.032**"royal" + 0.029**"queen" + 0.024**"princ" + 0.020**"babi" + 0.020**"harri" + 0.018**"palac" + 0.018**"visit" + 0.017**"day" + 0.014**"william" + 0.014**"meghan")"]]		
15	[(0, '0.012**"servic" + 0.010**"pay" + 0.008**"compani" + 0.007**"uk" + 0.007**"cost" + 0.006**"busi" + 0.006**"licenc" + 0.005**"work" + 0.005**"free" + 0.005**"govern")", (1, '0.010**"food" + 0.006**"water" + 0.006**"eat" + 0.005**"patient" + 0.005**"product" + 0.005**"day" + 0.005**"diseas" + 0.004**"cook" + 0.004**"plastic" + 0.004**"drink")", (2, '0.010**"countri" + 0.007**"south" + 0.007**"forc" + 0.007**"war" + 0.007**"imag" + 0.007**"day" + 0.007**"iran" + 0.006**"peopl" + 0.005**"land" + 0.005**"china")", (3, '0.026**"peopl" + 0.013**"work" + 0.011**"women" + 0.010**"school" + 0.009**"time" + 0.008**"children" + 0.007**"feel" + 0.007**"live" + 0.007**"imag" + 0.006**"univers")", (4, '0.085**"getti" + 0.078**"imag" + 0.036**"pa" + 0.030**"protest" + 0.027**"afp" + 0.027**"reuter" + 0.014**"email" + 0.014**"london" + 0.011**"sketch" + 0.010**"day")", (5, '0.021**"leagu" + 0.015**"player" + 0.014**"win" + 0.013**"club" + 0.013**"game" + 0.013**"play" + 0.013**"final" + 0.012**"season" + 0.011**"liverpool" + 0.011**"cup")", (6, '0.015**"august" + 0.015**"london" + 0.013**"septemb" + 0.011**"england" + 0.009**"build" + 0.009**"citi" + 0.009**"street" + 0.008**"council" + 0.008**"day" + 0.007**"hous")", (7, '0.015**"credit" + 0.013**"love" + 0.012**"star" + 0.009**"video" + 0.008**"instagram" + 0.008**"share" + 0.008**"pictur" + 0.007**"email" + 0.007**"subscrib" + 0.007**"fan")", (8, '0.025**"trump" + 0.016**"presid" + 0.007**"claim" + 0.007**"report" + 0.006**"court" + 0.006**"media" + 0.006**"law" + 0.006**"case" + 0.005**"tell" + 0.005**"donald")", (9, '0.010**"royal" + 0.010**"queen" + 0.009**"music" + 0.008**"princ" + 0.007**"harri" + 0.007**"song" + 0.006**"includ" + 0.006**"design" + 0.005**"palac" + 0.005**"visit")", (10, '0.015**"win" + 0.013**"time" + 0.007**"team" + 0.007**"lose" + 0.007**"start" + 0.006**"ve" + 0.006**"good" + 0.006**"day" + 0.006**"open" + 0.006**"point")", (11, '0.036**"polic" + 0.013**"attack" + 0.013**"man" + 0.013**"imag" + 0.011**"offic" + 0.010**"court" + 0.010**"murder" + 0.009**"victim" + 0.009**"incid" + 0.009**"arrest")", (12, '0.027**"brexit" + 0.026**"parti" + 0.017**"vote" + 0.015**"deal" + 0.015**"mp" + 0.014**"minist" + 0.013**"johnson" + 0.013**"elect" + 0.012**"labour" + 0.012**"eu")", (13, '0.015**"film" + 0.013**"bbc" + 0.012**"play" + 0.011**"seri" + 0.010**"star" + 0.010**"tv" + 0.010**"episod" + 0.008**"season" + 0.007**"charact" + 0.006**"actor")", (14, '0.018**"famili" + 0.012**"tell" + 0.010**"life" + 0.010**"leav" + 0.009**"day" + 0.009**"son" + 0.009**"die" + 0.008**"mother" + 0.008**"babi" + 0.008**"daughter")"]]	0.54365754522 6414	-8.66792360709 2357
14	[(0, '0.010**"servic" + 0.009**"pay" + 0.007**"compani" + 0.007**"uk" + 0.007**"work" + 0.007**"peopl" + 0.006**"cost" + 0.006**"govern" + 0.005**"busi" + 0.005**"licenc")", (1, '0.009**"food" + 0.006**"peopl" + 0.005**"health" + 0.005**"eat" + 0.005**"patient" + 0.005**"time" + 0.005**"day" + 0.004**"diseas" + 0.004**"good" + 0.004**"water")", (2, '0.027**"trump" + 0.020**"presid" + 0.010**"countri" + 0.008**"war" + 0.008**"forc" + 0.007**"china" + 0.007**"protest" + 0.007**"nation" + 0.006**"iran" + 0.006**"donald")", (3, '0.011**"peopl" + 0.009**"school" + 0.009**"women" + 0.007**"children" + 0.006**"case" + 0.006**"court" + 0.006**"tell" + 0.005**"time" + 0.005**"student" + 0.005**"claim")", (4, '0.055**"getti" + 0.032**"pa" + 0.025**"septemb" + 0.024**"august" + 0.021**"afp" + 0.020**"reuter" + 0.018**"imag" + 0.018**"london" + 0.014**"protest" + 0.012**"day")", (5, '0.016**"leagu" + 0.016**"win" + 0.012**"player" + 0.012**"play" + 0.012**"game" + 0.011**"final" + 0.011**"club" + 0.011**"team" + 0.010**"season" + 0.010**"time")", (6, '0.011**"imag" + 0.009**"build" + 0.008**"citi" + 0.007**"hous" + 0.007**"park" + 0.007**"road" + 0.006**"street" + 0.006**"local" + 0.006**"peopl" + 0.006**"council")", (7, '0.014**"credit" + 0.012**"famili" + 0.011**"love" + 0.010**"time" + 0.010**"tell" + 0.009**"day" + 0.009**"life" + 0.008**"leav" + 0.007**"email" + 0.007**"news")", (8, '0.011**"star" + 0.008**"wear" + 0.008**"black" + 0.007**"dress" + 0.007**"love" + 0.007**"share" + 0.007**"pictur" + 0.006**"instagram" + 0.006**"white" + 0.006**"fan")", (9, '0.020**"imag" + 0.009**"music" + 0.009**"royal" + 0.008**"queen" + 0.007**"princ" + 0.006**"time" + 0.006**"song" + 0.006**"copyright" + 0.006**"harri" + 0.005**"photograph")", (10, '0.015**"day" + 0.012**"south" + 0.011**"west" + 0.010**"wale" + 0.010**"england" + 0.009**"scotland" + 0.009**"london" + 0.009**"rain" + 0.009**"north" + 0.008**"uk")", (11, '0.035**"polic" + 0.013**"man" + 0.013**"attack" + 0.010**"offic" + 0.010**"imag" + 0.010**"murder" + 0.009**"death" + 0.009**"arrest" + 0.009**"incid" + 0.009**"victim")", (12, '0.025**"brexit" + 0.025**"parti" + 0.016**"vote" + 0.014**"deal" + 0.014**"mp" + 0.012**"minist" + 0.012**"elect" + 0.012**"johnson" + 0.012**"labour" + 0.012**"eu")", (13, '0.014**"film" + 0.014**"bbc" + 0.012**"tv" + 0.011**"play" + 0.011**"seri" + 0.010**"episod" + 0.009**"star" + 0.008**"season" + 0.008**"itv" + 0.007**"charact")"]]	0.55441510049 16393	-8.59857093994 5565

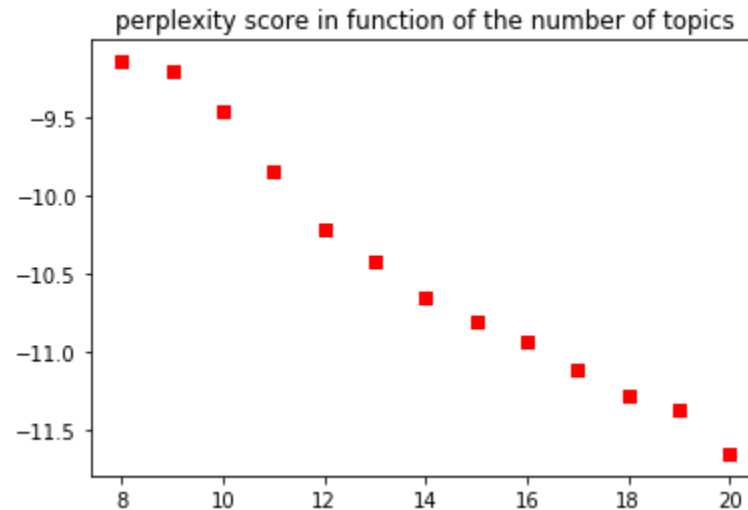
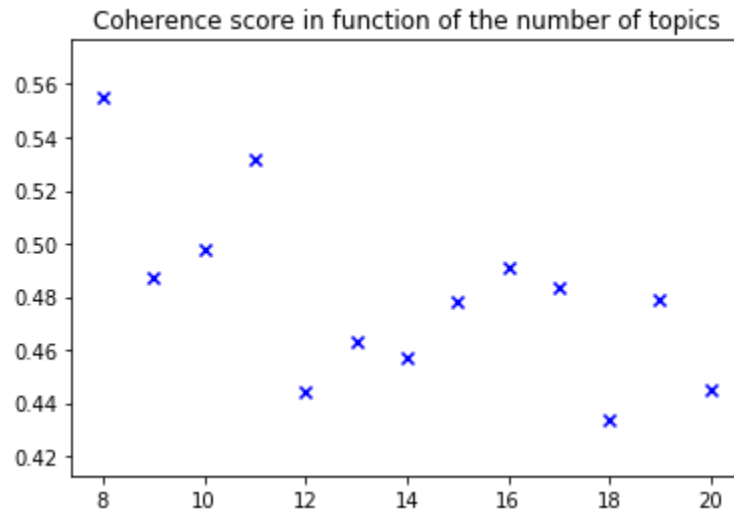
13	<p>[(0, '0.010**servic" + 0.009**pay" + 0.008**compani" + 0.008**uk" + 0.006**cost" + 0.006**licenc" + 0.005**work" + 0.005**govern" + 0.005**busi" + 0.005**free"), (1, '0.009**food" + 0.006**eat" + 0.005**patient" + 0.005**diseas" + 0.005**water" + 0.005**health" + 0.005**day" + 0.004**product" + 0.004**hair" + 0.004**cook"), (2, '0.010**countri" + 0.008**day" + 0.007**war" + 0.007**forc" + 0.007**imag" + 0.007**south" + 0.007**china" + 0.006**iran" + 0.006**peopl" + 0.005**attack"), (3, '0.017**peopl" + 0.012**time" + 0.011**work" + 0.009**life" + 0.009**children" + 0.008**women" + 0.008**school" + 0.008**feel" + 0.007**live" + 0.007**don"), (4, '0.049**getti" + 0.031**pa" + 0.024**septemb" + 0.022**august" + 0.019**afp" + 0.018**reuter" + 0.015**london" + 0.014**imag" + 0.013**day" + 0.012**england"), (5, '0.016**leagu" + 0.016**win" + 0.012**player" + 0.011**game" + 0.011**play" + 0.011**final" + 0.011**club" + 0.010**team" + 0.010**season" + 0.009**time"), (6, '0.015**imag" + 0.008**street" + 0.008**london" + 0.008**build" + 0.008**hous" + 0.007**email" + 0.007**citi" + 0.007**park" + 0.006**road" + 0.006**council"), (7, '0.012**credit" + 0.012**love" + 0.008**star" + 0.008**pictur" + 0.007**babi" + 0.006**leav" + 0.006**share" + 0.006**tell" + 0.006**itv" + 0.006**reveal"), (8, '0.025**trump" + 0.015**presid" + 0.007**media" + 0.006**report" + 0.006**claim" + 0.006**donald" + 0.005**law" + 0.005**comment" + 0.005**protest" + 0.005**public"), (9, '0.010**bbc" + 0.009**imag" + 0.009**film" + 0.007**star" + 0.007**music" + 0.006**royal" + 0.006**queen" + 0.006**play" + 0.006**perform" + 0.005**tv"), (10, '0.033**episod" + 0.030**season" + 0.015**play" + 0.013**charact" + 0.013**game" + 0.012**seri" + 0.012**dog" + 0.010**hbo" + 0.010**cat" + 0.009**spice"), (11, '0.029**polic" + 0.011**man" + 0.010**attack" + 0.009**die" + 0.009**court" + 0.009**death" + 0.009**tell" + 0.009**offic" + 0.009**imag" + 0.008**murder"), (12, '0.026**brexit" + 0.025**parti" + 0.017**vote" + 0.014**deal" + 0.014**mp" + 0.013**minist" + 0.012**johnson" + 0.012**elect" + 0.012**labour" + 0.012**eu")]</p>	0.55656317888 63326	-8.49501589411 3969
12	<p>[(0, '0.010**servic" + 0.008**pay" + 0.007**compani" + 0.007**uk" + 0.005**work" + 0.005**cost" + 0.005**busi" + 0.005**plan" + 0.005**govern" + 0.005**licenc"), (1, '0.008**food" + 0.005**water" + 0.005**day" + 0.005**eat" + 0.004**health" + 0.004**time" + 0.004**good" + 0.004**product" + 0.004**patient" + 0.004**diseas"), (2, '0.011**countri" + 0.007**war" + 0.007**forc" + 0.007**china" + 0.006**day" + 0.006**iran" + 0.006**govern" + 0.006**nation" + 0.006**south" + 0.006**attack"), (3, '0.023**peopl" + 0.012**work" + 0.010**school" + 0.008**time" + 0.008**live" + 0.007**imag" + 0.007**children" + 0.006**univers" + 0.006**life" + 0.006**women"), (4, '0.024**brexit" + 0.024**parti" + 0.016**vote" + 0.013**deal" + 0.013**mp" + 0.013**minist" + 0.012**johnson" + 0.012**elect" + 0.011**labour" + 0.011**eu"), (5, '0.015**leagu" + 0.015**win" + 0.011**player" + 0.010**play" + 0.010**final" + 0.010**game" + 0.010**club" + 0.010**team" + 0.009**time" + 0.009**season"), (6, '0.029**getti" + 0.024**imag" + 0.017**pa" + 0.014**septemb" + 0.014**london" + 0.013**august" + 0.011**afp" + 0.011**reuter" + 0.010**day" + 0.008**protest"), (7, '0.012**love" + 0.012**credit" + 0.008**time" + 0.007**star" + 0.007**pictur" + 0.007**tell" + 0.006**leav" + 0.006**day" + 0.006**ve" + 0.006**stori"), (8, '0.025**trump" + 0.014**presid" + 0.008**white" + 0.008**women" + 0.006**black" + 0.006**wear" + 0.006**media" + 0.006**comment" + 0.005**donald" + 0.005**dress"), (9, '0.009**film" + 0.009**bbc" + 0.008**play" + 0.007**star" + 0.007**seri" + 0.007**music" + 0.006**episod" + 0.005**time" + 0.005**tv" + 0.005**season"), (10, '0.021**famili" + 0.018**babi" + 0.015**children" + 0.015**royal" + 0.013**queen" + 0.013**mother" + 0.012**day" + 0.012**princ" + 0.011**child" + 0.011**son"), (11, '0.024**polic" + 0.010**court" + 0.010**man" + 0.009**attack" + 0.009**imag" + 0.009**offic" + 0.008**tell" + 0.007**death" + 0.007**investig" + 0.007**die")]</p>	0.56417215446 45778	-8.39256651438 7617
11	<p>[(0, '0.010**servic" + 0.008**pay" + 0.007**compani" + 0.007**uk" + 0.006**cost" + 0.006**busi" + 0.006**plan" + 0.006**work" + 0.005**bbc" + 0.005**licenc"), (1, '0.009**food" + 0.005**patient" + 0.005**eat" + 0.005**day" + 0.004**product" + 0.004**hair" + 0.004**health" + 0.004**good" + 0.004**diseas" + 0.004**plastic"), (2, '0.027**polic" + 0.013**attack" + 0.010**man" + 0.009**offic" + 0.008**imag" + 0.008**die" + 0.008**kill" + 0.008**death" + 0.008**murder" + 0.007**incid"), (3, '0.020**peopl" + 0.010**work" + 0.009**women" + 0.008**school" + 0.008**children" + 0.008**time" + 0.008**imag" + 0.007**life" + 0.006**live" + 0.005**univers"), (4, '0.024**brexit" + 0.024**parti" + 0.015**vote" + 0.014**deal" + 0.013**mp" + 0.013**minist" + 0.012**elect" + 0.012**johnson" + 0.011**labour" + 0.011**eu"), (5, '0.015**leagu" + 0.015**win" + 0.011**player" + 0.010**play" + 0.010**final" + 0.010**game" + 0.010**club" + 0.010**team" + 0.009**time" + 0.009**season"), (6, '0.033**getti" + 0.023**imag" + 0.019**pa" + 0.016**london" + 0.015**septemb" + 0.014**august" + 0.013**afp" + 0.012**reuter" + 0.012**day" + 0.012**protest"), (7, '0.010**love" + 0.009**credit" + 0.007**time" + 0.007**tell" + 0.007**pictur" + 0.006**leav" + 0.006**day" + 0.006**famili" + 0.006**star" + 0.005**babi"), (8, '0.022**trump" + 0.014**presid"</p>	0.55602893225 18506	-8.22889306314 5003

	+ 0.009**court" + 0.007**claim" + 0.006**report" + 0.006**case" + 0.006**tell" + 0.006**media" + 0.005**law" + 0.005**donald"), (9, '0.008**film" + 0.007**play" + 0.007**star" + 0.006**music" + 0.006**seri" + 0.006**bbc" + 0.005**royal" + 0.005**episod" + 0.005**season" + 0.004**queen"), (10, '0.008**imag" + 0.008**day" + 0.008**park" + 0.007**south" + 0.007**water" + 0.006**hous" + 0.005**travel" + 0.005**west" + 0.005**local" + 0.005**area"))]		
10	[(0, '0.008**pay" + 0.008**servic" + 0.007**uk" + 0.007**compani" + 0.006**work" + 0.006**cost" + 0.005**busi" + 0.005**time" + 0.005**plan" + 0.005**licenc"), (1, '0.006**food" + 0.006**peopl" + 0.006**health" + 0.005**time" + 0.005**water" + 0.004**day" + 0.004**patient" + 0.004**eat" + 0.003**work" + 0.003**good"), (2, '0.023**poli" + 0.011**attack" + 0.009**man" + 0.008**offic" + 0.007**die" + 0.007**court" + 0.007**kill" + 0.007**death" + 0.007**tell" + 0.007**imag"), (3, '0.017**imag" + 0.012**school" + 0.009**children" + 0.009**london" + 0.009**peopl" + 0.008**work" + 0.008**servic" + 0.007**live" + 0.007**copyright" + 0.007**hous"), (4, '0.024**brexit" + 0.024**parti" + 0.016**vote" + 0.013**deal" + 0.013**mp" + 0.013**minist" + 0.012**johnson" + 0.012**elect" + 0.011**labour" + 0.011**eu"), (5, '0.015**win" + 0.014**leagu" + 0.011**player" + 0.010**play" + 0.010**final" + 0.010**game" + 0.010**club" + 0.009**team" + 0.009**time" + 0.009**season"), (6, '0.025**getti" + 0.016**pa" + 0.016**imag" + 0.013**day" + 0.013**septemb" + 0.012**august" + 0.011**afp" + 0.011**reuter" + 0.010**london" + 0.009**protest"), (7, '0.010**love" + 0.009**credit" + 0.008**time" + 0.007**tell" + 0.006**pictur" + 0.006**day" + 0.006**leav" + 0.006**famili" + 0.006**life" + 0.006**star"), (8, '0.014**trump" + 0.010**peopl" + 0.009**presid" + 0.007**women" + 0.005**media" + 0.005**report" + 0.004**countri" + 0.004**time" + 0.004**case" + 0.004**tell"), (9, '0.008**film" + 0.007**star" + 0.006**play" + 0.006**music" + 0.005**seri" + 0.005**royal" + 0.005**episod" + 0.005**bbc" + 0.004**queen" + 0.004**time"))]	0.54722383185 22137	-8.07267877309 6847
9	[(0, '0.010**servic" + 0.007**pay" + 0.007**work" + 0.007**uk" + 0.006**peopl" + 0.006**compani" + 0.005**time" + 0.005**cost" + 0.005**busi" + 0.005**bbc"), (1, '0.006**peopl" + 0.006**food" + 0.005**time" + 0.005**health" + 0.004**day" + 0.004**work" + 0.004**good" + 0.004**don" + 0.004**eat" + 0.003**feeli"), (2, '0.024**poli" + 0.011**attack" + 0.009**man" + 0.008**offic" + 0.008**imag" + 0.008**die" + 0.007**court" + 0.007**death" + 0.007**kill" + 0.006**telli"), (3, '0.007**film" + 0.006**star" + 0.006**play" + 0.005**music" + 0.005**seri" + 0.005**queen" + 0.004**royal" + 0.004**perform" + 0.004**time" + 0.004**bbc"), (4, '0.023**brexit" + 0.023**parti" + 0.015**vote" + 0.014**minist" + 0.013**deal" + 0.013**mp" + 0.012**johnson" + 0.011**elect" + 0.011**labour" + 0.011**eu"), (5, '0.014**win" + 0.014**leagu" + 0.010**player" + 0.010**club" + 0.010**play" + 0.010**final" + 0.010**game" + 0.009**team" + 0.009**time" + 0.008**season"), (6, '0.015**imag" + 0.011**getti" + 0.010**london" + 0.010**pa" + 0.009**day" + 0.008**septemb" + 0.007**august" + 0.006**peopl" + 0.006**south" + 0.006**afp"), (7, '0.010**love" + 0.009**credit" + 0.008**time" + 0.007**tell" + 0.006**famili" + 0.006**pictur" + 0.006**leav" + 0.006**day" + 0.006**life" + 0.005**star"), (8, '0.016**trump" + 0.011**presid" + 0.009**peopl" + 0.007**women" + 0.005**media" + 0.005**report" + 0.005**imag" + 0.004**tell" + 0.004**countri" + 0.004**case"))]	0.55032876482 53006	-7.99173891019 4271
8	[(0, '0.008**servic" + 0.007**peopl" + 0.007**pay" + 0.007**work" + 0.006**uk" + 0.006**compani" + 0.005**time" + 0.005**govern" + 0.005**bbc" + 0.005**cost"), (1, '0.005**food" + 0.005**time" + 0.005**peopl" + 0.005**day" + 0.004**health" + 0.004**don" + 0.004**good" + 0.004**work" + 0.003**feel" + 0.003**eat"), (2, '0.018**poli" + 0.009**attack" + 0.008**court" + 0.008**tell" + 0.007**man" + 0.007**imag" + 0.007**offic" + 0.006**report" + 0.006**die" + 0.006**investig"), (3, '0.011**trump" + 0.007**peopl" + 0.006**presid" + 0.005**imag" + 0.005**time" + 0.005**women" + 0.005**bbc" + 0.005**film" + 0.004**write" + 0.004**work"), (4, '0.021**parti" + 0.021**brexit" + 0.014**vote" + 0.013**minist" + 0.012**deal" + 0.012**mp" + 0.011**johnson" + 0.011**elect" + 0.010**labour" + 0.010**tori"), (5, '0.014**win" + 0.013**leagu" + 0.010**game" + 0.010**player" + 0.010**play" + 0.010**final" + 0.010**club" + 0.009**season" + 0.009**team" + 0.009**time"), (6, '0.014**imag" + 0.011**getti" + 0.010**london" + 0.009**pa" + 0.009**day" + 0.008**septemb" + 0.007**august" + 0.005**south" + 0.005**peopl" + 0.005**afp"), (7, '0.009**love" + 0.008**star" + 0.007**credit" + 0.006**time" + 0.006**famili" + 0.006**pictur" + 0.005**telli" + 0.005**leav" + 0.005**day" + 0.004**babi"))]	0.54938660781 72877	-7.97348199826 2609

Without tf-idf :



With tf-idf :



The analysis :

As you can see from the results, combining Tfidf with LDA generates less coherence in the results (**that are mainly non-interpretable**). After some digging on the internet, I found out that It is recommended not to combine the two. To explain, tf-idf gives a certain weight (word_score_1) per word and LDA gives another weight per word w.r.t. a particular topic (word_score_2), what the model does is that it multiplies word_score_1 and word_score_2 for a particular word in a particular topic. And since tf-idf gives common words lower weights, multiplying will give priority to unique words over common words. That's why we find for example: when number of topics = 8, the keywords of the first topic are mainly **names** that are irrelevant to us and cannot be interpretable.

```
[(0, '0.013**"khan" + 0.008**"robinson" + 0.008**"custodi" + 0.008**"normandi" + 0.007**"bailey" + 0.006**"carbon" + 0.006**"magistr" + 0.005**"broadband" + 0.005**"sentenc" + 0.005**"sadiq")]
```

What to do ?

For the upcoming weeks, I will mainly try to focus on the data cleaning since it is the most important part. The text count **alone** is only relevant for really high number of occurrences.

Sorry\n\nSorry, you clicked a link to an article that we no longer retain a link to. NewsNow retains a record of links to stories for 30 days, so it may be that the link you clicked is out of date. Please close this window and try clicking a link to another article.	2475
What term do you want to search? Search with google	2249
We use cookies to give you a better service e.g. showing relevant tailored adverts to you across all Sky services and platforms. By using this website or closing this message you agree to our use of cookies.	1419
The i's Essential Daily Briefing\n\nWe know that sometimes it's easier for us to come to you with the news. That's why our new email newsletter will deliver a mobile-friendly snapshot of inews.co.uk to your inbox every morning, from Monday to Saturday.\n\nThis will feature the stories you need to know, as well as a curated selection of the best reads from across the site. Of course, you can easily opt out at any time, but we're confident that you won't.\n\nOliver Duff, Editor\n\nBy entering your email address and clicking on the sign up button below, you are agreeing to receive the latest daily news, news features and service updates from the i via email. You can unsubscribe at any time and we will not pass on your information.	858
How Oath and our partners bring you better ad experiences To give you a better overall experience, we want to provide relevant ads that are more useful to you. For example, when you search for a film, we use your search information and location to show the most relevant cinemas near you. We also use this information to show you ads for similar films you may like in the future. Like Oath, our partners may also show you ads that they think match your interests. Learn more about how Oath collects and uses data and how our partners collect and use data.\n\nHuffPost is part of Oath. Oath and our partners need your consent to access your device and use your data (including location) to understand your interests, and provide and measure personalised ads. Oath will also provide you with personalised ads on partner products. Learn more. Select 'OK' to continue and allow Oath and our partners to use your data, or select 'Manage options' to view your choices.	818

I will try to find another alternative to filter out the unnecessary. I am open to any suggestions.

Possible patterns :

- Exploring text length and combining it with the number of occurrences
- I found that when the title is the same as the text, it is definitely not an article (used)
- Hand-coding the subdomains for all media paths given by Ina and check which ones are relevant (already done for the top 20 for UK) **(time consuming)**