

CODEBOEK Kranten: VARIABELEN



Krantenmedia

De Standaard
De Morgen
De Tijd
Het Nieuwsblad
Het Laatste Nieuws
Gazet van Antwerpen
Het Belang van Limburg
Metro

PERIODE

1 januari 2003 – 31 december 2014

GOPRESS VIEWER

Tool om de tekst van een krantenartikel te bekijken (mits toegang tot GoPress, bv. via licentie bibliotheek).
Aanvraag login via ine.kuypers@uantwerpen.be.
<http://steunpuntmedia-softlayer.mediahaven.com/gopress/home>

OVERZICHT VARIABELEN

Articles (themabestand) op het niveau van één krantenartikel	People (actorenbestand) op het niveau van één persoon in een krantenartikel	Organisations (actorenbestand) op het niveau van één organisatie in een krantenartikel
Artikel_id	Artikel_id	Artikel_id
Sectie	Sectie	Sectie
Paginanummer	Paginanummer	Paginanummer
Titel	Titel	Titel
Ondertitel	Ondertitel	Ondertitel
Content	Content	Content
Journalist	Journalist	Journalist
Lengte	Lengte	Lengte
Lengte_relatief	Lengte_relatief	Lengte_relatief
Uniek	Uniek	Uniek
Medium	Medium	Medium
Edities	Edities	Edities
Datum	Datum	Datum
Woorden_totaal	Woorden_totaal	Woorden_totaal
Artikelen_totaal	Artikelen_totaal	Artikelen_totaal
Steunpuntcodes	Steunpuntcodes	Steunpuntcodes
Steunpuntsbpcodes	Steunpuntsbpcodes	Steunpuntsbpcodes
Zeticoncodes	Zeticoncodes	Zeticoncodes
Wikicodes	Wikicodes	Wikicodes
Locaties	Locaties	Locaties
Landen	Landen	Landen
Restcategorie	Restcategorie	Restcategorie
	Persoon_naam	Organisatie_naam
	Persoon_functie	Organisatie_functie
	Geslacht	
	Etniciteit	
	Politieke_partij	
	Religie	
	Beroep	
	Nationaliteit	

	Geboortedatum	
	Geboorteplaats	
	Sterftedatum	
Extra informatie over variabelen		
Naam variabele	Label/omschrijving	
Artikel_id	Referentie naar het oorspronkelijk krantenartikel	
Sectie	Katernen	
Paginanummer		
Titel		
Ondertitel		
Content	Niet in de output, maar wel mogelijk om erop te zoeken	
Journalist		
Lengte	Aantal woorden	
Lengte_relatief	Aantal woorden/woorden_totaal	
Uniek	1=uniek 2=duplicaat (bv. regiokatern)	
Medium	De Morgen De Tijd de_gazet_van_antwerpen de_standard Het Laatste Nieuws Het_laatste_nieuws het_belang_van_limburg het_nieuwsblad metro	
Edities	Lokale/regionale/nationale pagina's	
Datum	dd-mmm-yyyy	
Woorden_totaal	Totaal aantal woorden per artikel	
Artikelen_totaal	Totaal aantal artikelen per krant	
Steunpuntcodes	Tekstueel; volgorde heeft geen waarde Zie codeboek (hoofdcategoriën) http://www.steunpuntmedia.be/wp-content/uploads/2015/04/Codeboek-TV-Nieuwsarchief.pdf	
Steunpuntsbpcodes	Zie codeboek (volgorde heeft geen waarde) http://www.steunpuntmedia.be/wp-content/uploads/2015/04/Codeboek-TV-Nieuwsarchief.pdf	
Zeticoncodes	Zie codeboek http://www.iptc.org/std/NewsCodes/IPTC-MediaTopic-NewsCodes.xls	
Wikicodes	Wikipedia search	
Locaties		
Landen	Afgeleid van de variabele 'Locaties'	
Restcategorie		

BETROUWBAARHEID STEUNPUNTCODES			
	id	Precision	Recall
0	celebrity	0.453704	0.505155
1	ruimtelijke structuur	0.764706	0.722222
2	gerecht	0.658879	0.839286
3	rampen	0.652174	0.576923
4	financiën	0.621622	0.696970
5	politieke organisatie	0.678363	0.716049
6	vermaak	0.400000	0.210526
7	derde wereld	1.000.000	0.090909
8	cultuur	0.518072	0.864322
9	rechten en vrijheden	0.555556	0.454545
10	criminaliteit	0.668016	0.812808
11	mobiliteit	0.700935	0.842697
12	sport	0.742222	0.927778
13	oorlog en vrede	0.727273	0.571429
14	economie	0.519553	0.715385
15	sociale zaken	0.484848	0.678788
16	onderwijs	0.725000	0.690476
17	telecommunicatie	0.605263	0.522727
18	werk	0.741379	0.632353
19	media	0.617647	0.428571
20	consument	0.588235	0.612245
21	godsdienst	0.657143	0.605263
22	energie	1.000.000	0.600000
23	migratie	1.000.000	0.454545
24	europa	0.843750	0.692308
25	internationale contacten	0.400000	0.181818
26	royalty	0.916667	0.846154
27	toerisme	1.000.000	0.352941
28	milieu	0.428571	0.461538
29	wetenschappen	0.312500	0.200000
30	natuur	0.680000	0.680000
31	landbouw	1.000.000	0.363636
32	demografie	1.000.000	0.200000
33	defensie en wapens	0.857143	0.500000
34	ruimtevaart	1.000.000	0.333333
		9 0.619376	0.701883
instances:	1415		
micro f1:	0.658053		
macro f1:	0.585354		
errors:	1588		
accuracy:	0.490372		
misrate:	0.509628		
skipped:	0		
mean f1:	0.695485		
instances:	micro_f1	accuracy	misrate
1415	65.805	49.037	50.963