

Basic statistics for BiographyNed

October 19, 2012

Informatie	Aantal keer aanwezig	Percentage(%) of Gemiddelde
Naam	76359	100.00%
Geboortedatum	45559	59.66%
Geboorteplaats	41852	54.81%
Sterfdatum	51598	67.57%
Sterfplaats	39616	51.88%
Sekse	56780	74.36%
Rubriek	60925	79.79%
Claim to faim	100	0.13%
Religie	14100	18.47%
Vader	300	0.39%
Moeder	241	0.32%
Partner	10	0.01%
Beroep	24982	32.72%
Tekst	60135	78.75%
Woorden (totaal)/persoon	22054838	288.83
Woorden (langste tekst)/persoon	17489215	229.04
Woorden (totaal)/teksten	22054838	366.76
Woorden (langste tekst)/teksten	17489215	290.83
XML-files per individu	213315	2.79
Total	76360	100%

Table 1: Statistieken over individuen

Informatie	Aantal keer aanwezig	Percentage(%) of Gemiddelde
Naam	159061	74.57%
Geboortedatum	84188	39.47%
Geboorteplaats	73573	34.49%
Sterfdatum	91920	43.09%
Sterfplaats	67873	31.82%
Sekse	92545	43.38%
Rubriek	72707	34.08%
Claim to faim	100	0.05%
Religie	16918	7.93%
Vader	300	0.14%
Moeder	241	0.11%
Partner	10	0.00%
Beroep	26658	12.50%
Tekst	79012	37.04%
Tekstlengte (woorden/bio)	22054838	103.39
Tekstlengte (woorden/tekst)	22054838	279.13
Total	213315	100%

Table 2: Statistieken over XML-files

Bron	Tot	Naam	geb.d.	geb.pl	sterf.d	sterf.p	seks	Beroep	Tekst	W (tot)	W/tekst
0	94119	42	29	29	36	29	44	0	1	20519	42
1.1	577	100	100	0	99	0	0	100	100	967345	1676
1.2	21391	100	0	0	0	0	0	0	100	7373063	344
1.3	23961	100	0	0	0	0	0	0	100	5656267	236
1.4	2146	100	100	0	100	0	100	96	100	3458494	1611
1.5	1218	100	86	0	95	0	100	0	100	1266992	1040
1.6	102	100	100	91	100	92	100	0	100	48518	475
1.7	282	100	88	0	94	0	0	100	100	441299	1564
2.x	23437	100	78	76	76	58	100	0	0	0	0
3.01	1279	100	77	75	77	75	0	0	100	887420	693
3.02	20121	100	77	82	78	75	94	92	94	504349	26
3.03	2420	100	83	0	87	0	0	0	100	47133	19
3.04	1362	100	0	0	0	0	0	0	0	0	0
3.05	12092	100	63	25	74	24	0	0	0	0	0
3.06	1553	100	97	93	96	81	100	98	46	723934	1002
3.07	3681	100	100	100	100	100	83	99	73	185636	69
3.08	460	100	91	9	91	0	0	0	0	0	0
3.09	401	100	94	89	94	83	0	0	100	149478	372
3.10	409	100	96	0	95	0	100	0	100	2863	7
3.11	115	100	0	0	0	0	0	0	1	0	0
4	2189	100	94	93	99	96	0	0	92	321528	158

Table 3: Statistieken per bron (percentages, behalve W (=aantal woorden))

Informatie	Bron	Percentage
Rubriek	0	77%
Claim2Faim	1.6	98%
Godsdienst	0, 3.07	15%, 70%
Vader	3.09	75%
Moeder	3.09	60%
Partner	3.09	2%

Table 4: Informatie die maar in 1 of 2 bronnen voorkomt met percentages

Bron	id
unknown	0.0
XXX	0.1
Biografisch Woordenboek van het Socialisme en de Arbeidersbeweging in Nederland	1.1
Nieuw Nederlandsch Biografisch Woordenboek (Instituut voor Nederlandse Geschiedenis)	1.2
Van der Aa (Instituut voor Nederlandse Geschiedenis)	1.3
Biografisch Woordenboek van Nederland	1.4
Digitaal Vrouwenlexicon van Nederland	1.5
Database Vrouwenbiografieën	1.6
Biografisch Woordenboek Gelderland	1.7
Digitale Bibliotheek voor de Nederlandse Letteren	2.x
wikipedia	3.01
Rijksbureau voor Kunsthistorische Documentatie	3.02
Het Geheugen van Nederland	3.03
Instituut voor Nederlandse Geschiedenis	3.04
RAA	3.05
KNAW - Levensberichten en Herdenkingen	3.06
Parlementair Documentatie Centrum	3.07
Huygens Instituut voor Nederlandse Geschiedenis	3.08
Joods historisch museum	3.09
Nationaal Archief	3.10
Internationaal Instituut voor Sociale Geschiedenis	3.11
Persoon	4

Table 5: Bronnen en hun Id nummer

Informatie	Info	Tekst+Info	Tekst+Info %	W/tekst	Geen Info	Tekst-Info	Tekst-Info %	W/tekst
Naam	159061	79012	49.7%	279	54254	0	0.0%	0
Geboortedatum	84188	28423	33.8%	299	129127	50589	39.2%	268
Geboorteplaats	73573	23112	31.4%	111	139742	55900	40.0%	349
Sterfdatum	91920	28989	31.5%	298	121395	50023	41.2%	268
Sterfplaats	67873	21744	32.0%	112	145442	57268	39.4%	343
Sekse	92545	25854	27.9%	239	120770	53158	44.0%	299
Rubriek	72707	0	0.0%	0	140608	79012	56.2%	279
Claim to faim	100	100	100.0%	479	213215	78912	37.0%	279
Religie	16918	2159	12.8%	70	196397	76853	39.1%	285
Vader	300	300	100.0%	408	213015	78712	37.0%	279
Moeder	241	241	100.0%	419	213074	78771	37.0%	279
Partner	10	10	100.0%	691	213305	79002	37.0%	279
Beroep	26658	24711	92.7%	247	186657	54301	29.1%	294

Table 6: Indicatie van aanwezigheid van tekst voor verschillende informatie (per XML-file)

Informatie	Afwezig	Tekst	Tekst (%)	W/(tot. tekst)	W/(langste tekst)	Gem. aantal teksten
Naam	1	0	0.0%	0	0	0.00
Geboortedatum	30801	24969	81.1%	180	160	1.14
Geboorteplaats	34508	23664	68.6%	188	169	1.12
Sterfdatum	24762	20032	80.9%	140	129	1.10
Sterfplaats	36744	24663	67.1%	182	165	1.11
Sekse	19580	9017	46.1%	113	113	1.00
Rubriek	15435	11870	76.9%	131	130	1.03
Claim2Faim	76260	60035	78.7%	363	289	1.31
Religie	62260	47118	75.7%	319	255	1.28
Vader	76060	59835	78.7%	363	289	1.31
Moeder	76119	59894	78.7%	364	289	1.31
Partner	76350	60125	78.7%	366	291	1.31
Beroep	51378	36077	70.2%	343	278	1.28

Table 7: Aanwezigheid teksten, gemiddeld aantal woorden (totaal en langste tekst), gemiddeld aantal teksten (als aanwezig) voor ontbrekende informatie per persoon

Informatie	Aanwezig	Tekst	Tekst (%)	W/(tot. tekst)	W/(langste tekst)	Gem. aantal teksten
Naam	76359	60135	78.8%	367	291	1.31
Geboortedatum	45559	35166	77.2%	499	384	1.44
Geboorteplaats	41852	36471	87.1%	483	370	1.44
Sterfdatum	51598	40103	77.7%	480	372	1.42
Sterfplaats	39616	35472	89.5%	495	378	1.45
Sekse	56780	51118	90.0%	411	322	1.37
Rubriek	60925	48265	79.2%	425	331	1.38
Claim2Faim	100	100	100.0%	2332	1449	2.44
Religie	14100	13017	92.3%	539	420	1.45
Vader	300	300	100.0%	1046	690	1.82
Moeder	241	241	100.0%	1131	729	1.90
Partner	10	10	100.0%	2875	1546	3.00
Beroep	24982	24058	96.3%	403	310	1.36

Table 8: Aanwezigheid teksten, gemiddeld aantal woorden (totaal en langste tekst), gemiddeld aantal teksten (als aanwezig) voor gemarkeerde informatie per persoon