



Correction to: MELHISSA: a multilingual entity linking architecture for historical press articles

Elvys Linhares Pontes¹ · Luis Adrián Cabrera-Diego² · Jose G. Moreno³ · Emanuela Boros² · Ahmed Hamdi² · Antoine Doucet² · Nicolas Sidere² · Mickaël Coustaty²

Published online: 18 January 2022
© The Author(s) 2022

Correction to: International Journal on Digital Libraries <https://doi.org/10.1007/s00799-021-00319-6>

The following are the errors in the original publication of the article:

In the Abstract, third line from the bottom, the phrase “over two historical document” corrected to “over two historical document corpora”.

In Author Contributions, the name “JG M” corrected to “JGM”, and the phrase “MC helped in supervision” added at the end.

In Data availability, the first link was broken as the backslashes were kept. Now, they have been removed.

In Reference 49, the surname of the first author corrected to “Agirre” instead of “AAgire”.

The original article can be found online at <https://doi.org/10.1007/s00799-021-00319-6>.

✉ Antoine Doucet
antoine.doucet@univ-lr.fr

Luis Adrián Cabrera-Diego
luis.cabrera_diego@univ-lr.fr

Jose G. Moreno
jose.moreno@irit.fr

Emanuela Boros
emanuela.boros@univ-lr.fr

Ahmed Hamdi
ahmed.hamdi@univ-lr.fr

Nicolas Sidere
nicolas.sidere@univ-lr.fr

Mickaël Coustaty
mickael.coustaty@univ-lr.fr

¹ Trading Central Labs, Sophia Antipolis, France
² L3i, La Rochelle Université, La Rochelle, France
³ IRIT, University of Toulouse, Toulouse, France

The following references were missed in the reference list, but they have been cited in Fig. 1:

Gazette of the United-States. (New York, New York, U.S.A). In: *Chronicling America: Historic American Newspapers*. Library of Congress (02-Jan-1790). <https://chroniclingamerica.loc.gov/lccn/sn83030483/1790-01-02/ed-1/seq-4/>. Accessed on April 2021.

Gazette of the United-States. (New York, New York, U.S.A). In: *Chronicling America: Historic American Newspapers*. Library of congress (03-Mar-1790). <https://chroniclingamerica.loc.gov/lccn/sn83030483/1790-03-03/ed-1/seq-4/>. Accessed on April 2021.

Vossische Zeitung. (Berlin, Germany). Staatsbibliothek zur Berlin (11-Feb-1857). <https://dfg-viewer.de/show/?set%5Bmets%5D=https://content.staatsbibliothek-berlin.de/zefyfsSNP27112366-18570211-0-0-0-0.xml>. Accessed on April 2021.

Chariton Courier. (Keytesville, Chariton County, Missouri, U.S.A). In: *Chronicling America: Historic American newspapers*. Library of congress (13-Feb-1890). <https://chroniclingamerica.loc.gov/lccn/sn88068010/1890-02-13/ed-1/seq-3/>. Accessed on April 2021.

Le Libérateur du Sud-Ouest: organe régional du Parti populaire français. (Bordeaux, France). Bibliothèque nationale de France (3-Dec-1936). <https://gallica.bnf.fr/ark:/12148/bpt6k55631820>. Accessed on April 2021.

Les Affiches de Paris (Paris, France). Bibliothèque nationale de France (31-Dec-1750). <https://gallica.bnf.fr/ark:/12148/bpt6k10531388>. Accessed on April 2021.

In Table 3, the years were overlapped.

In Table 5 first line, 1A repeated twice instead of 1B.

In Table 11, there are various errors, so the correct Tables 3, 5, and 11 are given below:

The original article has been corrected.

Table 3 Number of mentions by period of time in the CLEF HIPE dataset

Splits	English										French					
	German															
	1750–1800	1800–1850	1850–1900	1900–1950	1950–2000	1750–1800	1800–1850	1850–1900	1900–1950	1950–2000	1750–1800	1800–1850	1850–1900	1900–1950	1950–2000	> 2000
<i>Train</i>																
ORG	8	56	70	74	–	–	–	–	–	–	–	–	–	–	–	–
LOC	12	84	105	111	–	–	–	–	–	–	–	–	–	–	–	–
PERS	16	112	140	148	–	–	–	–	–	–	–	–	–	–	–	–
PROD	20	140	175	185	–	–	–	–	–	–	–	–	–	–	–	–
TIME	24	168	210	222	–	–	–	–	–	–	–	–	–	–	–	–
<i>Dev</i>																
ORG	6	26	22	26	10	54	26	44	26	26	4	14	8	32	24	4
LOC	9	39	33	39	15	81	39	66	39	39	6	21	12	48	36	6
PERS	12	52	44	52	20	108	52	88	52	52	8	28	16	64	48	8
PROD	15	65	55	65	25	135	65	110	65	65	10	35	20	80	60	10
TIME	18	78	66	78	30	162	78	132	78	78	12	42	24	96	72	12
<i>Test</i>																
ORG	2	20	34	42	6	32	14	30	10	10	6	16	16	26	16	6
LOC	3	30	51	63	9	48	21	45	15	15	9	24	24	39	24	9
PERS	4	40	68	84	12	64	28	60	20	20	12	32	32	52	32	12
PROD	5	50	85	105	15	80	35	75	25	25	15	40	40	65	40	15
TIME	6	60	102	126	18	96	42	90	30	30	18	48	48	78	48	18

Table 5 Filter configurations used in this work

Filter	Edit distance	Mentions to filter by date	Middle candidates
1A	RapidFuzz W. Ratio	All	No
2A		None	
3A		Person	
4A		All	Yes
5A		None	
6A		Person	
1B	None	All	No
2B		None	
3B		Person	
4B		All	Yes
5B		None	
6B		Person	
1C	Weighted Levenshtein	All	No
2C		None	
3C		Person	
4C		All	Yes
5C		None	
6C		Person	

Table 11 Amount of mentions that match their surface form with an entry existing in the probability tables $p(\phi|lm)$

Splits	English			French			German													
	Mentions		Match Correction	Mentions		Match Correction	Mentions		Match Correction											
	No cor.	Multi		No cor.	Multi		No cor.	Multi												
<i>(a) CLEF HIPE 2020 dataset</i>																				
Train																				
ORG	-	-	-	554	266	465	473	247	82	83	172	173								
LOC	-	-	-	3333	2263	3022	3063	2009	1312	1338	1738	1767								
PERS	-	-	-	1343	429	734	829	849	161	187	317	363								
PROD	-	-	-	149	77	112	118	77	23	26	51	56								
Total	-	-	-	5406	3038	4348	4498	3209	1576	1632	2278	2359								
Dev																				
ORG	93	44	45	120	47	97	98	112	19	19	72	72								
LOC	332	210	211	851	617	759	767	711	494	504	613	625								
PERS	107	50	52	434	158	242	265	294	87	98	131	143								
PROD	16	9	9	37	17	30	30	35	13	15	29	31								
Total	549	313	317	1450	840	1135	1167	1157	613	636	845	871								
Test																				
ORG	45	20	20	105	58	90	90	99	37	37	72	73								
LOC	184	118	119	926	625	829	842	696	454	460	583	598								
PERS	46	8	8	271	90	161	167	188	31	38	57	72								
PROD	8	0	0	43	23	38	38	42	13	13	23	25								
Total	283	146	147	1345	796	1118	1137	1025	535	548	735	768								
Splits																				
Finnish					German					Swedish										
Mentions		No cor.		Match Cor.		Mentions		No cor.		Match Cor.		Mentions		No cor.		Match Cor.				
						Mono		Multi		Mono		Multi		Mono		Multi				
<i>(b) NewsEye dataset</i>																				
Train																				
ORG	92	13	13	54	55	352	106	110	308	311	165	21	22	60	60	72	5	5	29	45
LOC	837	398	399	673	691	1786	1174	1201	1662	1681	454	321	321	383	386	834	485	516	722	756
PERS	270	21	35	78	110	1057	244	254	660	697	73	28	28	55	56	395	48	63	153	183
PROD	82	5	5	38	42	108	35	37	87	90	3	1	1	3	3	146	20	20	95	123
Total	1281	435	450	843	898	3303	1559	1602	2717	2779	695	371	372	501	505	1447	558	604	999	1107
Dev																				
ORG	13	3	3	7	7	57	22	22	51	52	88	46	46	69	69	7	1	1	4	5
LOC	80	37	38	62	62	283	209	211	262	263	223	175	175	209	210	132	81	81	120	123

Table 11 continued

Splits	Finnish						French						German						Swedish						
	Mentions		No cor.		Match Cor.		Mentions		No cor.		Match Cor.		Mentions		No cor.		Match Cor.		Mentions		No cor.		Match Cor.		
	Mono	Multi	Mono	Multi	Mono	Multi	Mono	Multi	Mono	Multi	Mono	Multi	Mono	Multi	Mono	Multi	Mono	Multi	Mono	Multi	Mono	Multi	Mono	Multi	
PERS	31	2	2	13	13	163	59	62	116	122	96	27	30	54	56	49	14	24	24	24	37	37	37	37	
PROD	9	0	1	2	4	7	5	5	6	6	2	2	2	2	2	11	1	1	10	10	10	10	10	10	
Total	133	42	44	84	86	510	295	300	435	443	409	250	253	334	337	199	97	107	158	158	175	175	175	175	
Test																									
ORG	2	0	0	1	1	79	33	36	65	69	61	14	14	49	49	5	1	1	2	2	2	2	2	2	
LOC	78	47	47	65	65	481	304	310	442	444	269	184	184	230	231	121	75	81	108	108	113	113	113	113	
PERS	31	1	2	7	8	208	87	91	135	140	73	22	22	52	56	56	14	16	27	27	29	29	29	29	
PROD	5	0	0	1	4	15	6	6	14	14	3	0	0	3	3	13	2	2	11	11	12	12	12	12	
Total	116	48	49	74	78	783	430	443	656	667	406	220	220	334	339	195	92	100	148	148	156	156	156	156	

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.