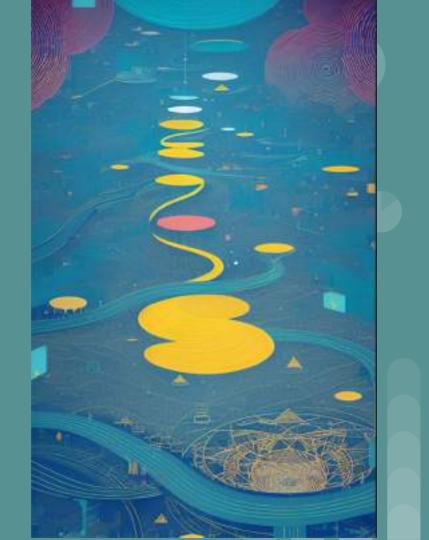
Natural Language Processing in Digital Humanities: a case study

Galanaki Maria-Anna, PhD student



Digital humanities & Natural Language processing

Digital humanities→ no commonly accepted definition (McGillivray, 2020

→ no standard description of the relation with Natural Language Processing (NLP)

⇒ an area that combines computing, digital resources and the traditional disciplines of humanities (Helsinki University, n.d.)

Suitable topics:

- Text analysis and processing related to humanities using computational methods
- Thorough error analysis of an NLP system using (digital) humanities methods
- Dataset creation and curation for NLP (e.g. digitization, digitalization, datafication, and data preservation).
- Research on cultural heritage collections such as national archives and libraries using NLP
- NLP for error detection, correction, normalization and denoising data
- Generation and analysis of literary works such as poetry and novels
- Analysis and detection of text genres (The 2nd International Workshop on Natural Language Processing for Digital Humanities – NLP4DH, 2022).

Like-farming research on Facebook & NLP

Emotional like-farming research on Facebook requires:

- -creation and curation of a multimodal dataset
- -extensive annotation (manual or not)
- -error detection and denoising parced annotation entries
- -use of appropriate python methods for a sociolinguistic text analysis & data visualisation
- -error analysis on Greek NLP models to revise practices
- -error detection to interpret design practises
- -NLU methods for sentiment analysis
 - ⇒ identify, interpret emotional exploitation like-farming characteristics
 - as fake news genre characteristics

What is like-farming?

No relevant research

Like-farming⇒ a little known fake news practice (Daniilidis, 2013)

⇒ part of engagement farming practices (Facebook, 2012)

Different from response farming→ click baits

Like-farming posts → stand alone Like-farming variations

→ spam, scam. fraud

Emotional exploitation

→ satisfy attention seeking designer

-Genius Puzzles

Some

-Facebook competition or giveaways

-Charitable donations ...

-Religious/political affinity

-Superstition (Charles, 2016)

Focus on emotional exploitation likefarming

Emotional exploitation like-farming news: fact checking organizations (Hellinika Hoaxes)

⇒ Posts mainly about vulnerable social groups (disabled people, sick children, elderly etc.)

Aim

- \rightarrow not at provoking social concern and action
- → turning them to objects of exploitation



Emotional like farming layout structure



Multimodal structure:

- -Group or Page logo & name
- -Time stamp
- -Friend tag
- -Post text with emojis/emoticons or not
- -Central-dominant image
- -Superimposed text on image
- -Social cues-metrics (like, share, comment number)
- -Reposting comment by intermediary
- -Comments ⇒ primary metadata in data set

Data set formation complexity

Data set structure is not limited in these metadata entries.

The fake news design layer adds complexity in data set structure.

-existing data sets→ primary, secondary and tertiary label tags with mixed criteria

→ mostly monomodal research approach

⇒ little relevance to our multimodal research subject

or debunked by fact checking organizations

			- 0		*
References	News Domain	Application area	Type of disinformation	Language	
Bertado et al., 2019	Technology	Fake detection	Fake reviews	English	
D.bioga et al., 2016; Gissemidio et al., 2018; Bonckelli and Marcelloni, 2015; Alkhodair et al., 2015	Society, politica	Rumour detection	Ramous	English and German	
Nitrs and Gilbert, 2015; Soumya and Shankar, 2015; Bondielli and Marcelloni, 2015; Zhang and Ghorbani, 2015	Society	Veracity dassification	Rumours	English	
Santia and Williams, 2008; Sowmya and Shankur, 2019; Bondielli and Marcelloni, 2015; Beis et al., 2019	Politics, society	Veracity classification	Fake news articles	English	
Tatchini et al., 2007; Sovereya and Shankar, 2013; Bonde Bi and Marcelloni, 2013	Science	Fake detection	Hoaxes	English	
Wang, 2018; Bondielli and Marcelloni, 2015; Zhang and Ghorbani, 2015; Aldwain and Alwahedi, 2003; Torabi and Taboada, 2015	Politics	Fake detection	Fake news articles	Ergish	
Vlachos and Riedel, 2014; Bondielli and Marselloni, 2019	Politics, society	Fact disolong	Faké news articles	English:	
Thome et al., 2018; Bondielli and Manselloni, 2029; Torebi and Taboada, 2019	Society	Fact directing	Fake news articles	Erglish	
Ferreira and Vlachos, 2016; Bondielli and Marcelloni, 2019; Torabi and Taboada, 2019	Society, technology	Rumour detection	Rumours	English:	
2 Shu et al., 2018; Sondielli and Marcelloni, 2005; Zhang and Ghorbani, 2009	Society, politics	Fake detection	Fake news articles	English	
Morrie and Aslait, 2017; Zhang and Ghorbani, 2015; Ahmed et al., 2018	Politics	Fake detection	fake news articles	English	
5 (Burloot and Baldwin, 2001; Home and Adal), 2011; Zhang and Ghorbani, 2019	Politics, economy, technology, society	Fake detection	Satire	English	
7 Harne and Adals, 2017; Zhang and Ghorbans, 2019	Politics	Fake detection	Fake news articles	Ergish	
il Torabi and Taboada, 2015	Society	Fact checking	Fake news articles	Trylish	
Ott et al., 2011; Viviani and Resi, 2017; Ahmed et al., 2018	Touton	Falle detection	Fakereviews	English	
Redel et al., 2017, Aldwari and Ahvahesii, 2018; Thota et al., 2018	Politics, society, technology	Fake detection	fake news articles	English	
2 Posadas-Durin et al., 2029	Science, Sport, Economy, Education, Entertainment, Politics, Health, Security, Society	Fake detection	Take news articles	Spanish	
l Dutta et al., 2019	Politics, society	False detection	Fake news articles	English	
5 (Rashkin et al., 2017; Barrikn-Cedeno, 2815	Politics.	Fact checking	Fake news articles	English	
7 Barrián Cedeno, 2019	Politics.	Fact checking	Fake news acticles	English	
9 Nemgaard et al. 2019	Politics	Fake detection	Fake news articles	English	
0 long et al., 2013	Politics	Fake detection	Fake news articles	English	
1 Papadopoulou et al., 2009	Society	Falle detection	Fake news content:	English, Russian, Spani	
2 Bouldou et al., 2018	Society	Veracity plassification	Houses	English, Spenish, Dutch	
3 (weet al., 2019)	Politics, society, crime, sport, economy, technology, health	Fake detection	Fake news articles	English	
4 Derg et sl., 3007	Society	Clickbalt detection	Clickett	Chinese	
5 Tametal., 2029	Politics, technology, science, crime, fraud and scam, fauxtography	Numour detection	Namours	English	

Larguage	Size	Physical news content	Rating scal Media platform	Sportz	nei Avalla	bilit Extraction
Z English	18,912 reviews (5,456 fake reviews and 3,456 trastful reviews)	Test	2 values (f Mainstream	Yes	Yes	No
Parama vinces						
4 English and German	100 rumpurous conversations (355 are true, 68 are false and 351 remained unvenified) and	Text	3 values (t Social media (twitter)	Yes	Yes	No
5 English	60 million streaming tweets and 1,049 events annotated with credibility scores	Test	5 values (CSocial media (twitter)	Yes	Yes	fes (Octo
l English	1,283 Facebook news from 9 facebook news pages (75.18% mostly true)	Test	4 values (/Social media (facebook)	Yes	Yes	Yes (Sept)
7 English	15,500 posts from 32 pages (14 conspiracy and 18 scientific) among which 8,523 (57.6%) are hos-	Test	2 values (if Social media (facebook)	Yes	Yes	Per (July 2
1 English	12,836 human labeled short statements from POLITEACT.COM?s API (1,856 pants-fine)	Text	6 values (c Wainstream + social media (facebook, twitter)	Yes	Yes	Fes (2007)
f English	221 statements from two websites (politifact.com@and Channel 4)	Text	5 values (t Mainstream	Yes	Yes	No
© English	385,445 claims extracted from Wikipedia	Text	3 values (: Mainstream	No	765	No
Tirgin	100 claims, and 1,585 associated article headlines (47.7% true, 15.2% false, 37.1 unverified)	Test	I values (t Mainstream + social media (twitter)	Yes	Yes	No
2 English	422 news (21) fake news and 213 real news)	Text, images	2 values (f Mainstream + social media (twitter)	Yes	Yes	No
3						
4 English	325 stones from political news websites	Test	3 values (r Mainstream	Yes	Yes	Tes (2034)
5 English	4,000 real news	Text	2 values (r Mainstream	Yes	Yes	No
6	sangles and 333 satire news samples					
7 English	2,283 news samples from Facebook	Test	4 values (-Social media (facebook)	Yes	Yes	Fes (2016)
8 English	1,692 news article (1,386 from BuzzFeed dataset and 302 from Snopes dataset)	Test	4 values fc Vainstream	Yes	Yes	No
9			false, and (fully) false)			
1) English	800 reviews (400 truthful and 400 gold-standard deceptive reviews from TripAdvisor)	Test	2 values (t Social media (TripAdvisor)	Mo	Yes	No
I English	49,572 articles (3,678 agree, 640 disagree, 8,909 discuss, and 3,654 unrelated)	Test	4 values (Mainstream	Yes	765	No
72 Spanish	971 news (450 true topics and 480 fake topics)	Test	2 values (t Mainstream	Yes	Yes	tes (Janu
3						
4 English	6,337 articles (50% real, 50% take)	Test	Z values (t Mainstream	Yes	Yes	No
5						
% English	33,363 news articles (18,483 from PolitiFact.com and 21,580 from unreliable sources (5,750 frus	f Text	6 values fc Mainstream	Yes	Yes	No
2 English	31,294 articles (5,797	Test	2 values (g Mainstream	Yes	Yes	No
26	propagandistic and 45,557 trustworthy)					
29 English	713000 articles	Text	2 values (t Mainstream	Yes	Yes	fee(Febry
ID English	3472 (1397 fake (1645) tweets) and 2005 real (5665) tweets))	Text	2 values (f social media (twitter)	Yes	Yes	Yes Jianu
English, Russian, Spanish, Arabic, German, Catalan, Japanese, and Portuguese	380 videos and 77258 tweets	Videos, test	2 values social media (YouTable, Facebook, Twitter)	Yes	Yes	Fes [April
22 English, Spanish, Dutch, French	15629 posts	Test, images, videos	2 values - social media (Twitter)	Yes	185	Fes (2012-
S English	39700 articles	Test	4 values Wainstream	Yes	Yes	Yes [April
4 Chinese	14922 headines	Text	2 values Mainstream + social media (Weshat)	Yes	Yes	No
5 English	1022 rumours and 4 million tweets	Test	2 values social media (Twitter)	Yes	Yes	Tes (May



Data set structure

Data set under construction

→ all instances in data set are **fact checked** by researcher

or debunked by fact checking organizations

Data set formation stages:

Identification: 479 fact checked like-farming fake news (2017-2022) from 49 different identified like-farming post resources

Screening: 436 identified emotional like-farming fake news posts in Greek or automatically translated for Greek Facebook users. 43 posts excluded due to fact checking ambiguity

Eligibility: Maximum 5 posts selected for each theme type from every resource (n=153). 3 posts excluded due to narrowing down the study protocol

Included: 150 posts

Data set structure & annotation complexity

Data set structure is not limited in these metadata entries.

- -emotional like-farming posts with:
 - → stollen, manipulated content or repurposed fake-news content
 - → content similarities across former ghost interconnected like-farming pages
 - → internet links possibly stripped from content after a while
- ⇒ complex production and consumption design that multiplies annotation criteria and data set curation standards (post date, like-farming page status, watermarks etc)

Annotation and processing complexity

Like farming understanding

-intrasemiotic & intersemiotic analysis→ annotation indicators based on:

Systemic Functional Grammar

Visual Grammar

Multimodal Critical Discourse Analysis

Computer Mediated Discourse Analysis

-Natural Language Processing→

-current python library/toolkit affordances (e.g. Pandas, NLTK, Textstat)

-previous NLP research on other similar type fake-news, mainly click baits

⇒ transdisciplinary & interdisciplinary approach in

Inpumpored Trial city bills	Tracestory, A., Parentinos, K. Remark, S. & Re	PROFIT PLANE	Control to the Land of the Land of the Park	The state of the second the second to the second state of the seco
Apriluis Missor	Mayakimpoc	CONTRACT	Nelectoria.	this proper transfer to the policy continuent; a finite contract, having in the state, through the state, 2007)
bram May	Mogreson		Policipis.	
Queens;	Delain sengage		Trave surgicul	
Ninterana	Davel dellinger			
Nec flooder	Neithburg wirk Wyne Miliesz and prinner between a	(militing)	House had annual print, Supragative Warned Says 200 Strains Disease SCOttomare	
SivesOr (Asserdences)	THE PERSON WAS A STREET OF THE PERSON OF THE	PP-FF graintery exhapsion/lines, LTM, Wulf, TTM;		
The World	Equil splan-banqueives anningen etalos			
Addutt pophilopal	Einform Minagers modernic from one specif		Professio, military magnetics (Spec, Transmission), Habrer, 2006.	
Adue overhiero	"mandature; later," us "tols firms occupably	7.1.7.7.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1	Spirit March Strate (SSA)	Direction Result, special discourse representants between plants of the contract input, which (Direction E. St
200	Instrumentary (117)	W.	HAMES STREET FOR STREET, STREE	Teilia promutemati Mariach, Lanages, Sautan, Voesb Protes, 25(7)
Transcriptoracing and	Les Rommandians, Elementosis, Microsia, O. v.	Daniel Maria	Note: Note: 10 th company Suff Company Specification Specification	
Dynotronis, devot journing	Mall recomposes - Case baseons; DIT or no		Not former of fragmer region for human security	
Avadopii min film)	Julies C+F point, o'Noti-es your man Pep	Control of the Contro	The second secon	
Bigans	Mode spuripum, Mac: runnima, Navecc;		Substitution and the State of State (SSS)	Proprietario espetar en Bargos espetar enpelan completo entranca espeta Dec
Annuar FS	- Programme and the contract of the contract o	solitivist state.	Tourney and Company of Company (1996)	administration in the melided science suffering all bedood melines in order one, subject over
4.000,000,00	Tegs averaged, average is sponsored	125	Interest references with a result that	
NET Divelopms	Property to low like you plant lightings	EL.	CONTROL OF THE PARTY OF THE PAR	
NOTIFICATION OF THE PARTY OF TH			Approximate and the second sec	
BADRATTA popopopul Familiana			To Prumbradio (publicand recipiosis front districtions Paris Statemente, 2011) (1900)	
Josepher As at length last past			Abblic terrolitimic incitor kina solar, Disa, Tel S Benerjes, (CC1)	was to the same of
Indidentifymenology (pasto	apped		Spinister) arises (Supershalter) managety to recently a large inter-executing and around (Disa, Fel & Bare in	MD .
			Anne rigary mires Migarray or perceptorpoint form to a late, 2000 forces & latery, 2000 days Recommendings	
ACCOUNTS NOT THE REAL PROPERTY.			Tompeninson biorcorporaspect art (for myclim)/vice & date: 2005/vice & bate; 2001	
Stor analysis			A CONTRACTOR OF THE PARTY OF TH	1
			Boyana spirmer letter, evenethys, ysunist form, moreonium as signiferant, conferming-metionium is basic. (22)	
Manufactic responsibility	-	E permittante ettek barr (mitac oce ocum)	A female M females of Sentence Service (SULT)	Εξερνική από διβλιογραφία και το χαρακτέριστικό σε τέλλους.
M		Μυρότερη πικούτητα πληροφορίας απέ τ	umusic cubit Americale Essentic reimprodopies: Assetsionates sourtheese	пікпрофоркахії чачії і рементатвіл (1994), абритає ампичицієє (Chen, Contray, and Rybin, 2011 «автапрофії тіє афтунцатиліє башіє тіє аметрацивніє пирацібає
			Territor specimen	parintipuo ses ngolekrynt authonisparinte (Moint-Rosphovska, 2018, 188), фтрайорія на падам
			Auriflyni nei Airmes	
		2. Eventurian Antipology	Toomselik Klemek	
			Elementarial anniform mon Earth	
		5. Eministrative non-partial system	Property Symposius (Control of Control of Co	erospeix repolleturintyms; surropert osuetsiλλουση του δημοφιλούς Norlage (1997: 68)
			Emeritéries que rétauries d'une	
		5. Translation indposes/Audions	Australgues regula tim teprophylining Egylyni ungelleking	
		2. Trimberror Francisco Control	Every inter unappeals our prophenius Sodure:	British Lubrary (Dutrers, 2002, Mores-Kosemwide, 2012)
			Settement springer straffension improdissportation	
		Cat to the	Monteyoryani ya 10 "vinda"	ursig, myneodrage, + speedingusyni neu sudkroun my sirsen; Pelauriamyria, 2016.74
Minisphero, roacepolism	_	N. Shirippi vite	Ελάφου ή αυτόποιακό παραγράμεσο Τροσωποιός, πουσαπότριστούς, επικοπότριας ή ανακδοπικού τέπου αναπομαπτέπους.	(Rememann et al., 2018)
		6. Eyttytumiská omtyloskovitnich molgala		
			Amiculate	
		Arriver techniques missing	Transport emergeneration where	
		THIS DOWN	Highest Telephonologic Seatomer, 2018. http://doi.org/10.1016/10.0016/	*
		Parkets tapitames	Prophotoly regions their years money many the real profession and the Province of the property of the property and send-orang papers.	
		Mit motoris birers	Therefore industry	
		Minimum Myrry Myptoricani	Therefore industry	
		Mit motoris birers	Parentener (n'aprent Kambanon millos en reliablesse Parentenes, prophysioli antonioni tapoglisses	
		Min sometic bilance Mingraphicane Epoplandingsta	Therefore industry	

Using Pandas methods to:

-display data frame columns,

types, shape,

format,

length and unique column name

						prises of recipio. No artistore in outcomes.	project with treaspire. You are enforce to our company.
н	Tracking Secure	1000.00	0004	100	10411	Typu amonthe type at motified God had all the statistic light part are from who with Arter	. If you are unlike type amonthisy facilities of the 0.01 offerent are those who with loose
p	April .	40.0	TIMOLATO AUTOMOCAS	, turn	1640	From every arteriority my last, prescribed the CRX and or not have for the former of the passer, and have it is the name of last.	From ethic amount to the paper from the child and as not found that the control of the control to the control of the control o
ir	Miss los Notes for No. Missespers	900.00	peax	arty (M)/65	MARKEMEN	The phoins is before the parament," The count who issued the photo part for influency. The strang count was almost at the frequent of stops and studing from 1 days do family married count to self attention, against a studied to self at teams. Such against a finish and the married from the first form of second to the other flows and and the first flow and the self-parameters and the self-parameters are almost to the self-parameters and the self-parameters are self-parameters.	The plants active/serite general () The ryne () who tool the phole and the billioning This ribert) man any administration for froger for cells in 2014 pages on any 2 tops any and regard from beet for many brane. (Ame of obtains immore that the cell many drawn for the pages or short to see celling on the part from drawn for the pages or short to see celling on the part from facility plant. Among their more formatty, they may be any of the pages of the more formatty. They may be seen of the companion.
8	Naparna.	146.00	200	1,00	ine	wepting for wrote. It has an whice the chronic fun-	The motive help daughter in this way to been corner. More will do excepting to the child. It if there will wish a loss connect of year word, a page of states. A year of wall for year will had been you are all natures?
pi.	West	900.00	TANSARD AUTOMATION	15,848	Notes	Type an other type annet May Doctried at the out of bloom an filter of the other Anne.	Figure or other type entertailing limit that all the arts of those are from other order from
	7790	3400000	HANGARD AUTOMICAZ	5.60	7611	E you and notice type are noticed by Control of the Arts of Report and Royal Arts of the	Figures make type energylap (and hear of the act officient and formally write.
и	majorne Zorani	1000.00	900	tun	1940	Reserving to my traplate. We are a very difficult abuston. In this the heige are a threat place players and user fulfi.	Please projective try to the lates. We alway any official shapes, N the fire steep on a travel pain payers are seen in m.
m.	No.	0.0	THEO, KTO ALTOMOTICAL	160	7600	No record a report to hope Proper ISS/SNE 1810/SNE SNE MINIST COMMENT, AND VALUE AND A HARD A position	
18	House	100000	TEMPLETES AUTOMOLIAS	AAR	1940	Fyrix are online, later direct of they disclined using partiest (Albert State of the partie direct	Figure are ordine high directivities that their army patient below bless from with white bless
p	Marin: Tons for In: Names	750	2000	MUN.	mate	N triplet the time	No line passe time to the dis-
н	TOWNS .	100	THRELETY AUTOMOTIVE	1,46	Manistration	Conference on more of the 20 years, begans all from arts for not found of the conference of the confer	Said Tempor or yours of the 20 years Purples of Mass of the St. No.
	100 200	100.00	960	acte (NORT)	continues	My reason in Albert Minday is my first tage of word, from letting in an a account of fermion medical critical and it is a public graph many production on the country on your and format in which the fermion is in first. See interest an approximation of all recognition	Wy earners Simul of halos is vig that day of word, from bally it are a district of halosed healest is those with TEL collaboration was quantitized the decision i propose self-lives, I worked and treated at an hore. On I decision an approximate of Sys & Life.

-change index,

drop columns,

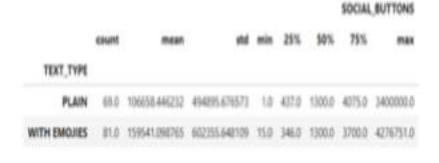
delete NaN values

110	105000	60x Ty. Tyry, for Ty. Riconyme	300.00	900	Wh hos	2000	-Chir rafige di folicio Seron tapagojal, in folicio Seron gayesi kao a physi Seron ta tadido trag. O ten talisama a tidiago, si talimini ratipioni, ta milienta tapagojali kai ta Seron o fraz stro diliki, si rikelana addimini te, tydi, turi tyogguni min a gruejdi pityonin trag tajanic, turi talihor tras in diaga attiv dayonatani.
154	541	tun.	561	54	tak	345	MA
116	. Ne	Wh	NI	NA	166	141	MA
116	166	165	No	565	166	346	No.
w	:36	W	56	366	186	M	NA NA
# 1	II Jenne	0					
ar							
	POST (ME)	NAME	SOCIAL BUTTONS	LANGUAGE	THEOLOGY	THEME, THE	HOT, IMMOLTEST
	2019/2021	To Timesorie	1074.0	TANKATO ALTONOCAL	AM	Harts	jako igen 5 kompilarno spomogilisto pa no vigo, par je Olista solonii kana strprzeimych Anae (vija
t	zruzo	Davids	97333	THREATS ALTORITICAL	906 306	Aud	Το μικό για στο γετάθηκε χέτε τρέδησε πρέξησε αγώς επεροφού δεφρεί οι διουπούσε με αλτέρδο χέριστος αδιούε τος υρισέα το μετάσει το πολέφου με είναι με το πολέφου χέριστος το πολέφου το πολέφου
1	19/1,002	Makes	7900038	THISLITE AUTOMOCAL	Nan	Medi	As then writer triproprint pays which is limit on Hammion Limit to conflict of substitution on a polyton $Autros$
1	HINGS:	Inten	42925100	WASLESS AUTOMOCKE	900 900	1666	As sint as the Thirty-Sopjets in a Apily-Silk is that we have the softent in
	261600	Mr Drave	24000000	THISLESS	WN 906	PORT	Parentyphin se no oblikel peint Dric vs Aleyl amil (no polyne Ariste

-apply a groupby function

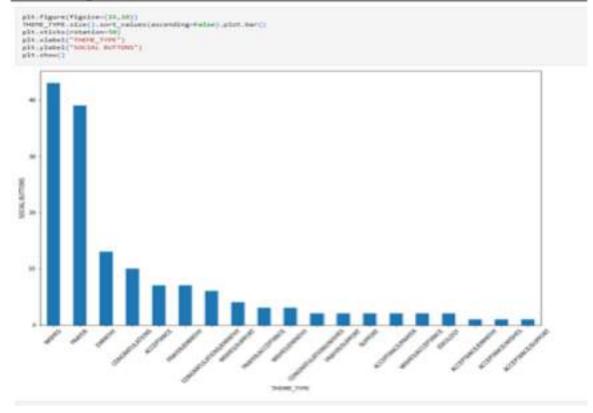
and perform a descriptive statistic on axis

```
# Groupby
TEXT_TYPE = df.groupby("TEXT_TYPE")
# Summary statistic of all
TEXT_TYPE.describe().head()
```



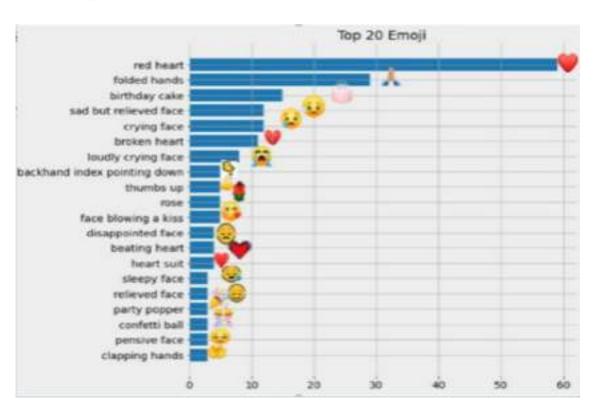
-print indexed column text and concatenate

-visualise sorted values using matplotlib.pyplot



- -decode,
- -count unique,
- -total count,

plot statistics of emojis/emoticons



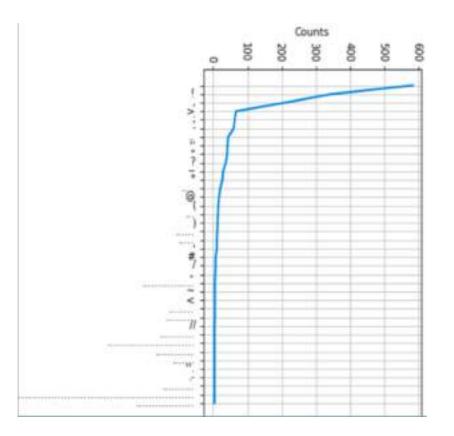
-join -split -print text length

ded to When William and page of the first process are not to be proceed an explorer date. Table Agric I Annoquillatis ausmaggiffitte par tar wing you 🛡 100 Mits, followed blank to be wing you 💌 100 Mits, followed by your way and will be a providing on the control of the part of the par a binding and wheth bearings in a longer to be the strong or the dispersion and of the contract of the strong or t a public A to Spengativity was not obtain you the State of the State o to the or or discover the law is discover the law is also be at the contract of the contract o a nighter failer as the first on providing to posite to accordance or upgate that they asked with they according to posite the first on according to the transport of the security of the secu or exists, a force, \$0.000 for event toke tapping are supplied as the thinker and tapping the event factors, by markler are the time of the time of the time to be the time of the time of the time to the time of the time of the time to time. as the desirable of transporter. The a form has the desirable for any or extra the contract of transporter and produce of the contract of transporter and tran a suppression of the first pair are also are expect, examples, or also in high one of the parties of the partie a persection are trike by expell on principles, preserved, that their plays now we expelled, no preserved on principles on against their paper on deposits on their paper on t we, prime to dright till my fifteen required and till needle on pain and manachine required on capital man required to require the required till manachine required the manachine required to paint and manachine required to require the required till manachine required to prime the required till manachine required to require the required till manachine required till on other popular transports and the presentation to final transport and the present and the pr to complete adjust no telling an agreen state of the new Workshopman and a transformatic number of the plant sects nine and meaning deliberal to appropriate the party of the plant sects nine and the plant of the p has not about the specific to the telegraphy and it has been part and it has not part and in the part of the specific till regions which the telegraphy in the specific till regions and the specific till regions are the specific till regions and the specific till regions are the specific till regions and the specific till regions are the specific till regions and the specific till regions are the specific till regions and the specific till regions are the specific till regions and the specific till regions are the specific till regions and the specific till regions are the specific till regions and the specific till regions are t He six or on an estimate in natural to find an indicate per foliated per foliated per foliated per plants have provide an engage or agree to the per plants for the per plants are agreed as a provide per plants. The per plants are agreed as a provide per plants are agreed as a per plant are agreed as a per plants are agreed as a per plant are agreed as a per plants are agreed as a per plant are agreed as a per plants are ag tigates, etc. this or recognition to principle to princip Station III place on the Ear No. and all other contract to the contract of the or students, address; on adjusted to different and differe man - 179 page to a medicur and manifor all a facility and a facil or of degree to account to account to account the parameter or account to account the parameter and the specific terminal account to account the parameter account to account the parameter account th as better midden ... as on it that the place are with an important many or man out when annual control of the particular plants of the particular plan Nation of the state of participation elegans, appropriate page article representation on the financial page of the state o And advertise transmissing and their receives the finite and extension to their new arrangement to detect and change for page between the detection of their receives and their finite and their per or placket of historical. As the exactor, compresses, expresses, expresses for an exactor are exceptions on the first and the first and the second are expressed as a position for the exact to exact the exact to exact the exact to exact the exact the exact the exact to exact the exa while a gold with annual control of the factor and as willow elementaries under this or from the Representation of the contract o A to but draw on which it designed about special and excitated an excitated to excitate and to become on the special and the s The piles by appropriate the high on the attachment. The expectate that the expectate that is expected that a plane on a special property than a property that is expected to be expected to a plane of the expectation of the THE THE CONTRACTOR AND ADDRESS OF THE PARTY Also be Millers, and the major, the part of the part o to in attent on the disconnection made with a finite or Representation than an adjustment on the adjustment of the contract of to explorate an among to the element from the element (because the element of the defined, and finite are unitarial according ton applicate days. One has required most one in the pass on required finite are unitarial according to the first properties of the contract of the pass o a plin balance, not report country of the same terr country out if you are contracted out to be a series of the same and the same and the same are the country out of the same are the same and the same are the same allow fights beligned about a spirit you page and Audaba . Deltain prompts and display the ten begand fights and the complete on the complete and the complete of the complete a ten effection as first. Table mean per arter. Name tables a rate on page on training and region and region of the respect Early as the "Affection and the region and the respect Early as the contract of the region and the region a ent tablican symples was promised was promised and registered and a to describe address which pur the effect for the effect for the effect for the effect for the effect on the effect on the effect of the effe to play your money advances, notice below pages your states are good 11 Nationarrous dates for your five or a flad you to dated the property of the pagesta physical vice and deliverage and pagestally account of the pagesta physical vice and the pagesta physical vice and the pagesta pagestally account of the pagestally acco AND REPORTED AND AND ADDRESS OF THE PARTY ADDRESS OF THE PARTY AND ADDRESS OF THE PARTY ADDRESS OF THE PARTY AND ADDRESS OF THE PARTY AND ADDRESS OF THE PARTY AND ADDRESS ages ... and we fill Might men Walter their grinn grinning and an article gain or described gain or described gain and described an agent of their agent and their grinning and agent and agent of their agent age 1 1 or in lay me from your grider or and the party of a sale as explore any miles are with any o to have through any in another income and in any in the second and in any in the second and in the sec

NLP using NLTK

Cleaning text

- -inspect nltk stopword list insert new list & set final stopword list
- -lowercase when needed
- -extract http & punctuation using RegEx, filter text using for loops +plot

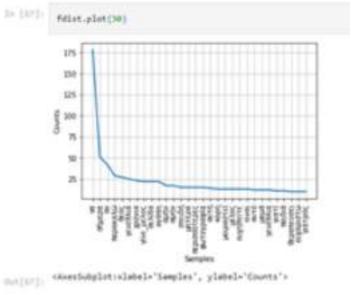


NLP using NLTK

-get frequency distribution of words or emojis/emotico ns in cleaned text

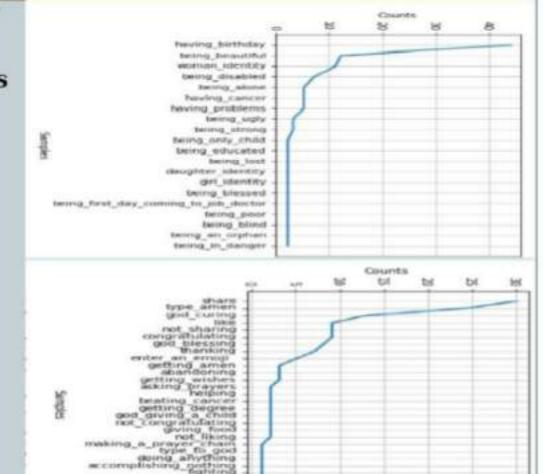
-plot & interpret





- Passive representations of vulnerable social groups
- Constructions of otherness and group identity
- -Relational processes: narrow identity representations

 The confined in body
- -Material processes:
- Mediating for the vulnerable with Facebook tokens
- \blacksquare The non responding Others



thanking god share picture

god helping

surgery about to happen

NLP using NLTK

- -use Bigram Collocation Finder
 to get bigram association
 measurements + frequency
 distribution+ dispersion plots
- -use n-grams to inspect text+focus on certain features
- -use concordances for SFL+CDA on targeted features

```
E - Text (text_tubers)
 T. concordance! 'wy' !
Displaying 25 of 42 matches;
 aw dyere 5 deutepohents moosevysHeite vt.
va edvenue pen advotto apprengts apriv as stars on line sharepolaritate aprix still
officeric subdeputate times yaddesic made no atoms on line abpreciatoryteen for surju-
oto Saudrio the continue was time dry on Bay Speve poli too Ba thy metalice dis-
aydva Adepun Bétochnyn sensuls purlou au Egetc à Beirepäkents un voderts tou ff
Elect ve pac Bhigory emergi pac nyahi ny etner enyahiny yenye '' se enyaptarin
 dys to many ter the mesoscopy man and on quarter variations; you us to have your
Air aproxity offices who the the many poor out or figure subby applied consumerations was a
were due nameddes recordinate experience on the Spinsois decide while you expired on
dystractors compasseers sty capts pay av closs online objectsophrydres apply eithe
amortonia ani updata kua despho aydhia ao Bahata pia sponsigh kua ayte meo pu p
adout on atom autual mos yaddoos apris no store on line wireipologists apris siff
trampidate to don reprodutes strates or store on line stratechardors appressed
table taketumm the totalent again on 18 me our makket on appropriate years also
per duto da Speite oto un unu meir es ou a Cut pou ta édepe dhe avámoda, eva d
λά και κοινωποιήστε το δυτογραφία της αν δεν ντρέπεστε όπως άλλοι κακοί άνθρωπ
withou out in the ware him maintenaction on arm spaces struct oboded worken him the
two apopen marger you appear propers or bee traditions on entremotionis to due
maken and grant may return may be purpled out to the purple and the contract of the contract o
matta pou par maste pas suge impunabil su differe escate to pupike pou provedique
 dyn to aphy was the mooneway now chi or bursed variations; you we to him vive
τώ, πείτε ένα χρόνια πολλά στα αγάλια αν θέλετε γίνα μέλος στο μεταδοτικό είσε
 segment fitter ver subgettens tor potagos or otres such seat or apartificate seat day
shoptions toy pateon by sives subd can by pperaceran part delime to paper too it
agrae tour was pass other Some post Bayers no Bakare natural natural natural tar Aprilia p
 L -- Tent(Chank Edhans)
* consistence [ ] per ]
tringlaying to of to matches;
oc va cukayet autooc mar ypagour wyar yn Byette ympic agakta wyny marcadaraye
- Access paper on equivalence " agree " on Parity paper egibts made surrobbourse
Singue own eric alleftion mides addle person on dispers yants on updates don moto pri t
c pa divers gaple on aphless for aphr pa so agregate, encounting to decoupable
se en Autospinhia stenut bilaut seu seu un un anu neit du Apelte mepissétépet sindu
a av Békers pia spoosvyší áva spáv sto pa pou metc de Bpeite septoodtepa elote.
year sounds you now sover you evenue an exercise teas upon you ago, not not
these retranscripts muto to bests one by how many at my of Col post to apope the
n dynas not tondhe pou vive pilos one po pou mess va negroularge ev va surve
ena de en una nepresóraga unua póbas era pa pou mena alesa abon cumpiedantos mondi-
tost sumpostretor on attest pulhor thus an also make on Atyo Bu minus and extertedu-
topotepec duto yere adhoc est dow oto an app mete sadvo as synthem hive ohe to
Now Alves on Olove past yive pelos one po por more year reproduction stadows pe
symmetriffensor and to Albo maidta on on pry to notic stance althoughts athe pass dige
```

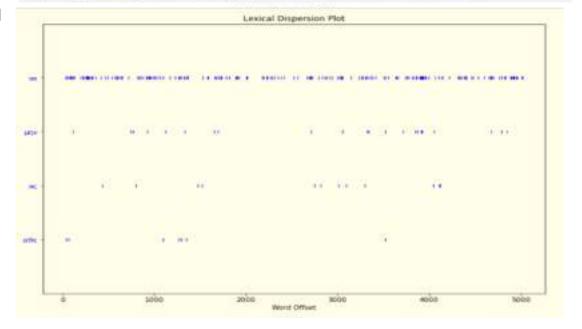
from witk, twet import foot

NLP using NLTK

-use dispersion plots for getting a group understanding of targeted

features

```
z the following command can be used to increase the size of the plot using width and hight specifications
plt.figure(figsize=(12, 9))
targets=('vo','μην','oc','eife')
dispersion_plot(text_tokens, targets, ignore_case=True, title='tanical Dispersion Fiot')
```



NLP using Textstat

- -sentence count
- -readability scores
- → even wrong outcomes
- give an understanding about
- like-farming production design
- & Facebook algorithm circumventing practices

```
1110-1416
        Text stat governey for their data maximum toxic)
        AMARAGUALLY Conveyous Smoot upon all the open bests
        Affirmal ages, will the above firsts, externs the authorited action great small required to commerciate the total
        RENGERAL Floor, parent allians the source to be returned as a first definite to false.
        tactutet test standard/hest data sedies toot, first materials at
        Actificity pages with
        "Title and $6% grade"
        Manufactivity Commission manufactor will the stone again
        Wheney were not the above here, returns the authorist colonia grade bond required to conservant the best
        Windformal Flood action allows the sames to be returned us to Plant. Definite by Palos.
        testatut test standard[test data atelese book, float actoristalias]
        newfolk years old
        "3rel and 3rd grade"
        PERSONALLY CONCERNS SHOOT ARREST ATT The second process.
        Without made off the about tools, retains the authorite count grade tools required to communant the tour
        Minthons from adject of time the come to be returned as a fight, Defaults to false.
        Testiful hest standard[test mits earlies foot, flost mitselfslas]
        --- th-il swars old
        "50th and 80th grade"
```

NLU using Vader & TextBlob

→ text cleaning for TextBlob, Vader sentiment analysis (NLU) → interpreting data from libraries

* different training process

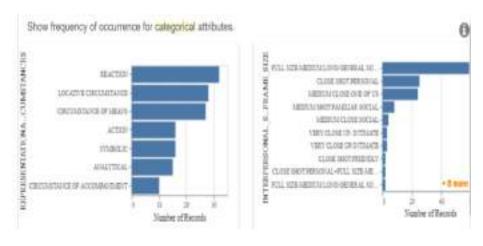
```
print(df["Emoji test VaderSentiment (ABEL"]-value counts())
             127
MADOR
Western
              22
#noutral
#Nome: Plain text VaderSentiment LABEL, dtype: intid
-- dps εδώ τα amojis αλλάζουν προς το θετικότερο άνα πρνητικό μήνυμα
           134
            15
neg
Name: Emoji text VaderSentiment LABEL, dtype: int64
print(df["emoil textblob bayes sentimentlabel"], value counts())
WOOD'S
             2:0:3
              34
#moutral
#Nowe: testblob polarity bayes, dtype: into4-->XXPII EMOCI KIIMIMO
-->άρα τα emojis αλλάζουν ΕλΩ τον τόνο μετώνοντας την ουδετερότητα...
           97
pos
           30
neig
mentral
Name: amoji textblob bayes sentimentlabel, dtype: int64
```



Visualisations:

matplot.lib, py.plot + softcode libraries

- quick inspection of data patterns
- -grouped data columns statistics
- ⇒ comparing and focusing analysis
- ⇒ spotting mistakes
- ⇒ using basic pandas methods for cleaning annotation noise



Data appropriation

→emoji/emoticon translation (EMOJI & DEMOJI libraries)

→in context emoji//emoticon

translation

```
partnessent (ANAL Therefore any Partness and Company)

# Of you have It commits away for any image of a partness from their three who tage throbe from the committee from the partness for a present property and the company for the company
```

```
textSetext4.str.replace: " . " - ]
tecthictects, kts.reglace(":", " ")
test/brest6.str.replace("face", "7)
testExtest7. ste. inner(). ste. replace("ta", "")
testhetests str Inam'() atr neglacel'hackbass index scirting door', "like there")
textIdetestS.str.lmost().str.replace('face with open month', 'astumished')
textilutestip.str.lisser().str.registe('sseen boolth sorker', 'surse')
textitutesti0.str.Somer().str.reglace('four Inst Lincor', 'good lack')
test12stast11.str.lour().str.replace('four loaf closer', 'good lack')
testilatestilate.lose().ate.replace('souble outlanation mark', '11')
testlibriestli, str. imser[] atr. replace['spec Secto', 'sepplication']
textiStesti4.str.lower().str.reglace('folded bunds', 'preying')
textifictestif-str.loser()_str.replace('thombs up', 'like')
```

Tensions between real +perceived representation affordances

→ decorative > clarifying word clouds

→ no adequate emoji representation in plots

 \rightarrow BERTopic

modeling technique > n-grams

	Topic # 01	Topic # 02	Topic # 03
0	amen	birthday	years
1	god	today	twins
2	write	wish	20
3	love	happy	good
4	page	wished	pregnancy
5	share	wishes	children
6	like	old	man
7	please	make	single
8	girl	please	days
9	sav	vears	congratulate

```
photo', 0.22849291214251222),
share', 0.16967127849622485),
hi', 0.14398329748979846),
wish', 0.14139273208018738),
birthday', 0.13754442275301546),
friends', 0.1322772861818923),
disabled', 0.1322772861818923),
think', 0.12400148124660015),
problem', 0.10937868529310978),
'christmas', 0.10937868529310978)
```

Discussion

NLP is a professional digital competence that should be more advanced in Higher Education

Challenges to be addressed:

- error analysis practices→ reflect on non English NLP restrictions + linguistic contribution in digital humanities
- research priorities +NLP goals: corpus criteria, data set formation practices
- demand for scaffolded NLP +NLU didactic strategies

different data mining approaches, cleaning text methods

different python libraries use: informal, formal speech analysis

Discussion

Tensions to overcome:

- instrumental approach of NLP (a tool that can be handled with elementary skills)
- over relying on emerging digital literacy skills or in other coding languages experience
- strong classification boundaries + context frame (Bernstein, 1975: 88-89) impede interdisciplinary pedagogy and didactics

Tensions to to be anticipated for scaffolding practices:

- different understandings on situated awareness may harm task performance and deprive learner agency
- perceived and real NLP affordances (positive or negative) can be either way productive

Discussion

Needs to attend

- danger of imposing a computational/technocratic point of view on digital humanities studies and a tool didactic approach on AI
- → need for bridging the divide: different communication approach +educational initiative in digital humanities
 - digital humanities curriculum should be based on digital humanities needs analysis in research and teaching
- → need to form a digital humanities Grammar: coding, syntax, pragmatics (ontologies, networks and mixed approaches)
- → need for an informed educational approach to connectivism and its relation to knowledge, learning and truth
- → need for curriculum alignment and professional development strategies that are research domain specific

THANK YOU!