

Poisoned texts report

Contents

1	Summary	2
1.1	Basic info	2
1.2	Shared or genuine lexical items	2
1.3	File by file coincidence	2
2	Style	3
2.1	Heatmap	3
2.2	Most frequent words	3
2.3	Most frequent words first position in sentence	4
2.3.1	Single words	4
2.3.2	Two first words	4
2.3.3	Three first words	4
2.3.4	Four first words	4
2.4	Most frequent upos first position in sentence	5
2.4.1	Single upos	5
2.4.2	Two first upos	5
2.4.3	Three first upos	5
2.4.4	Four first upos	5
3	Similar sentences	6
4	Annex (texts)	6



Reference: Cabedo, A. (2022) Poisoned texts. V.1.0 beta. Available at: <https://github.com/acabedo/poisonedtexts>

GNU General Public License v3.0

In this report you will find tables and charts covering the selected texts. Please, take into account that the interpretation it has not only one direction, and that this is a task for a specialist.

1 Summary

In this section, there are two brief summaries. The first one shows the number of paragraphs, sentences and ngrams for each text file. The second one presents the quantity of ngrams shared more than once, just once between at least two files or even ngrams appearing only at one of the texts being analysed.

The selected ngrams for the analysis were 3.

1.1 Basic info

filename	paragraphs	sentences	words
spinosaurus_cat.txt	2	4	102
stegosarurus_cat.txt	3	18	607
tyrannosaurus_cat.txt	3	13	377

1.2 Shared or genuine lexical items

ngram	type	freq	total
spinosaurus_cat.txt	once	0	0.00
spinosaurus_cat.txt	only	75	100.00
spinosaurus_cat.txt	shared	0	0.00
stegosarurus_cat.txt	once	2	0.40
stegosarurus_cat.txt	only	495	99.20
stegosarurus_cat.txt	shared	2	0.40
tyrannosaurus_cat.txt	once	2	0.66
tyrannosaurus_cat.txt	only	301	99.01
tyrannosaurus_cat.txt	shared	1	0.33

1.3 File by file coincidence

filename	variable	value	total	percent
tyrannosaurus_cat.txt	stegosarurus_cat.txt	3	304	0.99
stegosarurus_cat.txt	tyrannosaurus_cat.txt	3	499	0.60
spinosaurus_cat.txt	spinosaurus_cat.txt	0	75	0.00
stegosarurus_cat.txt	spinosaurus_cat.txt	0	499	0.00
tyrannosaurus_cat.txt	spinosaurus_cat.txt	0	304	0.00
spinosaurus_cat.txt	stegosarurus_cat.txt	0	75	0.00
stegosarurus_cat.txt	stegosarurus_cat.txt	0	499	0.00
spinosaurus_cat.txt	tyrannosaurus_cat.txt	0	75	0.00

filename	variable	value	total	percent
tyrannosaurus_cat.txt	tyrannosaurus_cat.txt	0	304	0.00

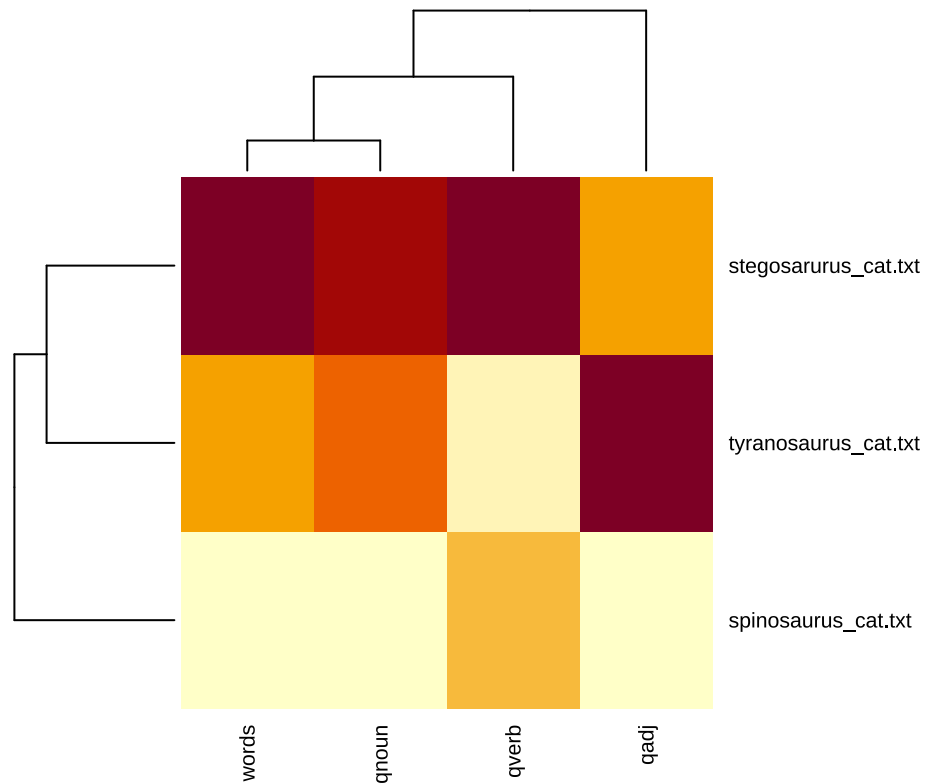
2 Style

This section covers several variables covering common styling characteristics for individuals. These variables must be selected previously on the app. Accordingly, they will be projected on a heatmap.

The frequencies have been automatically scaled with a reference to 100 words, because every sentence has a different quantity of words. Furthermore, the variable *words* refers to the mean of words by sentence for every text.

doc_id	words	qnoun	qpropn	qconj	qadv	qprep	qadj	qverb	qdet	qcomma	qcolon
spinosaurus_cat.txt	25.5	16.7	12.8	2.5	4.6	14.9	6.5	6.0	5.4	1.9	0.0
stegosaurus_cat.txt	33.7	21.0	2.2	2.8	6.4	13.6	8.6	7.7	15.3	4.0	0.0
tyrannosaurus_cat.txt	29.0	19.7	3.1	3.1	5.1	14.1	11.5	5.4	14.3	4.0	0.2

2.1 Heatmap



2.2 Most frequent words

doc_id	ngram	freq
spinosaurus_cat.txt	el	4
spinosaurus_cat.txt	i	3
stegosarurus_cat.txt	que	4
stegosarurus_cat.txt	de	4
tyrannosaurus_cat.txt	i	5
tyrannosaurus_cat.txt	el	4

2.3 Most frequent words first position in sentence

2.3.1 Single words

doc_id	token	freq
spinosaurus_cat.txt	Espinosaure	1
spinosaurus_cat.txt	Fou	1
stegosarurus_cat.txt	Avui	2
stegosarurus_cat.txt	El	2
tyrannosaurus_cat.txt	Els	2
tyrannosaurus_cat.txt	L'	2

2.3.2 Two first words

doc_id	ngram	freq
spinosaurus_cat.txt	Espinosaure (1
spinosaurus_cat.txt	Fou descobert	1
stegosarurus_cat.txt	Avui en	2
stegosarurus_cat.txt	Tot i	2
tyrannosaurus_cat.txt	S' han	2
tyrannosaurus_cat.txt	Com els	1

2.3.3 Three first words

doc_id	ngram	freq
spinosaurus_cat.txt	Espinosaure (Spinosaurus	1
spinosaurus_cat.txt	Fou descobert per	1
stegosarurus_cat.txt	Avui en dia	2
stegosarurus_cat.txt	Així doncs ,	1
tyrannosaurus_cat.txt	Com els altres	1
tyrannosaurus_cat.txt	Diversos altres gèneres	1

2.3.4 Four first words

doc_id	ngram	freq
spinosaurus_cat.txt	Espinosaure (Spinosaurus)	1

doc_id	ngram	freq
spinosaurus_cat.txt	Fou descobert per primera	1
stegosarurus_cat.txt	Així doncs , no	1
stegosarurus_cat.txt	Avui en dia ,	1
tyrannosaurus_cat.txt	Com els altres tiranosàurids	1
tyrannosaurus_cat.txt	Diversos altres gèneres de	1

2.4 Most frequent upos first position in sentence

2.4.1 Single upos

doc_id	upos	freq
spinosaurus_cat.txt	ADV	1
spinosaurus_cat.txt	AUX	1
stegosarurus_cat.txt	DET	8
stegosarurus_cat.txt	ADV	5
tyrannosaurus_cat.txt	DET	6
tyrannosaurus_cat.txt	PRON	3

2.4.2 Two first upos

doc_id	ngram	freq
spinosaurus_cat.txt	ADV PRON	1
spinosaurus_cat.txt	AUX ADJ	1
stegosarurus_cat.txt	DET NOUN	4
stegosarurus_cat.txt	ADV ADP	3
tyrannosaurus_cat.txt	DET NOUN	4
tyrannosaurus_cat.txt	DET DET	2

2.4.3 Three first upos

doc_id	ngram	freq
spinosaurus_cat.txt	ADV PRON PRON	1
spinosaurus_cat.txt	AUX ADJ ADP	1
stegosarurus_cat.txt	ADV ADP NOUN	3
stegosarurus_cat.txt	DET DET NOUN	3
tyrannosaurus_cat.txt	DET DET NOUN	2
tyrannosaurus_cat.txt	PRON AUX VERB	2

2.4.4 Four first upos

doc_id	ngram	freq
spinosaurus_cat.txt	ADV PRON PRON AUX	1
spinosaurus_cat.txt	AUX ADJ ADP ADJ	1

doc_id	ngram	freq
stegosarurus_cat.txt	ADJ ADP NOUN ADP	1
stegosarurus_cat.txt	ADV ADP NOUN ADJ	1
tyrannosaurus_cat.txt	ADP NOUN ADP DET	1
tyrannosaurus_cat.txt	DET DET NOUN ADP	1

3 Similar sentences

This last section presents the common verbatim concordances for the coinciding *ngrams* among at least two texts.

```
## [1] "Please, select two filenames if you want this section to be available"
```

4 Annex (texts)

In this section all the texts appear

sentence_id

spinosaurus_cat.txt_1: Espinosau (Spinosaurus) és un gènere de dinosaure teròpode megalosauroïdeu que visqué al nord Àfrica a mitjans del Cretaci (en l'època Cenomaniana, fa entre 95 i 93 milions d'anys).

spinosaurus_cat.txt_2: Fou descobert per primera vegada l'any 1912 a Egipte pel paleontòleg alemany Ernst Stromer.

spinosaurus_cat.txt_3: També se n'han trobat restes al Níger.

spinosaurus_cat.txt_4: L'espinosau podia arribar a mesurar 17 m de llargària i 5 m d'alçada, el que el fa ser el teròpode més llarg conegut (fins i tot més llarg que el Tyrannosaurus, Giganotosaurus i Carcharodontosaurus).

stegosarurus_cat.txt_1: L'estegosaure (Stegosaurus, rèptil amb teulada) fou un dinosaure herbívor el qual va viure a Nord-amèrica i a Europa[2] durant el Juràssic superior, entre 156 i 155 milions d'anys enrere, del grup dels ornitòmisis.

stegosarurus_cat.txt_2: Els primers fòssils van ésser identificats per Othniel Charles Marsh el 1877, essent una de tantes espècies descobertes durant les anomenades Guerres dels Ossos a l'oest nord-americà.

stegosarurus_cat.txt_3: Avui en dia s'han trobat desenes d'esquelets complets i diversos fòssils més als rics jaciments de Wyoming, Utah, Colorado i Portugal,[2] i es pot dir amb tota seguretat que és un dels dinosaures més coneguts.

stegosarurus_cat.txt_4: Els estegosaures tenien una longitud de fins a 9 metres i un pes de dues tones.

stegosarurus_cat.txt_5: El cos, sostingut de forma obligada per quatre potes finalitzades en peülles, es feia més estret cap al davant, fins a finalitzar en un minúscul cap allargat dotat d'un feble bec i dents petites d'escassa capacitat masticatòria.

stegosarurus_cat.txt_6: Ja al segle XIX es va convertir en el paradigma del dinosaure babau, a causa del fet que el seu cervell era poc més gran que una nou.

stegosarurus_cat.txt_7: Tot i que avui se sap que el cervell no creix de forma proporcional a la resta del cos i que a la majoria de dinosaures tenia una mida normal per aquests, els estegosaures el continuen tenint lleugerament més petit del que es podria esperar.

stegosarurus_cat.txt_8: Així doncs, no tindrien els sentits molt desenvolupats.

stegosarurus_cat.txt_9: Comparativa de mida entre un estegosaure i un home Els trets més notables d'aquests animals són les seves quatre punxes de fins a 60 centímetres a la cua i sobretot la filera d'amples plaques que coronaven la seva esquena.

stegosarurus_cat.txt_10: La seva posició ha sigut força discutida anteriorment, arribant a suggerir-se que es distribuïen paral·leles sobre el tronc o àdhuc enganxades al cos a manera d'armadura.

stegosaurus_cat.txt_11: Tot i això, el coneixement actual permet assegurar que es distribuïen en dues fileres sobre el tronc, de forma aleatòria i en posició vertical.

stegosaurus_cat.txt_12: El seu paper defensori era menyspreable a causa del seu escàs gruix i forta vascularització, de manera que probablement tenien un paper més dissuasiu que altra cosa: sobre el tronc de l'animal, aquest semblaria encara més gran als seus predadors com ara l'al·losàurid i el ceratosaure.

stegosaurus_cat.txt_13: S'ha suggerit darrerament que la vascularització de les plaques podria tenir-les de colors vius a voluntat de l'animal, bombant la sang amb força vers elles, i podent així tenir un paper de reclam sexual o augmentant el seu poder dissuasiu.

stegosaurus_cat.txt_14: Una altra teoria les considera una eina termoreguladora.

stegosaurus_cat.txt_15: Avui en dia, els dinosaures són considerats animals de sang calenta, de manera que les plaques no servien per escalfar l'animal com es va arribar a creure, sinó per refredar-lo.

stegosaurus_cat.txt_16: Com que no mastegava el seu aliment, l'estegosaure s'empassava enormes quantitats de plantes que fermentaven dins dels seus grans intestins: això originava enormes quantitats d'escalfor que eren emeses a l'exterior mitjançant les plaques, salvant l'animal de rostir-se amb el seu propi metabolisme.

stegosaurus_cat.txt_17: Un experiment amb una maqueta que reproduïa la forma d'un estegosaure corroborà aquesta idea.

stegosaurus_cat.txt_18: També s'observa que les plaques són progressivament més amples segons augmenta la mida entre els membres de la família Stegosauridae, de la qual aquesta espècie és la més gran.

tyrannosaurus_cat.txt_1: Els tiranosaures (Tyrannosaurus) són un gènere extint de grans dinosaures teròpodes que visqueren en allò que avui en dia és l'oest de Nord-amèrica, que aleshores formava part del continent insular de Laramídia, durant el Maastrichtià (Cretaci superior), fa entre 66 i 68 milions d'anys.

tyrannosaurus_cat.txt_2: Tenien una distribució molt extensa en comparació amb la dels altres tiranosàurids.

tyrannosaurus_cat.txt_3: Foren els darrers tiranosàurids a aparèixer i uns dels últims dinosaures no aviaris que caminaren sobre la Terra abans de l'extinció del Cretaci-Paleogen.

tyrannosaurus_cat.txt_4: L'espècie T. rex és un dels teròpodes més coneguts.

tyrannosaurus_cat.txt_5: Com els altres tiranosàurids, el tiranosaure era un carnívor bípede amb un crani massiu, equilibrat per una llarga cua pesant.

tyrannosaurus_cat.txt_6: En comparació amb les grans i potents potes posteriors, les potes anteriors del tiranosaure eren petites, però particularment potents per la seva mida, i presentaven dos dits amb urpes.

tyrannosaurus_cat.txt_7: Tot i que altres teròpodes igualaven o superaven el tiranosaure en mida, fou el tiranosàurid més gran conegut i un dels predadors terrestres més grossos, fent fins a 13 metres de llarg,[2] fins a 4 metres d'alt al maluc,[3] i fins a 6-8 tones de pes.[4] Sent de llarg el carnívor més gran del seu medi, el tiranosaure podria haver estat un depredador alfa, caçant hadrosaures i ceratops, tot i que alguns experts pensen que podria haver estat principalment un carronyaire.

tyrannosaurus_cat.txt_8: S'han identificat més de 30 exemplars de tiranosaure, alguns dels quals són esquelets gairebé complets.

tyrannosaurus_cat.txt_9: S'han trobat teixits tous i proteïnes en almenys un d'aquests exemplars.

tyrannosaurus_cat.txt_10: L'abundància de material fòssil ha permès una recerca significativa de molts aspectes de la seva biologia, incloent-hi el cicle vital i la biomecànica.

tyrannosaurus_cat.txt_11: Els costums alimentaris, la fisiologia i la velocitat potencial del tiranosaure són alguns temes de debat.

tyrannosaurus_cat.txt_12: La seva taxonomia també és controvertida; alguns científics creuen que Tarbosaurus bataar, d'Àsia, és una segona espècie de tiranosaure, mentre que d'altres mantenen el tarbosàurid com un gènere separat.

tyrannosaurus_cat.txt_13: Diversos altres gèneres de tiranosàurids nord-americans també han estat sinonimitzats amb Tyrannosaurus.
