WHY YOU SHOULD ~~NEVER~~ TALK ABOUT SEX, RELIGION, POLITICS

A TOPIC MODELING STORY

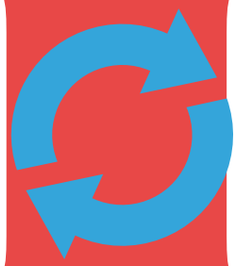
TED

IDEAS WORTH SPREADING

WHAT DO THE EXPERTS HAVE TO SAY?

DATA COLLECTION

TEXT PROCESSING



TOPIC MODELING

350 transcripts

Linguistic annotation

LDA
NMF

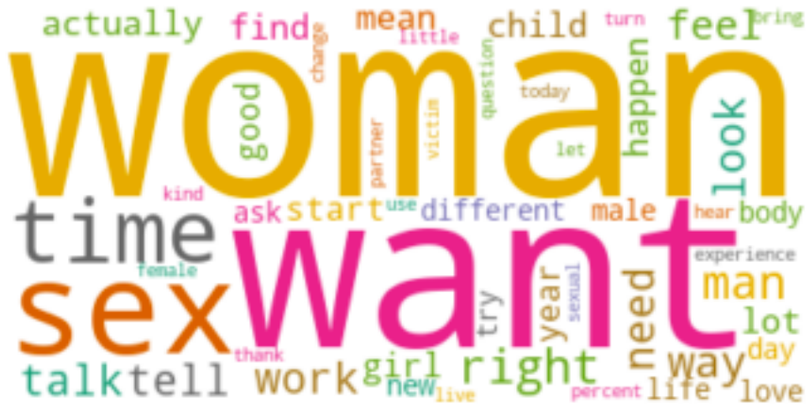
Requests

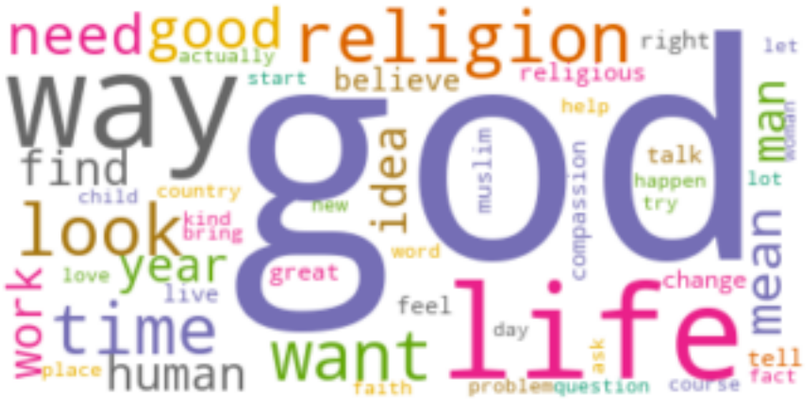
Beautiful Soup

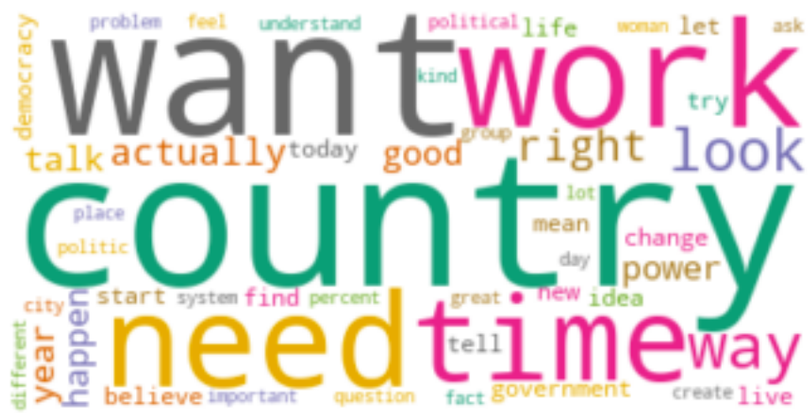
spaCy

scikit learn

gensim







Sex

Religion

Politics

LDA

infidelity
sex stories
orgasm

desire
family
reproduction
rape
porn

NMF

sex workers
violence
lifestyle

Usage → Tagger

Future → Recommender
Summarizer

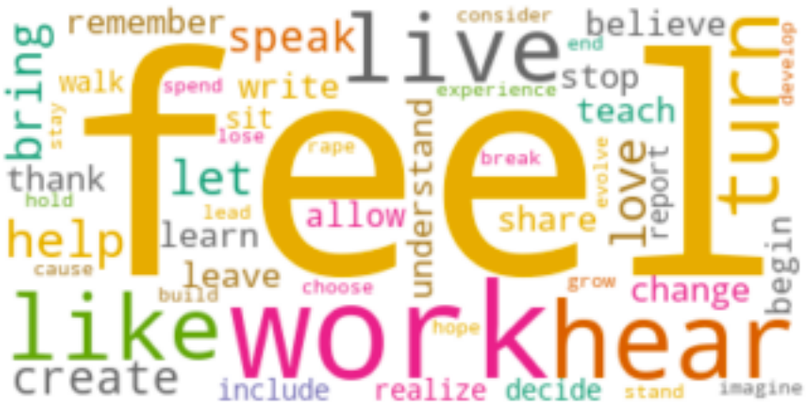
WHAT DO THE EXPERTS HAVE TO SAY?

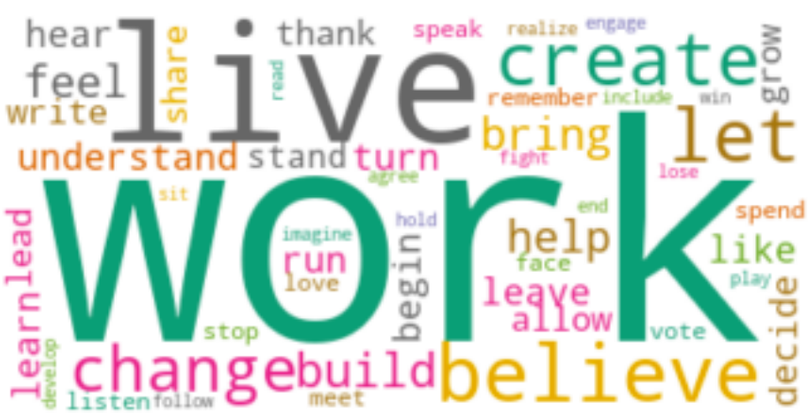
Ten things you
didn't know
about orgasm.
TED
IDEAS WORTH SPREADING

THANKS

P.S. LET'S ~~NOT~~ TALK ABOUT SEX, RELIGION, POLITICS

APPENDIX

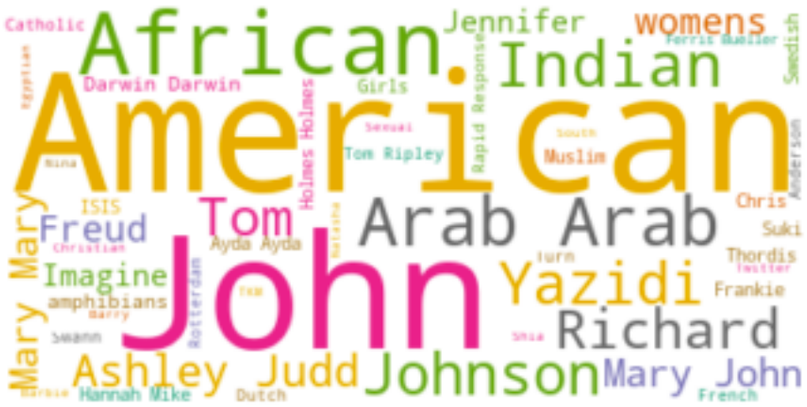


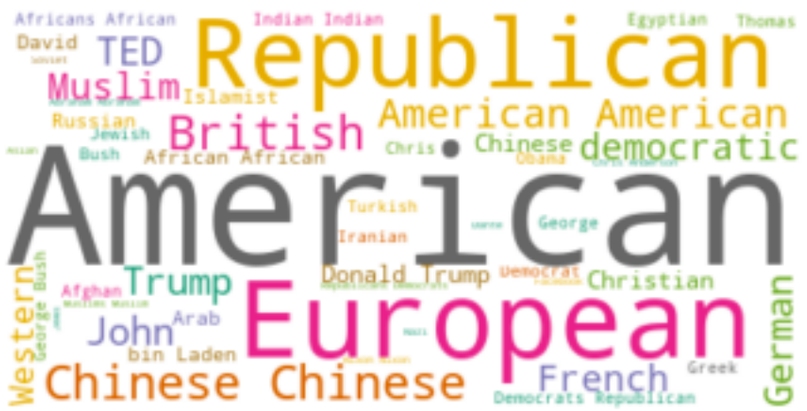


Sex

Religion

Politics





Sex

Religion

Politics

NON-NEGATIVE MATRIX FACTORIZATION

	topic_0	topic_1	topic_2	topic_3	topic_4	topic_5	topic_6	topic_7	topic_8	topic_9
0	woman	sex	child	male	kit	man	desire	gay	pornography	pollen
1	girl	worker	family	female	rape	woman	sexual	agenda	violence	plant
2	sexual	work	year	body	year	issue	partner	lifestyle	sex	insect
3	sex	law	life	penis	time	violence	affair	love	sexual	flower
4	young	money	world	animal	victim	lot	question	time	kid	different
5	man	street	percent	clitoris	state	sexual	good	right	year	little
6	body	industry	time	sperm	police	victim	sex	life	adolescent	lot
7	right	criminalization	today	sex	crime	gender	couple	state	conversation	specie
8	country	client	birth	human	old	culture	time	evil	study	picture
9	time	prostitution	woman	time	work	boy	relationship	lot	problem	nectar

LATENT DIRICHLET ALLOCATION

```
[(0,
  '0.019*"affair" + 0.011*"door" + 0.008*"sexual" + 0.008*"true" + 0.007*"relationship" + 0.007*"infidelity" + 0.007*"partner" + 0.006*"key" + 0.006*"good" + 0.006*"sex"'),
 (1,
  '0.029*"woman" + 0.015*"sex" + 0.015*"child" + 0.013*"kid" + 0.012*"girl" + 0.011*"family" + 0.011*"life" + 0.011*"year" + 0.009*"young" + 0.009*"parent"'),
 (2,
  '0.019*"sex" + 0.019*"pornography" + 0.015*"sexual" + 0.012*"man" + 0.011*"woman" + 0.011*"body" + 0.010*"relationship" + 0.009*"human" + 0.008*"lot" + 0.008*"year"'),
 (3,
  '0.021*"woman" + 0.012*"sex" + 0.008*"story" + 0.008*"hyman" + 0.007*"girl" + 0.007*"myth" + 0.007*"hymen" + 0.007*"room" + 0.007*"locker" + 0.007*"question"'),
(4,
 '0.012*"clitoris" + 0.011*"child" + 0.011*"desire" + 0.011*"body" + 0.010*"time" + 0.009*"world" + 0.009*"clit" + 0.008*"abuse" + 0.008*"erotic" + 0.008*"sex"'),
(5,
 '0.027*"sex" + 0.021*"kit" + 0.017*"rape" + 0.014*"work" + 0.013*"worker" + 0.011*"time" + 0.010*"skeleton" + 0.009*"year" + 0.007*"tissue" + 0.007*"law"'),
(6,
 '0.012*"baseball" + 0.011*"pizza" + 0.009*"atrazine" + 0.009*"different" + 0.008*"chemical" + 0.008*"sexual" + 0.008*"project" + 0.008*"cancer" + 0.007*"activity" + 0.007*"breast"'),
(7,
 '0.026*"woman" + 0.019*"man" + 0.012*"sex" + 0.012*"sexual" + 0.010*"girl" + 0.009*"time" + 0.008*"year" + 0.007*"violence" + 0.006*"child" + 0.006*"world"'),
(8,
 '0.011*"time" + 0.008*"woman" + 0.007*"man" + 0.007*"orgasm" + 0.007*"sex" + 0.007*"sexual" + 0.006*"disgust" + 0.005*"privacy" + 0.005*"science" + 0.005*"lot"'),
(9,
 '0.020*"male" + 0.020*"sex" + 0.015*"female" + 0.012*"gay" + 0.010*"pollen" + 0.010*"insect" + 0.008*"animal" + 0.008*"time" + 0.007*"plant" + 0.007*"different"')]
```