

Multimedia Appendix 4: Examples of Regular Expression Codes for Reducing Noise Patterns in the Patients Posts for Drug Reviews for Case Study “Identifying Factors Affecting Patients’ Attitudes towards Antidepressants”

Source: Zolnoori, M. (2017). “Utilizing Consumer Health Posts for Pharmacovigilance: Identifying Underlying Factors Associated with Patients’ Attitudes Towards Antidepressants.” Theses and Dissertations. 1733. <https://dc.uwm.edu/etd/1733>

Noisy patterns	Example	Regular expression code
Inserting space after full stop punctuations and the beginning of a new sentence	“I have suffered many side effects from the prednisone used to treat the <u>symptoms</u> . I am just now starting to feel better.”	comments = re.sub (r'([a-zA-Z])([.?!])([a-zA-Z])',r'\1\2\3', comments)
Removing Emails	com.uk@hotmail.comsevere	
Replacing dot before parenthesis with space	“But about a year later I tried others and could tolerate them much <u>better</u> . (Such as Lexapro and Prozac.) I hate it.”	Comment = re.sub(r'\.(?!(?:[^\]]*)\s*\.)', r'\1', Comment)
Removing repeated punctuations from the end of the sentence	“Think menopause is a huge factor in <u>this...i have</u> only been on generic zoloft for about a week and a half”	comment = re.sub (r'([.,:;!/?-])\1+', r'\1',comment)
Replacing “.” with “,” from the list of side-effects (ADRs)	“Night <u>terrors</u> . <u>Vivid</u> dreams. <u>Nightmares</u> . <u>auditory</u> hallucinations everyday!!!! never take this drug!!!”	comment = re.sub(r'\.(?!(?:i she he He She I we We They they it It You you)\b)\w+(?:\s+\w+)?\s*(?=[^\w\s])\$', r'\1', comment , flags=re.I)
Replacing hyphen with the dot when hyphen indicates the full stop of the sentence	“Severe teeth clenching/TMJ which caused excruciating <u>headaches - clenching</u> came on right after taking the dose and would gradually wear off”	Comment = re.sub (r'(\s+)(-)(\s+)', r'\1.\3 ', comment)
Replacing the comma with the dot, when comma indicates full stop of a sentence	“Effexor People are still having problems with this medication even after being off of it for 8 months or more, please don't take this medication.”	Comment = re.sub (r'(\s+)(,)(\s+)', r'\1.\3 ', comment)
Replacing dot after numbers when numbers indicate the order of items in a list	“After two to three hours vomit 2. Stiff muscles (Lack of oxygen?) 3. <u>Spacey</u> , out of sync with time (as in Where did that tree come from? What am I doing?) 4. Weird, (as in How did that tree get there?). 5. Weak muscles”	comment = re.sub(r'(\d)\.(?!(?:[a-zA-Z]))',r'\1)\2',comment)
Inserting space between parenthesis and the following words	"Was deadly sick for 2 wks and he kept me on it regardless, upped the dosage three times over a year because it wasn't <u>working</u> (<u>duh</u> , it was my thyroid!)That dr. is history, btw! This time, I was very leary of taking anything."	comment = re.sub(r'(\s+)(\w+)',r'\1\2',comment)
Replacing (, - ,) before “overall” with dot	and I was only on 2,5 months - <u>overall - DOES</u> do what it is supposed to <u>do - you</u> need to	comment= re.sub(r'[- ,](\s*[oO]verall?\s+)',r'. \1',comment)

	weight whether it is worth the side effects.	
Patients did not use any punctuations indicating sentence stop	I was hoping it would eventually go away so I continued to take it and after the 6th day of the same intense side effects I finally had to just chunk them in the trash.	No solution
Incomplete sentences	i recommend the <u>cookies.. PS.. i</u> have not gained any <u>weight.. and</u> if	No solution