

Changes made in files “clinical.csv” and in “CSV Annotations” as of 29-11-2018

The changes need to be done after running the clinical_notes_crawling and the annotation_crawling code and before the annot_mapping code. All the changes apply when you are using the Normal String replace code but only very few changes would be required when using Regex.

Removed the following characters from the dataset and replaced them with corresponding characters:

- "/" -> " "
- "-" -> " "
- "," -> " "
- "(" -> " "
- ")" -> " "
- ".." -> "."
- ":" -> " "
- "[" -> " "
- "]" -> " "
- "*" -> " "
- "\" , " "
- "&" -> " "
- "{" -> " "
- "}" -> " "
- "@" -> " "
- "A" -> " "
- "~" -> " "
- "..." -> "."
- 'x4' -> ' 4'
- 'x2' -> ' 2'
- 'x6' -> ' 6'
- 'x8' -> ' 8'
- 'x10' -> ' 10'
- 'x11' -> ' 11'

1. 10_255
 - a. 1.Mantle in clinical.csv changed to 1 Mantle
2. 12_550
 - a. 5.Obesity to 5 Obesity in clinical.csv
3. 13_28
 - a. Change 5-Mar to 3/5

4. **13_596**
 - a. Change “\nlow” to “low” in **CSV Annotations**, **offset** from **1427** to **1428** and **length** from **21** to **20**
5. **13_86**
 - a. Change **PainMed** in **CSV Annotations** to **OffPainMed**, **offset** from **3495** to **3492** and **length** from **7** to **10**
6. **14_223**
 - a. Change “cold” to **cold** in **CSV Annotations** and change **offset** to **1007** and **location** stays same as **quotes** is replaced with **spaces**
7. **14_298**
 - a. Change offset from 5921 to 5922 and length from 10 to 9 for “End-stage”.
8. **14_350**
 - a. Change “lightheadedness” to lightheadedness in clinical.csv
9. **14_490**
 - a. Change “breakthrough.” to breakthrough. In clinical.csv
10. **14_629**
 - a. Change “rise in his serum calcium” to “arise in his serum calcium” in CSV Annotations and offset from 1157 to 1156, length from 25 to 26
11. **14_730**
 - a. Change “exhausted.” to exhausted. In clinical.csv
12. **14_805**
 - a. “Breakthrough” to breakthrough in clinical.csv
13. **16_35**
 - a. Change in CSV_Annotations to 16_35, 3518, Duration, 6, 752, 5 days
14. **17_1003**
 - a. Change “tender” to “nontender”, offset from 2762 to 2759 and length from 6 to 9 in CSV Annotations.
15. **17_1036**
 - a. Change “tender” to “nontender” , offset from 3624 to 3621 and length from 6 to 9 in CSV Annotations.
16. **17_547**
 - a. Change “icteric” to “anicteric”, offset from 1646 to 1644 and length from 7 to 9 in CSV Annotations.
17. **17_605**
 - a. Change “palpable” to “nonpalpable”, offset from 2474 to 2471 and length from 8 to 11 in CSV Annotations
18. **17_611**
 - a. Change “x1” to “1” and offset from 1611 to 1612 and length from 2 to 1. 1635 offset to 1636 and 2 length to 1 in CSV Annotations.
19. **17_777**
 - a. Change “palpable” to “nonpalpable”, offset from 2758 to 2755 and length from 8 to 11 in CSV Annotations.

20. 17_811

- a. Change “tender” to “nontender”, offset from 2980 to 2977, length from 6 to 9 and offset from 3231 to 3228 and length from 6 to 9 in CSV Annotations.

21. 17_852

- a. Change “tender” to “nontender”, offset from 4937 to 4934, length from 6 to 9 in CSV Annotations

22. 17_98

- a. Change “tender” to “nontender”, offset from 2647 to 2644, length from 6 to 9 and offset from 2895 to 2892 and length from 6 to 9 in CSV Annotations

23. 1_1034

- a. Delete one record from CSV Annotations for ‘MS’
(1_1034,109546,Drug,9,5420,MS Contin)

24. 1_1035

- a. Change 10-Jun to 6/10 in two records in CSV Annotations
- b. Change Hypothyroidism17 to Hypothyroidism

25. 1_148

- a. Change ‘CAP’ to ‘UDCAP’, offset from 7169 to 7167 and length from 3 to 5 in CSV Annotations

26. 1_282

- a. Change offset from 535 to 536 and length from 22 to 21 for ‘IgG kappa paraprotein’ in CSV Annotations

27. 1_421

- a. Change “800 mg” to “800mg” and length from 6 to 5 in CSV annotations in clinical.csv and add a space between mg and 1 in clinical.csv

28. 1_583

- a. Change “10-Jun” to 6/10 in CSV Annotations (2times)

29. 1_598

- a. Change 6-Mar to 3/6 in CSV Annotations
- b. Change offset from 2113 to 2114 and length from 10 to 9 for ‘2 patches’ in CSV Annotation

30. 1_61

- a. Change 10-May to 5/10 in CSV Annotations
- b. Change “Velcade ...” to “Velcade” in CSV Annotations

31. 1_620

- a. Delete 1_620,64741,Drug,10,3309,MS Contin from CSV Annotations (duplicate, wrong)
- b. Change 10-May to 5-10 in CSV Annotations
- c. Change offset from 5014 to 5015 and length from 17 to 16 for ‘multiple myeloma’ in CSV Annotations
- d. Change “valacyclovir ... aspirin” to “aspirin” in CSV Annotations

32. 1_672

- a. Change offset from 301 to 302 and length from 16 to 15 for IgG paraprotein in CSV Annotations.

33. 1_694

- a. Change 10-Aug to 8/10 in CSV Annotations
- b. Change offset from 4010 to 4011 and length from 17 to 16 for 'multiple myeloma' in CSV Annotations

34. 1_764

- a. Change '2 weeks out of the month' to 'E2 weeks out of the month', offset from 1934 to 1933 and length from 24 to 25 in csv annotations

35. 1_838

- a. Change offset from 2514 to 2515 and length from 7 to 6 for '1 year' in CSV Annotations

36. 1_859

- a. Remove double quotes from "musculoskeletal pain about 6/10, especially in the back" in CSV annotations (not completed)

37. 1_874

- a. Change in CSV_Annotations to 1_874, 92218, Frequency, 18, 1697, daily as directed ,

38. 3_405

- a. Change '(> 7 cm' to '> 7 cm' and offset from 2901 to 2902 and length from 7 to 6 in CSV Annotations

39. 4_345

- a. Change offset from 2514 to 2515 and length from 20 to 19 for 'spleen not palpable' in CSV Annotations

40. 4_525

- a. Change 'pleen not palpable' to 'spleen not palpable', offset from 2520 to 2519, length from 18 to 19 in CSV Annotations

41. 4_78

- a. Change 'e sclerae are anicteric' to 'The sclerae are anicteric', offset from 2197 to 2195, length from 23 to 25 in CSV Annotations
- b. Change offset from 2259 to 2261, length from 24 to 22 for 'Thyroid is nonpalpable' in CSV Annotations

42. 5_283

- a. Change 'enal insufficiency' to 'renal insufficiency', offset from 1255 to 1254, length from 18 to 19 in CSV Annotations

43. 5_358

- a. Change offset from 3188 to 3189 and length from 42 to 41 for 'anterior or posterior cervical adenopathy' in CSV Annotations

44. 5_48

- a. Change offset from 4469 to 4470 and length from 39 to 38 for 'supraclavicular or axillary adenopathy' in CSV Annotations.

45. 5_502

- a. Change ' "lump\'' to 'lump in his left groin', offset from 2047 to 2048 and length from 24 to 23 in CSV Annotations.
- b. Change offset from 397 to 398, length from 35 to 34 for 'Recurrent follicular cell lymphoma' in CSV Annotation

46. 5_788

- a. Replace ' _ ' with a space for the following words in clinical.csv : 'Gait', 'Paresthesias', 'Constipation', 'Vomiting', 'Nausea', 'Abdominal', 'Dysphagia', 'Anxiety', 'Depression', 'Palpitations', 'Edema', 'Chest', 'Wheezing', 'Anemia', 'Back', 'Joint', 'Rashes', 'Hematuria', 'Dysuria', 'Sore', 'Congestion', 'Weight', 'Weight', 'Chills', 'Sweats',
- b. tamiflu

47. 5_867

- a. Change offset from 3049 to 3050, length from 42 to 41 for 'anterior or posterior cervical adenopathy' in CSV Annotations

48. 6_184

- a. Change 'D5W +50 mEq' to 'D5W+ 50 mEq' in clinical.csv

49. 7_160

- a. Change "goal" to "forgoal" and offset from 2509 to 2506 and length from 12 to 15 in CSV Annotations

50. 7_268

- a. Change offset from 2520 to 2521 for "active bleeding" and leave two spaces extra between active bleeding and change length from 18 to 17

51. 7_355

- a. Change offset from 5689 to 5690 and length from 40 to 39 for "LDH and reticulocyte count are elevated" in CSV Annotations

52. 7_356

- a. Change "CVAswhich are" to "CVAs whichare" in clinical.csv
- b. Change offset from 3656 to 3657 and length from 56 to 55 for " Marginal zone lymphoma involving spleen and bone marrow" in CSV Annotations

53. 7_472

- a. Change offset from 4262 to 4263 and length from 8 to 7 for "nodules" in CSV Annotation

54. 7_565

- a. Remove \n from 'hematocrit of 34.6, which has also decreased', offset from 3833 to 3834, offset from something(?) to 45 in CSV Annotations

55. 7_650

- a. Change "5 mcg/kg/min" to ".5 mcg/kg/min", offset from 2299 to 2298, length from 12 to 13 in CSV Annotations

56. 7_670

- a. Change “\nPOLYETHYLENE GLYCOL” to “POLYETHYLENE GLYCOL”. Offset from 7065 to 7066, length from 20 to 19 in CSV Annotations
- b. Change ‘goal’ to ‘forgoal’, offset from 2509 to 2506, length from 12 to 15 in CSV Annotations

57. 7_719

- a. Change offset from 2545 to 2544, length from 25 to 24 for ‘elevated ionized calcium’ in CSV Annotation

58. 7_800

- a. Change “25 mg.” to “25 mg” in CSV Annotations and clinical.csv and change length from 6 to 5

59. 7_801

- a. Change offset from 2050 to 2051, length 59 to 57 in ‘Marginal zone lymphoma involving spleen and \nbone marrow’ in CSV Annotations.

60. 7_92

- a. Change x3 days and x3 weeks to 3 days and 3 weeks respectively in CSV Annotations. Also increase their corresponding offset by 1 and decrease length by 1.

61. 8_451

- a. Remove _ and add space between ‘No’ and the following words in clinical.csv: Rashes, Dysuria, Chest Pain

62. 8_498

- a. Remove _ and add space between ‘No’ and the following words in clinical.csv: 'Hematemesis', 'Vomiting', 'Abdominal', 'Heartburn', 'Dysphagia', 'Depression', 'Claudications', 'Palpitations', 'Orthopnea', 'Edema', 'Chest', 'Blurry', 'Diplopia', 'Cough', 'Anemia', 'Stiffness', 'Back', 'Joint', 'Lumps', 'Erythema', 'Rashes', 'Dysuria', 'Urgency', 'Frequency', ['*nausea', 'cough*']