|  |  |  |
| --- | --- | --- |
| **Experiment 1: Document at various level for short dataset (law: non-law)** | | |
| **Experiment 1 (No. | No.)** | **law-texts-all.txt** | **non-law-all.txt** |
| 1. **Article level (4|4)** | 'tariff': 0.3329384971656326,  'secretari': 0.33147720211273746,  'prison': 0.24915911995674625,  'life': 0.23611708519236543,  'sentenc': 0.22952296370340217 | 'judg': 0.32073578367328565,  'evid': 0.3092444069593226,  'west': 0.28040281700236735,  'trial': 0.2772485745197228,  'case': 0.2281040582305716 |
| 1. **2-3 paragraphs** | 'tariff': 0.2852867613391492,  'prison': 0.24509619427164647,  'life': 0.2450961942716464,  'would': 0.24193396077834065,  'sentenc': 0.20411506883081784 | 'life': 0.42462507465364185,  'prison': 0.3661564361559012,  'sentenc': 0.2915494390879298,  'hindley': 0.25768521514592213,  'posit': 0.23308080059018935 |
| 1. **Paragraph level (450|3038)** | 'tariff': 0.3724236615213221,  'secretari': 0.32565842167448944,  'prison': 0.2623244008241644,  'state': 0.24618953873707555,  'life': 0.23430524638710098 | 'hindley': 0.6376706428026212,  'myra': 0.5862877432528727,  'west': 0.23128010770574473,  'case': 0.18245376198330635,  'trial': 0.1482359599207642 |
|  | | |
| **Experiment 2: Document at various level for small dataset (COVID-science: non-science)** | | |
| **Experiment 2 (No. | No.)** | **covid-science-small.txt** | **Covid-non-science-small.txt** |
| **Article level (10|10)** | 'covid': 0.46802988644393223,  'model': 0.33198252628296177,  'imag': 0.3267157192360456,  'class': 0.20559425407282628,  'dataset': 0.15776417661740808 | 'test': 0.6105961140643039,  'peopl': 0.31072871004100194,  'covid': 0.2511987082782914,  'symptom': 0.23022371656243645,  'fever': 0.17811639545747135 |
| **2-3 paragraphs** | 'support': -0.3289947178827896,  'granada': -0.24814037209037249,  'covid': -0.2276958139915428,  'research': -0.2193298119218597,  'programm': -0.2193298119218597 | 'market': 0.2986573559242643,  'covid': 0.21979867961934133,  'friday': 0.18596179123202244,  'infect': 0.1521249028447036,  'month': 0.1465324530795609 |
| **Paragraph level (7401|229)** | 'covid': 0.7920607903974568,  'model': 0.2954037408226817,  'imag': 0.2744367432214875,  'detect': 0.16366442454183422,  'class': 0.129553688452376 | 'trial': -0.46733382557099273,  'vaccin': -0.3789502030913742,  'stage': -0.3328546259512892,  'expect': -0.23326973177945665,  'underway': -0.2001556553367668 |
|  | | |
| **Experiment 3: Document at various level for large dataset (COVID-science: non-science)** | | |
| **Experiment 3 (No. | No.)** | **covid-science.txt** | **Covid-non-science.txt** |
| **Article level (20|20)** | 'covid': 0.5195000201940313,  'imag': 0.38194199249598826,  'model': 0.24414357176752643,  'dataset': 0.17846130209695074,  'class': 0.15297059818817785 | 'test': 0.5525012615032742,  'peopl': 0.32711032908806575,  'covid': 0.32350304714824774,  'symptom': 0.20147297742697454,  'travel': 0.15909033106510675 |
| **2-3 paragraphs** | 'support': -0.32899471788278867,  'granada': -0.24814037209037235,  'covid': -0.22769581399154418,  'research': -0.21932981192185935,  'programm': -0.21932981192185935 | 'market': 0.2986573559242645,  'covid': 0.21979867961934121,  'friday': 0.18596179123202242,  'infect': 0.15212490284470354,  'month': 0.14653245307956086 |
| **Paragraph level (14379|387)** | 'covid': 0.8353446259284905,  'imag': 0.35517890693720033,  'detect': 0.14215734819653986,  'model': 0.12671854341204644,  'dataset': 0.1093524203458417 | 'trial': -0.46149334680263143,  'vaccin': -0.3811212110990389,  'stage': -0.32791214235644467,  'expect': -0.23283592428120545,  'dose': -0.1998784280912528 |