|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Experiment 1: Document at various level for short dataset (law: non-law)** | | | | |
| **Experiment 1 (No. | No.)** | **law-texts-all.txt** | | **non-law-all.txt** | |
| **Top words** | **weights** | **Top words** | **weights** |
| 1. **Article level (4|4)** | **'tariff'** | 0.3329384971656329, | **'judg'** | 0.3207357836732813 |
| **'secretari'** | 0.3314772021127373 | **'evid'** | 0.3092444069593179 |
| **'prison'** | 0.2491591199567464 | **'west'** | 0.2804028170023648 |
| **'life'** | 0.2361170851923661 | **'trial'** | 0.2772485745197205 |
| **'sentenc'** | 0.2295229637034024 | **'case'** | 0.2281040582305726 |
|  |  |  |  |  |
| 1. **2-3 paragraphs** | **'tariff'** | 0.2852867613391 | **'life'** | 0.424625074653 |
| **'life'** | 0.245096194271 | **'prison'** | 0.366156436155 |
| **'prison'** | 0.2450961942716 | **'sentenc'** | 0.291549439087 |
| **'would'** | 0.2419339607783 | **'hindley'** | 0.257685215145 |
| **'sentenc'** | 0.2041150688308 | **'posit'** | 0.233080800590 |
|  |  |  |  |  |
| 1. **Paragraph level (450|3038)** | **'tariff'** | -0.372423678930 | **'hindley'** | -0.63764826639 |
| **'secretari'** | -0.325658507266 | **'myra'** | -0.58632670557 |
| **'prison'** | -0.262324361300 | **'west'** | -0.23136661906 |
| **'state'** | -0.246189129218 | **'case'** | -0.18241034492 |
| **'life'** | -0.234305014747 | **'trial'** | -0.14821551842 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Experiment 2: Document at various level for small dataset (COVID-science: non-science: Covid-psychology)** | | | | | | |
| **Experiment 2 (No. | No.)** | **covid-science-small.txt** | | **Covid-non-science-small.txt** | | **Covid-psychology-small.txt** | |
| **Top words** | **weights** | **Top words** | **weights** | **Top words** | **weights** |
| **Article level (10|10)** | **'covid'** | 0.46802988644393423, | **'test'** | 0.6105961140643029, | **'covid'** | 0.41992720259455585, |
| **'model'** | 0.3319825262829552, | **'peopl'** | 0.31072871004100155, | **'psycholog'** | 0.28694986504872455, |
| **'imag'** | 0.32671571923604975 | **'covid'** | 0.25119870827829194, | **'pandem'** | 0.24145037935803, |
| **'class'** | 0.20559425407282925, | **'symptom'** | 0.2302237165624363, | **'impact'** | 0.21712822111271726, |
| **'dataset'** | 0.15776417661741118 | **'fever'** | 0.17811639545747093 | **'health'** | 0.19835033050070103 |
|  |  |  |  |  |  |  |
| **2-3 paragraphs** | **'support'** | -0.3289947178827894, | **'market'** | 0.2986573559242644, | **'misinfodem'** | 0.5033743541941375, |
| **'granada'** | -0.24814037209037246, | **'covid'** | 0.21979867961934094, | **'covid'** | 0.36297115033235156, |
| **'covid'** | -0.2276958139915434, | **'friday'** | 0.18596179123202228, | **'articl'** | 0.21459251935197457, |
| **'programm'** | -0.21932981192185963, | **'infect'** | 0.15212490284470373, | **'commun'** | 0.14837863098037715, |
| **'research'** | -0.21932981192185963 | **'copper'** | 0.14653245307956103 | **'health'** | 0.14837863098037676 |
|  |  |  |  |  |  |  |
| **Paragraph level (7401|229)** | **'covid'** | 0.7920442330385795, | **'trial'** | 0.4673333939365011, | **'pandem'** | 0.3492162493102458, |
| **'model'** | 0.2953962153046564, | **'vaccin'** | -0.3789496907239388, | **'polici'** | 0.33206567897289235, |
| **'imag'** | 0.2745133138045724, | **'stage'** | -0.33285475306510043, | **'health'** | 0.29289264374192664, |
| **'detect'** | 0.16365565281326466, | **'expect'** | -0.2332697529258982, | **'covid'** | 0.26562150610270957, |
| **'class'** | 0.12952267447918445 | **'underway'** | -0.20015564175913142 | **'impact'** | 0.2463281943440908 |
|  | | | | |  | |
| **Experiment 3: Document at various level for large dataset (20 papers COVID-science: non-science: Covid-psychology)** | | | | | | |
| **Experiment 3 (No. | No.)** | **covid-science.txt** | | **Covid-non-science.txt** | | **Covid-psychology.txt** | |
| **Top words** | **weights** | **Top words** | **weights** | **Top words** | **weights** |
| **Article level (20|20)** | **'covid'** | 0.5195000201940321, | **'test'** | 0.5525012615032745, | **'covid'** | 0.3733564956224059, |
| **'imag'** | 0.3819419924959906, | **'peopl'** | 0.32711032908806575, | **'psycholog'** | 0.26785991090419736, |
| **'model'** | 0.24414357176752507, | **'covid'** | 0.32350304714824757, | **'social'** | 0.2517797817165504, |
| **'dataset'** | 0.17846130209695354, | **'symptom'** | 0.20147297742697395, | **'epidem'** | 0.19223792506781157, |
| **'class'** | 0.15297059818817663 | **'travel'** | 0.15909033106510667 | **'health'** | 0.18194229083665558 |
|  |  |  |  |  |  |  |
| **2-3 paragraphs** | **'support'** | 0.32899471788278933, | **'market'** | 0.2986573559242641, | **'misinfodem'** | -0.5033743541941365, |
| **'granada'** | 0.24814037209037237, | **'covid'** | 0.2197986796193419, | **'covid'** | -0.3629711503323515, |
| **'covid'** | 0.22769581399154337, | **'friday'** | 0.18596179123202256, | **'articl'** | -0.2145925193519742, |
| **'programm'** | 0.21932981192185955, | **'infect'** | 'infect': 0.15212490284470295, | **'commun'** | -0.14837863098037748, |
| **'research'** | 0.21932981192185955 | **'reuter'** | 0.1465324530795613 | **'mechan'** | -0.14837863098037712 |
|  |  |  |  |  |  |  |
| **Paragraph level (14379|387)** | **'covid'** | 0.8353445537976227, | **'trial'** | -0.46149246082032147, | **'covid'** | 0.44528765261781605, |
| **'imag'** | 0.3551934971949819, | **'vaccin'** | -0.38111970491120056, | **'pandem'** | 0.3598378349771223, |
| **'detect'** | 0.14217453078608513, | **'stage'** | -0.3279128249788164, | **'health'** | 0.33461199438050226, |
| **'model'** | 0.1267002478578127, | **'expect'** | -0.23283573850743006, | **'polici'** | 0.2627823706235591, |
| **'dataset'** | 0.1093369812928102 | **'dose'** | -0.19988003539528476 | **'impact'** | .2152347635291778 |
|  |  | |  | |  | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Experiment 4: Document at various level for large dataset (40 papers COVID-science: Covid-psychology)** | | | | |
|  | | | | |
| **Experiment 4 (No. | No.)** | **covid-science-large.txt** | | **Covid-psychology-large.txt** | |
| **Top words** | **weights** | **Top words** | **weights** |
| **Article level (40|40)** | **'covid'** | 0.5580672321131127, | **'student'** | 0.31520975301891685, |
| **'imag'** | 0.2828570178358147, | **'psycholog'** | 0.3088225210323284, |
| **'model'** | 0.2574998284866968, | **'covid'** | 0.3082634599864977, |
| **'dataset'** | 0.1425624890560568, | **'train'** | 0.22330563399536046, |
| **'infect'** | 0.13935041563870734 | **'pandem'** | 0.21884380058036093 |
|  |  |  |  |  |
| **2-3 paragraphs** | **'support'** | -0.32899471788278967, | **'misinfodem'** | 0.5033743541941371, |
| **'granada'** | -0.24814037209037254, | **'covid'** | 0.3629711503323513, |
| **'covid'** | -0.22769581399154254, | **'articl'** | 0.21459251935197415, |
| **'programm'** | -0.21932981192185977, | **'mechan'** | 0.14837863098037712, |
| **'research'** | -0.21932981192185977 | **'commun'** | 0.14837863098037704 |
|  |  |  |  |  |
| **Paragraph level (274941|183592)** | **'covid'** | 0.8597188592513563, | **'covid'** | 0.5542346968279828, |
| **'imag'** | 0.24459808570501076, | **'psycholog'** | 0.38587594842118456, |
| **'model'** | 0.21196118253925922, | **'pandem'** | 0.36424816089766787, |
| **'detect'** | 0.14846685076222263, | **'health'** | 0.2988945427299109, |
| **'volum'** | 0.09486963239930264 | **'impact'** | 0.15508253422230434 |