# Cosine Similarity between top five words at different levels of document

|  |  |  |
| --- | --- | --- |
| **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)** |  |  |
| 1. **2-3 paragraphs** |  |  |
| 1. **Paragraph level (450|3038)** |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **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** |
| **Article level (10|10)** |  |  |  |
| **2-3 paragraphs** |  |  |  |
| **Paragraph level (7401|229)** |  |  |  |
|  | | |  |
| **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** |
| **Article level (20|20)** |  |  |  |
| **2-3 paragraphs** |  |  |  |
| **Paragraph level (14379|387)** |  |  |  |

|  |  |  |
| --- | --- | --- |
| **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** |
| **Article level (40|40)** |  |  |
| **2-3 paragraphs** |  |  |
| **Paragraph level (274941|183592)** |  |  |