## Supplementary Materials

**Caption Clustering.** The captions extracted from the PDF articles were clustered using TF-IDF representation and then running K-means. We also estimate the number of cluster in the captions using the Elbow method. We found out the correct number of clusters to be 13. In the Table 2 we share the top terms in those clusters. Table 3 shows some samples belonging to there corresponding clusters.

| Cluster ID : Name | Top-5 Terms | # of Cluster Members |
|---|---|---|
| 0 : Infection statistics about the virus | data, virus, results, infection, analysis | 62155 |
| 1 : RNA | rna, lanes, protein, gel, viral, cells | 3355 |
| 2 : Research on Mice | mice, group, day, infected, days | 3824 |
| 3 : Protein Sequencing | sequences, amino, acid, residues, protein | 7591 |
| 4 : Demographic Understanding | characteristics, demographic, patients, clinical, baseline | 1683 |
| 5 : Oligonucleotide sequences | primers, study, pcr, sequences, oligonucleotide | 1004 |
| 6 : Pandemic Outbreak Modelling | model, parameters, values, data, epidemic | 2457 |
| 7 : SARS related research | patients, respiratory, acute, clinical, syndrome | 3615 |
| 8 : Confirmed Cases tracking | cases, number, confirmed, reported, cumulative | 2285 |
| 9 : Laboratory reports | continued, laboratory, previous, parameters, page | 570 |
| 10 : Influeza Related Research | influenza, pandemic, virus, ahn, seasonal | 3482 |
| 11 : Cell Specific Research | cells, infected, expression, control, infection | 11353 |

Table 2: Clusters of the extracted captions and there corresponding top-5 terms.

**Keyphrase extraction and similar paper recommendations** We have deonstrated our keyphrase extraction capabilities in the Table 4. Users can filter there search results based on them using our Faceted Search feature. Table 5 demonstrates our similar paper recommendation based on SciBERT re-ranking on title and abstract similarity.

| Caption | Cluster ID | Figure |
|---|---|---|
| Table 2. Comparative analysis of results tests for detection of Chikungunya virus in suspected cases of Chikungunya infection. | 0:Infection statistics |  |
| Figure 2.4 Expression of SARS-CoV 1-259 nsp13 in E. coli BL21 (DE3): A, the protein with a C-terminal 6His tag (construct 4); B, the protein without a 6His tag (construct 5). | 1:RNA |  |
| Figure 3. Long-term CTL memory responses in infected/treated animals versus control mice. | 2:Research on Mice |  |
| Fig 1. Alignment of predicted amino acid sequences within exon 1 of CXCL16 proteins from seven different mammalian species. | 3:Protein Sequencing |  |
| Table 2: Oligonucleotide sequences of the primers and probes used in the GPV FQ-PCR method | 5:Oligonucleotide sequences |  |
| Figure 3. Inferring epidemic dynamics in China (excluding Hubei province). (A) Prediction using stage II data. (B) Model-fitting and testing with stage II data. | 6:Outbreak Modelling |  |

Table 3: Random samples from the caption clusters. We show only the first sentence in the captions.

| Paper Title | Top-10 Keyphrases |
|---|---|
| **Evolution and variation of 2019-novel coronavirus** | nCoV, nucleotide, coronavirus, nucleotide substitution, outbreak, substitution rate, nucleotide substitution rate, phylogenetic trees, amino acid, amino |
| **Self-assembly of Severe Acute Respiratory Syndrome Coronavirus Membrane Protein** | Membrane, Protein, Coronavirus, VLP, SARSCoV, Syndrome Coronavirus, Coronavirus Membrane, Golgi, Respiratory Syndrome, Selfassembly |
| **Respiratory viral infections in institutions from late stage of the first and second waves of pandemic influenza A (H1N1) 2009, Ontario, Canada** | outbreak, LTCF, HN, influenza, pdm, late stage, viral infections, Ontario, Respiratory viral infections, wave |
| **Biogenesis and Dynamics of the Coronavirus Replicative Structures** | viruses, viral RNA synthesis, vesicle, synthesis, infection, Structures, Replicative, RNA viruses, Replicative Structures, RNA |
| **Infection with human coronavirus NL63 enhances streptococcal adherence to epithelial cells** | adherence, coronavirus NL, human coronavirus, coronavirus, bacterial pathogens, human coronavirus NL, pathogens, Infection, HCoVNL, NL |

Table 4: Top-10 keyphrases of research paper selected by CeKE-Citing model

| Paper Title | Top-5 Similar Papers |
|---|---|
| **Middle Eastern Respiratory Syndrome Coronavirus (MERS-CoV)** | 1. Middle East Respiratory Syndrome coronavirus (MERS CoV): Update 2013<br><br>2. A novel coronavirus capable of lethal human infections: an emerging picture<br>3. Structure, Immunopathogenesis and Vaccines Against SARS Coronavirus<br>4. Hantaviruses in the Americas and Their Role as Emerging Pathogens<br>5. Zika fever and congenital Zika syndrome: An unexpected emerging arboviral disease |
| **Coronavirus Receptors** | 1. Crystal structure of murine sCEACAM1a[1,4]: a coronavirus receptor in the CEA family<br>2. The nucleocapsid protein of the SARS coronavirus is capable of self-association through a C-terminal 209 amino acid interaction domain<br>3. C-terminal domain of the MERS coronavirus M protein contains a trans-Golgi network localization signal<br>4. Structural and functional analysis of the S proteins of two human coronavirus OC43 strains adapted to growth in different cells<br>5. Species-specific evolution of immune receptor tyrosine based activation motif-containing CEACAM1-related immune receptors in the dog |

Table 5: Titles of top-5 similar paper recommendations by SciBERT