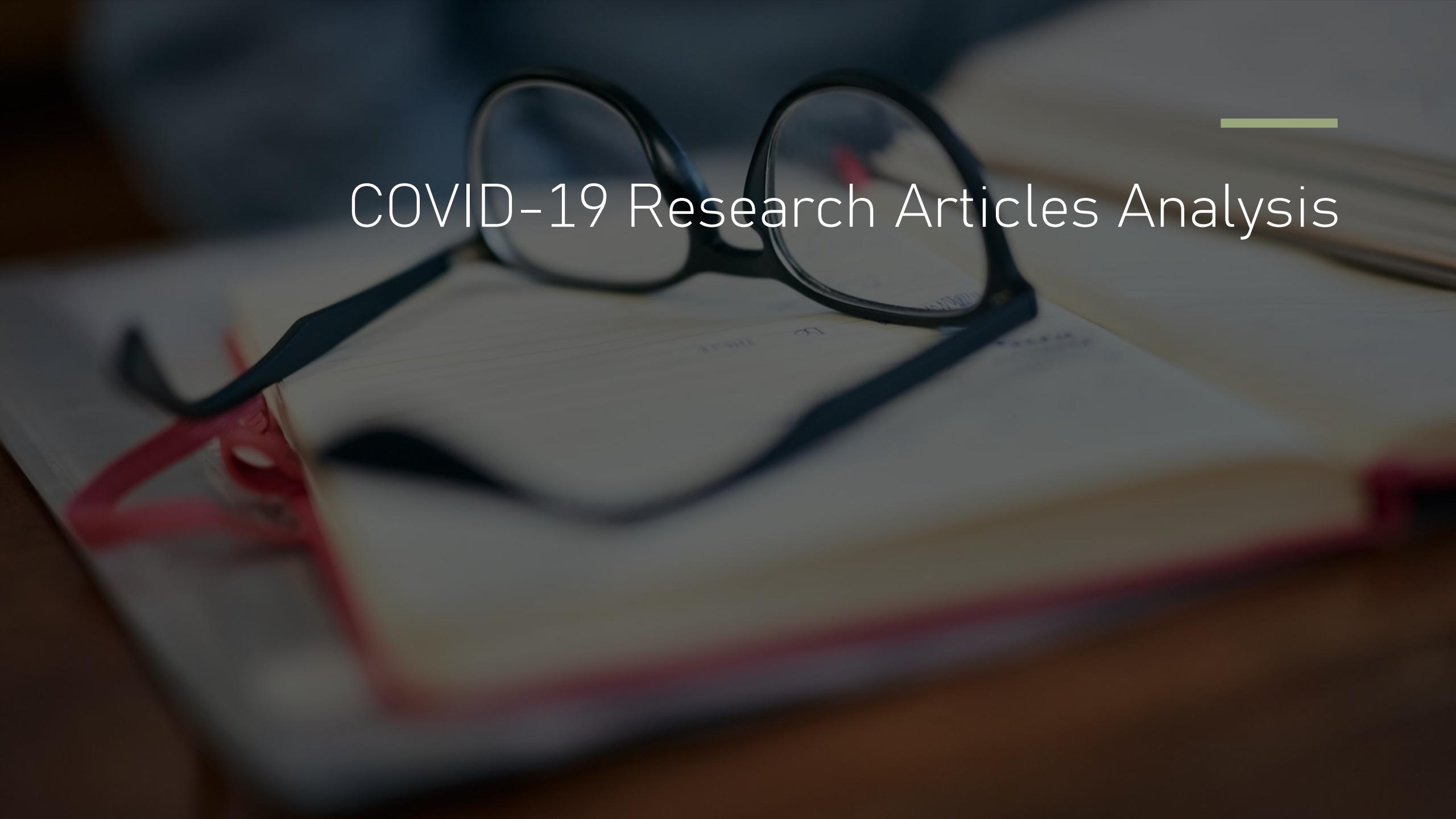


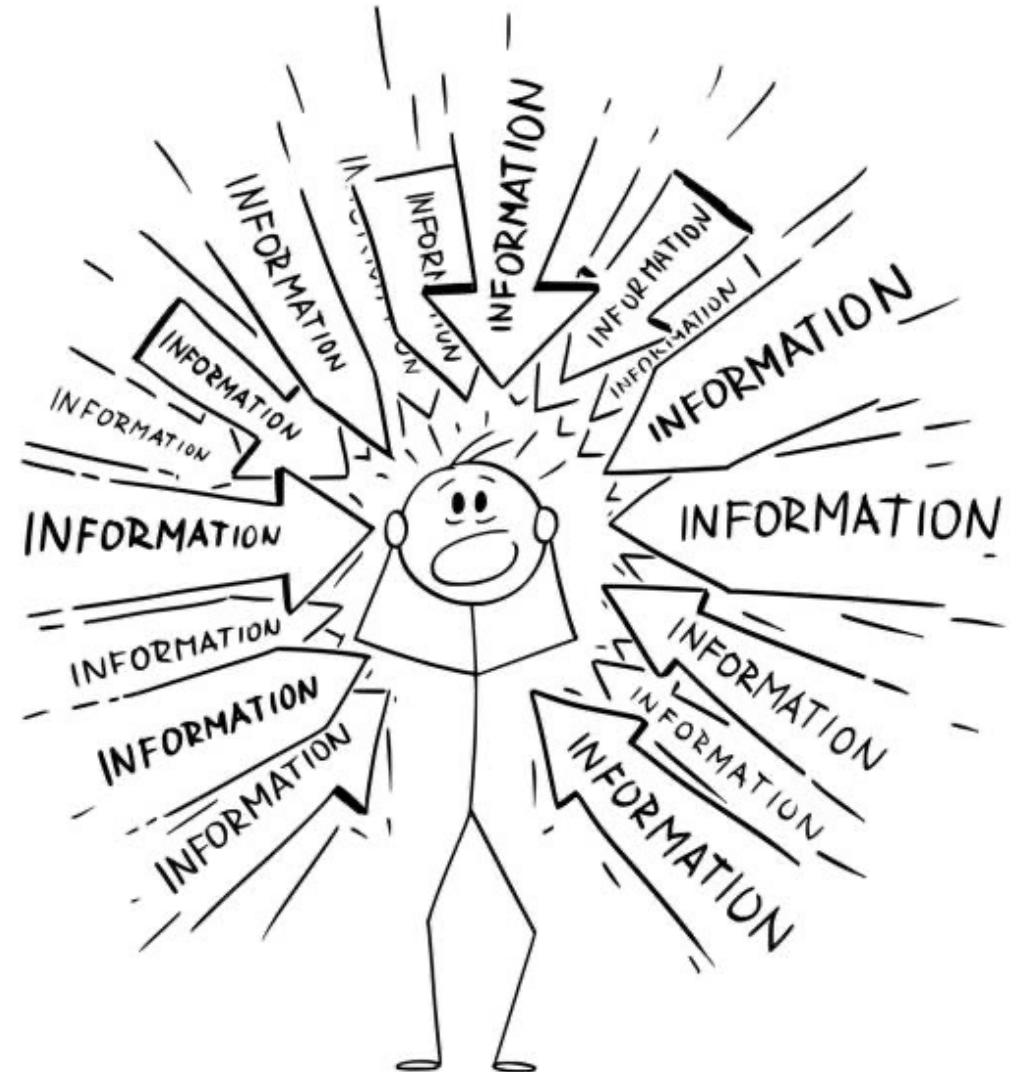
COVID-19 Research Articles Analysis



Number of scientific papers and researches on COVID-19 constantly increasing

It becomes much harder for medical professionals to digest and understand important aspects of the data

Solution is needed to extract useful insights from these vast resources and provide the medical community the required information





DATA

CORD - 19 Open Research Dataset

200 000 + research articles and constantly growing

Subset for analysis - 4 561 articles (2020 – 2021)



METHODOLOGY

- Data acquisition

CORD - 19 (available from kaggle.com)



METHODOLOGY

- Data acquisition

CORD - 19 (available from kaggle.com)

- EDA

Handling missing values, initial analysis, graphical representation

METHODOLOGY

- Data acquisition

CORD - 19 (available from kaggle.com)

- EDA

Handling missing values, initial analysis, graphical representation

- Text Preprocessing

Lemmatization, stop words and punctuation removal, tokenization



METHODOLOGY

- Data acquisition

CORD - 19 (available from kaggle.com)

- EDA

Handling missing values, initial analysis, graphical representation

- Text Preprocessing

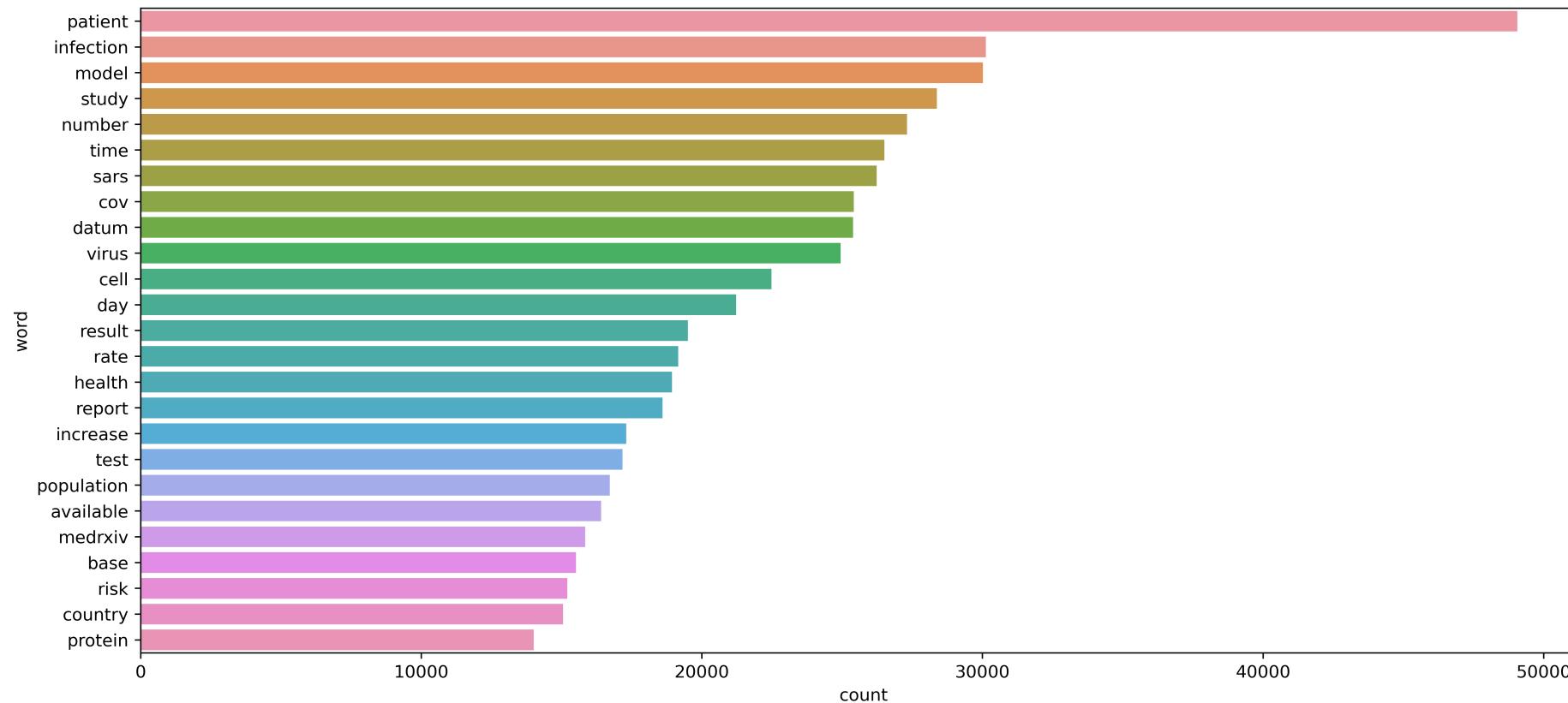
Lemmatization, stop words and punctuation removal, tokenization

- Topic Modeling

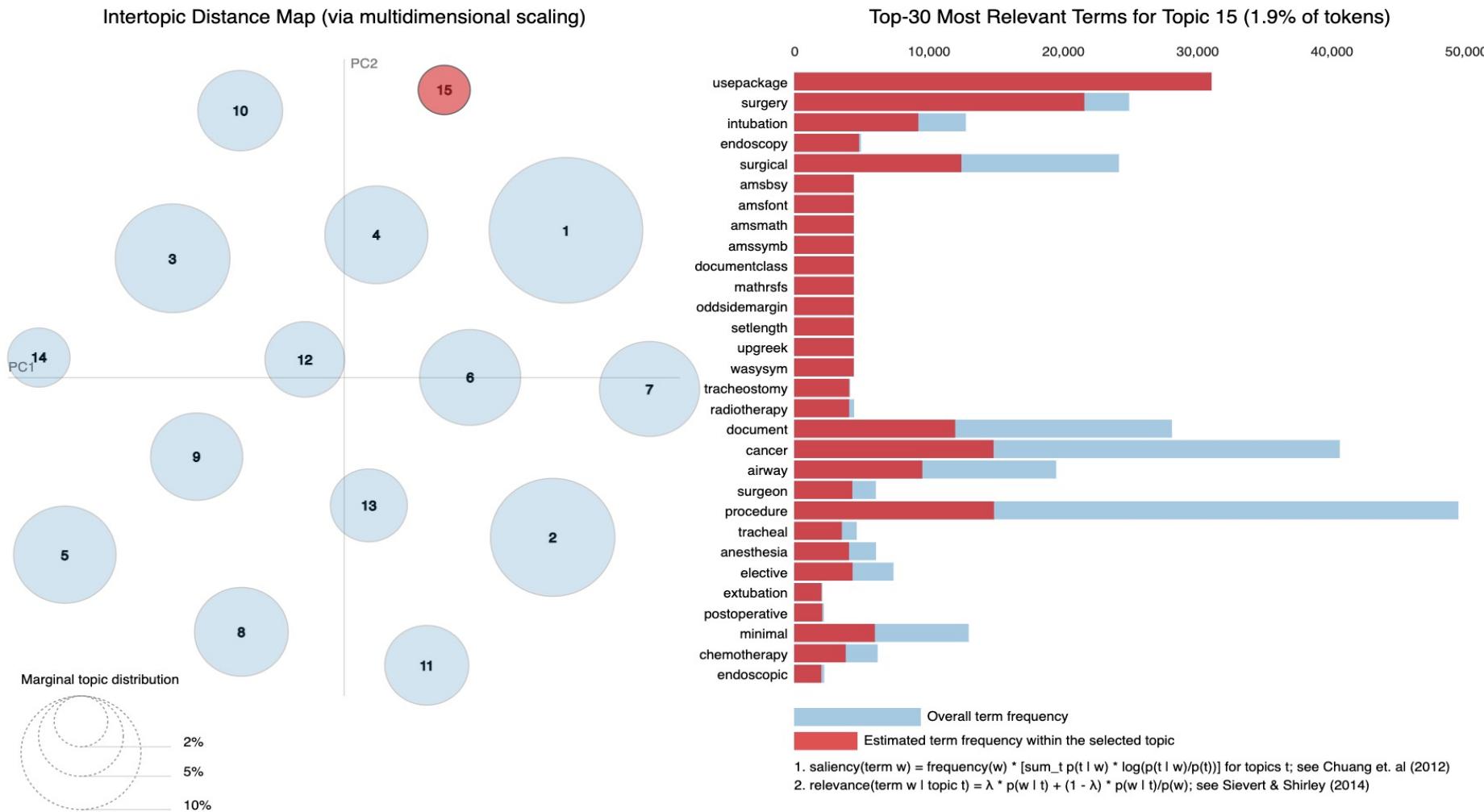
LDA with GridSearchCV to determine the optimal number of components

—

Top 25 most commonly used words

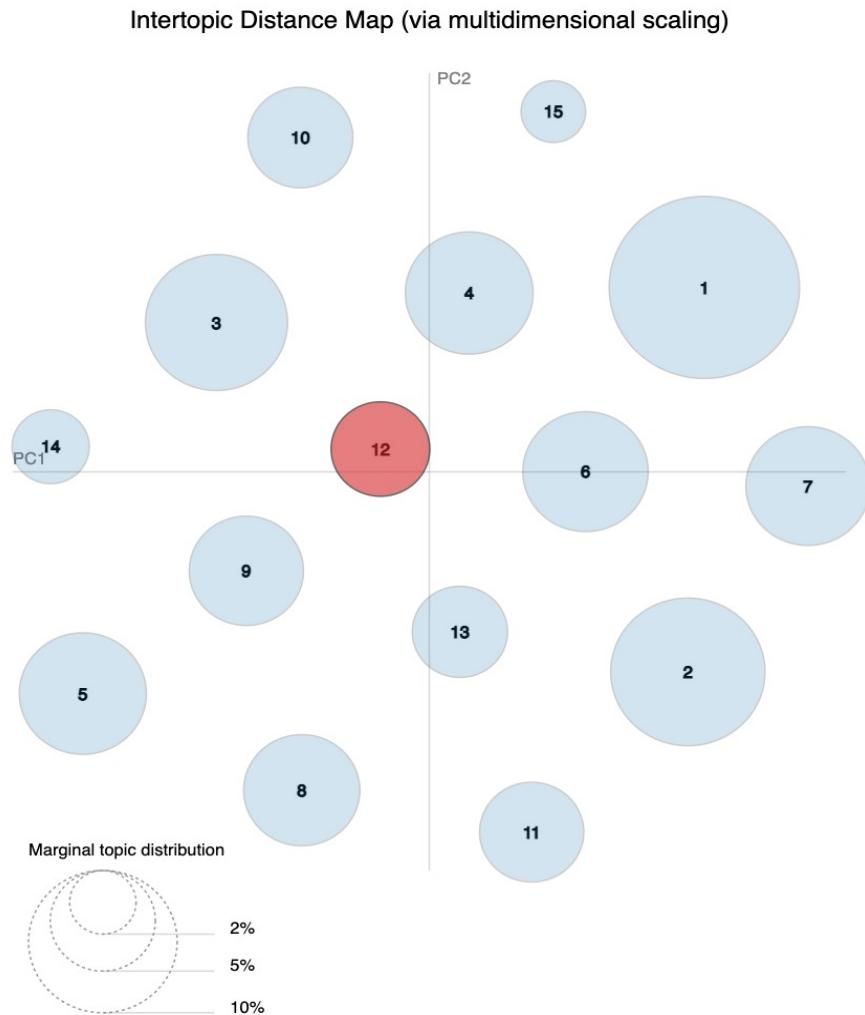


Results of Topic Modeling: LDA with 15 components

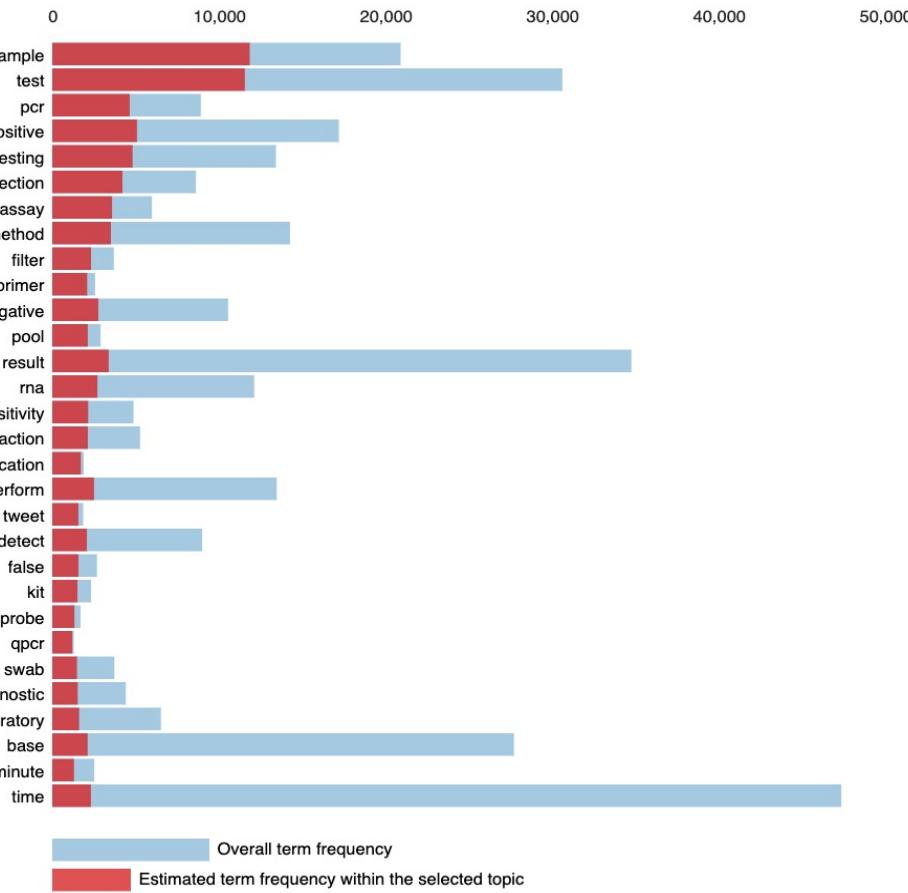




Results of Topic Modeling: LDA with 15 components



Top-30 Most Relevant Terms for Topic 12 (4.4% of tokens)



1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)



15 Topics

15 Topics

- **Title:** SARS-COV-2 infection (coronavirus disease 2019) for the gastrointestinal consultant
- **Abstract:** The current pandemic due to the severe acute respiratory syndrome coronavirus 2 has caused an extreme burden for health care systems globally, and the number of cases is expected to continue to increase...
- **Results:** Pneumonia is the most common serious clinical manifestation of COVID-19, with fever, fatigue, myalgia, and dry cough being the most common features[6]. Other common symptoms include the anorexia...
- **Text:** Since its emergence in December 2019, the coronavirus disease 2019 (COVID-19) has spread to over 146 countries and has been declared a pandemic by the World Health Organization (WHO). The virus...

15 Topics

- **Title:** SARS-COV-2 infection (coronavirus disease 2019) for the gastrointestinal consultant
- **Abstract:** The current pandemic due to the severe acute respiratory syndrome coronavirus 2 has caused an extreme burden for health care systems globally, and the number of cases is expected to continue to increase...
- **Results:** Pneumonia is the most common serious clinical manifestation of COVID-19, with fever, fatigue, myalgia, and dry cough being the most common features[6]. Other common symptoms include the anorexia...
- **Text:** Since its emergence in December 2019, the coronavirus disease 2019 (COVID-19) has spread to over 146 countries and has been declared a pandemic by the World Health Organization (WHO). The virus...

Topic 9 – Symptoms Analysis

15 Topics

- **Title:** SARS-COV-2 infection (coronavirus disease 2019) for the gastrointestinal consultant
- **Abstract:** The current pandemic due to the severe acute respiratory syndrome coronavirus 2 has caused an extreme burden for health care systems globally, and the number of cases is expected to continue to increase...
- **Results:** Pneumonia is the most common serious clinical manifestation of COVID-19, with fever, fatigue, myalgia, and dry cough being the most common features[6]. Other common symptoms include the anorexia...
- **Text:** Since its emergence in December 2019, the coronavirus disease 2019 (COVID-19) has spread to over 146 countries and has been declared a pandemic by the World Health Organization (WHO). The virus...

Topic 9 – Symptoms Analysis

Can be found by: ['study', 'respiratory', 'symptom', 'infection', 'patient', 'cough', 'chest',...]



FUTURE WORK

- Build an app that will be showing articles based on topics and words used in the text of those articles
- Build a recommendation system that will be showing research articles similar to the current one



Thank you