

ENHANCE INFORMATION RETRIEVAL AND SUMMARIZATION ACCURACY IN BIOMEDICAL LITERATURE SYSTEM

PRIYESH TANDEL (202101222)

SHYAM SAKTAWAT (202311048)

AYUSH KUMAR SAHU (202311066)

ABHISHEK CHOUDHARY (202311067)

G13 RUN TIME ERRORISTS

CONTENTS

01 PROBLEM STATEMENT

02 PROPOSED SOLUTION

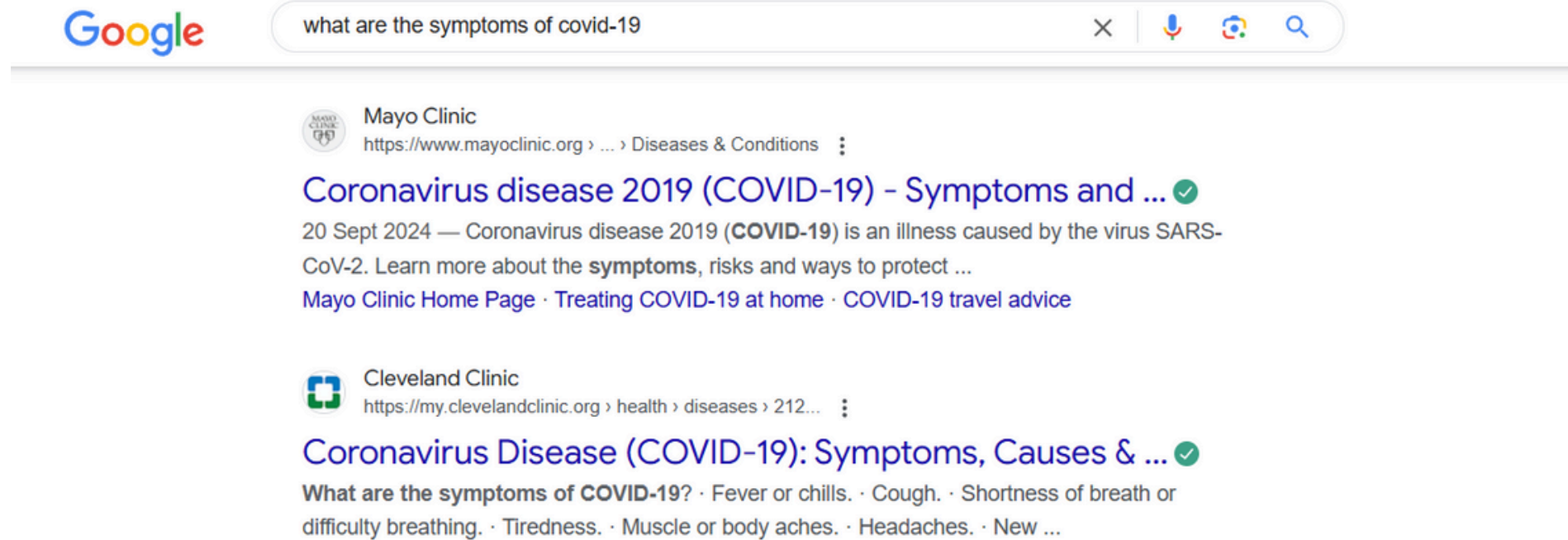
03 METHODOLOGY

04 DATASET

05 EVALUATION

06 RESULTS

PROBLEM STATEMENT



- THE HUGE AMOUNT OF COVID-19 RESEARCH PAPERS IS TOO MUCH FOR TRADITIONAL SEARCH SYSTEMS TO HANDLE.
- THERE IS AN URGENT NEED FOR TOOLS THAT CAN QUICKLY ANSWER COMPLEX MEDICAL QUESTIONS.
- CURRENT TOOLS FOR SUMMARIZING LARGE NUMBERS OF SCIENTIFIC PAPERS ARE NOT EFFECTIVE.

PROPOSED SOLUTION

EFFICIENT INFORMATION RETRIEVAL

- INTEGRATED FRAMEWORK COMBINES BM25, BIOSENTVEC, AND PAGERANK ALGORITHMS.

ADVANCED QUESTION ANSWERING

- CAPABLE OF ANSWERING COMPLEX MEDICAL QUESTIONS FROM THE RETRIEVED INFORMATION.

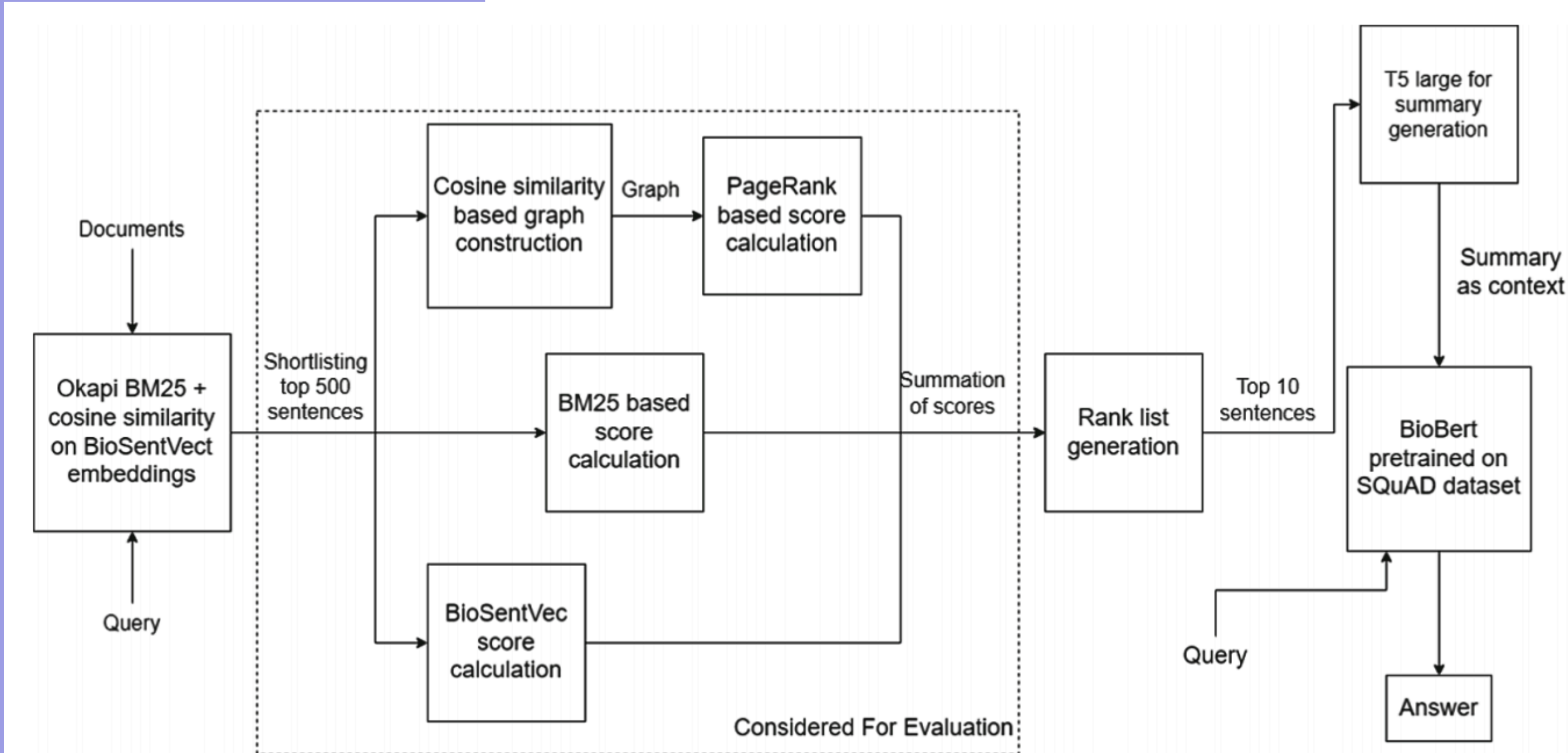
INTELLIGENT SUMMARIZATION

- GENERATES CONCISE SUMMARIES FROM MULTIPLE RESEARCH PAPERS.

INTEGRATED SOLUTION FOR INFORMATION ACCESS

- COMBINES IR, QA, AND SUMMARIZATION IN A SINGLE FRAMEWORK.

METHODOLOGY



Proposed workflow

DATASET

CORD-19

- CORD-19 IS A RESOURCE OF OVER 1,000,000 SCHOLARLY ARTICLES, INCLUDING OVER 600,000 WITH FULL TEXT, ABOUT COVID-19, SARS-CoV-2, AND RELATED CORONAVIRUSES.
- CONTAINS 19 COLUMNS.
- PRIMARY DATASET FOR INFORMATION RETRIEVAL IN THIS STUDY.

CovidQA

- QUESTION-ANSWERING DATASET, BASED ON THE CORD-19 DATA.
- 600+ PAIRS QUERIED FROM 26 STACK EXCHANGE COMMUNITIES, GROUPED IN 3 DISTINCT CATEGORIES
- USED TO ASSESS THE PERFORMANCE OF THE INFORMATION RETRIEVAL AND QUESTION-ANSWERING SYSTEM.

EVALUATION

- **PRECISION@1 ($P@1$):** FOR A SINGLE QUERY, $P@1$ IS 1 IF THE TOP SENTENCE CONTAINS THE ANSWER, AND 0 OTHERWISE.
- **RECALL@3 ($R@3$):** FOR A SINGLE QUERY, $R@3$ IS CALCULATED AS THE NUMBER OF RELEVANT SENTENCES IN THE TOP 3 DIVIDED BY THE TOTAL NUMBER OF SENTENCES CONTAINING THE ANSWER.
- **MEAN RECIPROCAL RANK (MRR):** FOR A SINGLE QUERY, THE RECIPROCAL RANK (RR) IS THE INVERSE OF THE RANK OF THE FIRST SENTENCE CONTAINING THE ANSWER.
- MRR IS THEN CALCULATED AS THE MEAN OF INDIVIDUAL RR VALUES ACROSS ALL QUERIES.

RESULTS

Evaluation Metrics:

Precision@1: 0.169

Recall@3: 0.267

MRR: 0.344

RESULTS

Query: What are the symptoms of COVID-19?

Summary: coronavirus disease 2019 (COVID-19) mainly presents with persistent fever, cough, and general malaise, which may become severe or fatal . neurological symptoms are frequent in COVID-19. upper respiratory tract symptoms including acute pharyngitis are the common symptoms .

Answer: persistent fever, cough, and general malaise

Query: Are masks effective against COVID-19?

Summary: face masks were not recommended for use by the UK public at the beginning of the COVID-19 pandemic . face masks are an important component in controlling COVID-19, and policy orders to wear masks are common . little is known about their social and behavioral consequences .

Answer: face masks are an important component in controlling covid - 19

Query: Can children get COVID-19?

Summary: the COVID-19 pandemic and quarantine undergone by children in many countries is a stressful situation about which little is known to date . parents in Saudi Arabia were generally positive toward children's COVID-19 vaccination . most frequent coping strategies were accepting what is happening (58.9%), collaborating with quarantine social activities (e.g., drawings on the windows, supportive applause)

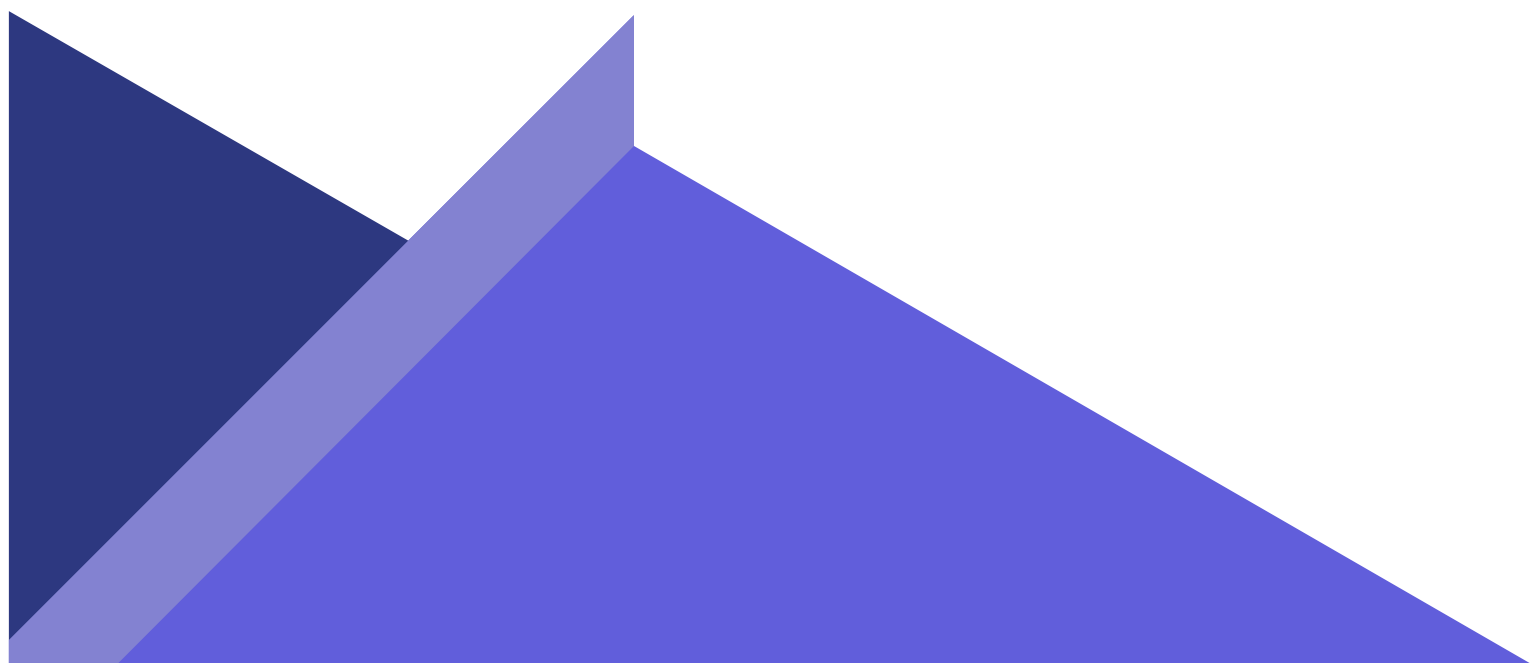
Answer: parents in Saudi Arabia were generally positive toward children ' s covid - 19 vaccination



Thank You

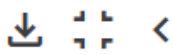
BY RUN TIME ERRORISTS

APPENDIX



CORD19

processed_covid_papers.csv (1.5 GB)



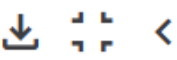
Detail Compact Column

10 of 10 columns

cord_uid	title	authors	publish_time	abstract	journal	license	source_x	abstract_len...	processed_a...	
zhw8vh3e	Artificial Intelligence (AI) applications for COVID-19 pandemic	Vaishya, Raju; Javaid, Mohd; Khan, Ibrahim Haleem; Haleem, Abid	2020-04-14	BACKGROUND AND AIMS: Healthcare delivery requires the support of new technologies like Artificial In...	Diabetes Metab Syndr	no-cc	PMC	1559	['background aims healthcare delivery requires support new technologies like artificial intelligence...	
542e5h1u	Coronavirus disease: challenges for psychiatry	Kelly, Brendan D.	2020-04-15	Coronavirus disease (COVID-19) presents two urgent health problems: the illness caused by the virus ...		cc-by	PMC	371	['coronavirus disease covid presents two urgent health problems illness caused virus anxiety panic p...	
72s8wp1a	Practical Considerations When Performing Neurodiagnostic Studies on Patients with COVID-19 and Other...	Haines, Seline; Caccamo, Amy; Chan, Fonda; Galaso, German; Catinchi, Alexis; Gupta, Puneet K.	2020-05-06	The coronavirus disease 2019, SARS-COV-2 (the cause of COVID-19), has led to a worldwide shortage of...	Neurodiagn J	no-cc	PMC	1581	['coronavirus disease sars cov cause covid led worldwide shortage personal protective equipment ppe ...	
ehw61gu6	COVID-19: Rätsel um den Verlauf bei Kindern: Coronavirus-	Reinhardt, D.	2020-05-14	Auch ein halbes Jahr nach dem Auftauchen von SARS-CoV-2 weiß die Fachwelt	MMW Fortschr Med	no-cc	PMC	221	['auch ein halbes jahr nach dem auftauchen von sars cov wei	



COVIDQA

community.csv (2.61 MB)



Detail Compact Column

10 of 10 columns

question_id	title	question	answer_id	answer	answer_type	wrong_answer	wrong_answer_t...	url
The database ID for the question. You can use this in the stackexchange query system to retrace the original question.	Title of the post	The question body	The database ID for the answer. You can use this in the stackexchange query system to retrace the original answer.	The correct answer, selected based on answer_type	Either Accepted or Reasonable Answer. The former is designated as accepted by the original question author, the	An incorrect answer, selected based on wrong_answer_type	Either Random, or Bad Answer. The latter is an answer that has a score of 0 or less.	Only the subdo the exact URL i given. but can l using the quest
 6235542k	642 unique values	642 unique values	 7195542k	642 unique values	Accepted52% Reasonable48%	640 unique values	Random85% Bad Answers15%	travel.stackexc health.stackex Other (406)
14057	Can pets catch the cold?	Last night I was drying my cat with a towel after shenanigans in the rain and she sneezed! Questio...	14083	Yes they can. The viruses that cause a cold include rhinoviruses (up to 80%), coronaviruses and infl...	Accepted	That is a Priapulid worm, also known as a "penis worm". In its own phylum, Priapulida. https://en.wi...	Random	biology.stacl e.com
89709	Is the Common Cold an Immune Overreaction?	It's my understanding that the majority of symptoms associated with the common cold (runny nose, inf...	89712	Can someone die of the common cold? No. The common cold is a clinical syndrome restricted to upper...	Accepted	The dash ("-") does not represent a negative charge or an electron. It simply indicates that the gr...	Random	biology.stacl e.com
89886	Air purifier agains bacteria and viruses?	We would buy a mobile air purifier in our home. Against the smoke of the city, etc. It would hav...	89887	The aforementioned filter will filter microbes Yes, it is routine practice to use microfiltration i...	Accepted	It's a bleu ray gelyfish, don't tauch is because it's toxic where did oyu found it from did you ofun...	Random	biology.stacl e.com