ENHANCING SEMANTIC INFORMATION RETRIEVAL (SIR) THROUGH ANTONYMS EXTRACTION FOR RETRIEVING PRECISE COVID-19 INFORMATION

M. Rahmah[1*], Muhammad Ahsan Raza[2], Z. Fauziah[3], Hassan Ibrahim[4], Mohd Khairul Azmi Bin Hassan[5]

[1, 3, 4] Faculty of Computing, University Malaysia Pahang Al-Sultan Abdullah,
26600 Pekan, Pahang, Malaysia

[2] Department of Information Sciences, University of Education Lahore,
Multan Campus 60000, Pakistan

[5] Department of Information Systems, International Islamic University Malaysia,
P.O. Box 10, 50728 Kuala Lumpur, Malaysia

Emails: drrahmah@umpsa.edu.my[1*], ahsan.raza@ue.edu.pk[2] , fauziahz@umpsa.edu.my[3], pez24004@adab.umpsa.edu.my[4], mkazmi@iium.edu.my[5]

ABSTRACT

The Semantic Web extends the capabilities of the traditional Web by enabling machines to process and interpret data through ontology knowledgebase. Integrating ontologies into the Web facilitates more accurate and precise searches, task automation, and optimized integration between systems. This research work focuses on semantic information retrieval (SIR) for COVID-19-related queries, leveraging ontologies to generate precise search results and antonyms to reduce irrelevant results. By conducting syntactic and semantic analysis, the system expands the search query using the context derived from the ontology. The query is further refined by extracting antonyms via the ontology relations. The refined query is then submitted to the search engine to retrieve more precise results. A ranking module further filters and prioritizes the most pertinent result links. The SIR approach is novel among existing information retrieval systems in that it eliminates irrelevant search results via antonyms, rather than displaying all the retrieved results based on the query, and in that it re-ranks the results semantically. The SIR algorithm demonstrates significant performance improvements for most queries, primarily due to the semantic analysis, antonyms addition and re-ranking processes. The query dataset achieved 100% precision and 80% recall, outperforming existing search engines in these metrics.

Keywords: Ontology; COVID-19; Semantic analysis; Information retrieval; Antonym analysis

1.0 INTRODUCTION

Internet users heavily rely on search engines to locate the information they need. Most traditional search engines extract data based on keywords provided in user queries, a method that, while effective to some extent, often yields imprecise and irrelevant results. This shortfall arises because these systems fail to interpret the contextual meaning of the query. For complex searches, users might turn to specialized web services, as conventional search engines cannot handle such requests automatically. Researchers have proposed machine-readable knowledge bases, such as ontologies, to enhance the user search experience [1]. An ontology defines the context of a specific domain [2] and helps in interpreting the meaning behind the domain's concept.

This research introduces a method that improves search results by analyzing the semantics and antonyms of a search query, with a specific application to COVID-19-related information. To achieve this, a COVID-19-specific ontology has been developed to serve as the knowledge base. The ontology incorporates predefined rules that refine the search process by filtering out irrelevant data.
Despite technological advancements and the availability of vast amounts of online data, finding accurate and relevant information remains a challenge. This issue is particularly acute in fields like education and healthcare. In education, extensive data on universities and colleges is uploaded regularly. Similarly, in healthcare, diseases like COVID-19 generate massive volumes of data daily. As this data volume continues to grow, search engines often fail to deliver satisfactory results due to irrelevant links and a lack of semantic understanding of the query's context.

In response to these challenges, we propose a comprehensive Semantic Information Retrieval (SIR) approach for COVID-19 information searches. The SIR system focuses on four critical aspects to improve search accuracy:

i. Analyze the query syntactically with the help of parsing.
ii. Expanding the query using domain-specific terms that are retrieved from the COVID-19 ontology.
iii. Extract the query antonyms to formulate a refined query.

iv. Ranking of semantically relevant results.

2.0 LITERATURE AND EXISTING SYSTEMS

Semantic web searches are gaining significant notice from users due to their ability to deliver highly relevant and accurate results. Various techniques for semantic information retrieval have been developed, each with unique features. This section reviews existing systems and their distinct methodologies. Semantic orientation in information retrieval is essential for enhancing the performance of search engines. Studies in the literature leverage semantic web technologies, such as ontology construction, to improve retrieval efficiency [1,3,6]. One notable example is the system proposed in [4], which employs fuzzy ontology to interpret user query contexts. The fuzzy ontology is created by combining domain-specific ontologies with the ConceptNet ontology. This system extracts contextually relevant concepts from the ontology and retrieves results through well-known search engines like Google and Yahoo, achieving up to 89% precision across various queries. Another system, RISSA [5], employs a socially driven method to process query keywords by considering the context of retrieved information. This approach resolves semantic ambiguities, resulting in more accurate outputs for Arabic Hadith. By incorporating both social and semantic dimensions into Hadith interpretation, the system successfully addresses morphological ambiguities in Hadith texts.

The CO-Search retrieval system, introduced in [7], is tailored for handling complex COVID-19-related queries. It uses a deep learning algorithm to derive query semantics and identify critical keywords. A re-ranking method prioritizes results based on their semantic relevance and the summaries of retrieved documents. The system demonstrates strong performance when evaluated on the TREC dataset. Another intelligent retrieval system for handling legal issues related to COVID is presented by [11]. The system collects the legal knowledge, trains the CNN model on the collected knowledge, and uses a semantic matching mechanism to retrieve the legal cases relevant to a user query. The system performance is evaluated over WeChat in terms of relevant case retrieval and compared with the baseline models. Overall, the system achieved better results than other models.

A novel mechanism to map natural language queries to documents is present in [8]. This method explores variations such as supervised, unsupervised, dense, and sparse approaches. The framework improves similarity scores between natural language queries and corresponding documents. Additionally, the LDA Bayesian model proposed in [9] effectively mines semantic correlations from metadata ontologies, achieving a remarkable 90% accuracy in results. In 2022, a detailed review of semantic models designed for first-stage retrieval is presented [10]. The review mainly focuses on each model's role in improving query understanding and relevance matching. The study explores neural and traditional approaches for generating candidate document sets efficiently in large-scale retrieval tasks. The research in [12] introduces a unique framework that combines text mining and deep learning methods to automate semantic data extraction and boost geological information searchability. The methodology employs Latent Dirichlet Allocation (LDA) to extract keywords and constructs topic graphs for theme identification alongside knowledge graphs to map geological entity relationships. The system uses visualization tools to enable users to explore extracted information through intuitive methods. The proposed system shows significant efficiency gains in processing vast mineral exploration reports through its capability to rapidly detect target mineral systems together with geological site locations and related rock compositions. The proposed method minimizes human intervention while improving decision-making processes in mineral exploration tasks. Artificial intelligence-driven methods show potential to transform unstructured geological data into structured knowledge which propels advancements in geological information retrieval and analysis. Similarly, authors in [14] describe the SE-BERT model, which retrieves text by applying abstractive summarization methods to the text passage. The model makes use of a generative pre-trained language model to summarize user queries and text passages. The summaries are concatenated and fed to the model as a single sequence. The model can process all summative content semantically. Using two TREC datasets, experiments show that the model achieves its highest performance for longer texts. This means a large summary enables the model to understand the relationship between the queries and the documents on a much deeper level.

These studies highlight various approaches for semantic information retrieval, each addressing specific challenges. However, most existing systems perform semantic analysis in a generalized manner [10], which may limit their effectiveness in specialized domains. After the detailed analysis of existing approaches, it has been observed that some information retrieval systems have not provided the detail of context extraction from the ontology, while others are specialized for particular domains(e.g., for the health domain only). On the other hand, existing approaches do not refine the query (after context retrieval and query expansion) to minimize the noise in the retrieved results. In last, prioritization of the results for an individual or organization is missing.

To address this limitation, a new system called Semantic Information Retrieval (SIR) has been developed, specifically designed for COVID-19-related queries. SIR combines several existing methodologies and incorporates new procedures to enhance the extraction of semantically relevant information.

2.1. Contributions of SIR

SIR focuses on providing precise results for medical queries related to COVID-19. After conducting syntactic and semantic analysis, the system expands a user query using keywords extracted from a domain-specific ontology. Furthermore, the antonyms of the query are identified and added to the query to formulate a refined query. The refined query is then sent to a search engine to retrieve the most relevant results. Finally, a ranking algorithm filters and ranks the retrieved links to ensure the most relevant results are presented on top of the list.

3.0 INITIAL PROCEEDINGS

As described in [2], an ontology is a "formal specification of a shared conceptualization," providing a structured hierarchy of terms and their relationships. Ontologies serve as knowledge bases, offering a common understanding of terms across different systems. In the medical domain, terminology often varies between hospitals, which can complicate query interpretation. To address this, user queries are parsed to eliminate stop words and identify parts of speech such as nouns and verbs. WordNet is used to retrieve noun synsets for further refinement, while nouns are employed to extract related terms and antonyms from the COVID-19 ontology.

The proposed SIR algorithm, illustrated in Fig. 1, outlines the entire process, ensuring efficient and accurate retrieval of COVID-19-related information. We first normalize the query to obtain the noun tokens (see lines 1-7). On lines 8-11, we extract the synonyms of noun tokens from the WordNet ontology to access the sense of each token. We proceed with the context retrieval of each token and its corresponding senses. If the token is matched with the ontology concept, we simply extract the context of the matched concept, otherwise, we proceed with matching the senses of a token with the ontology to retrieve the context (see lines 12-18). On lines 19-24, the antonyms are obtained from the ontology based on retrieved context. Lines 25 and 26 generate a refined query using the antonyms. This refined query is useful in reducing the unrelated results. In the last, results obtained from Google are re-ranked to prioritize the related documents as given on lines 27-31. Overall, this structured approach enhances query interpretation, relevance, and the overall quality of search results.

4.0 PROPOSED ARCHITECTURE

The process of SIR for COVID-19 includes four steps as shown in Fig. 2.

4.1. Query Parser (Step 1)

The Query Parser evaluates the semantic context and meaning of keywords in user queries through a step-by-step process involving multiple modules:

i. *Query Parsing:* Keywords from the user’s query are parsed to identify grammatical elements such as nouns, verbs, and adjectives. This analysis helps in interpreting the structure and meaning of the query. For this purpose, the Stanford Parser tool is employed, enabling precise syntactic parsing.
ii. Synset Retrieval: The parsed terms from the user query undergo semantic analysis to extract their corresponding synsets from WordNet. Synsets provide sets of synonyms that help identify related meanings and enhance the query's semantic relevance. This ensures that the query's intent is accurately represented. The WordNet API is used for this task, enabling efficient and precise retrieval of synsets to align with the contextual nuances of the user query.

Algorithm 1 SIR Algorithm

Input : Q , The user search query
CO, COVID ontology

Output : Q′ , The refined query

```
// Query parsing
FOR each term in Q
   Remove term from Q if it belongs to stopword list
   Obtain part-of-speech of term if its not a stopword
    IF term is Noun
        Add term (called as token) in Q
    END-IF
END-FOR
// Synset retrieval
FOR each Token (Ti) in Q
  Extract Synset (s) of Ti from WordNet
  Add s in Q
END-FOR
// Context retrieval
FOR each Token (Ti) in Q
  IF Ti is matched with CO //part-of or exact match
    Extract keywords (Ks) of match CO via class label
    Extract keywords (Ks) of match CO via instances
    Add Ks in Q to from expanded query EQ
  END-IF
END-FOR
// Antonym extraction
FOR each Keyword (Ki) in EQ
  IF Ki is matched with CO //exact match
    Extract antonyms (As) of Ki via antonym relation
    Add As antonym list AL
  END-IF
END-FOR
// Query refining
Calculate similarity weight of each As in AL based on original query Q
Append high weight As in EQ by putting a minus sign - immediately in front of each As to create a refined query Q′
// Links ranking
FOR each retrieved Link (Li) via Google for Q′
  Calculate weight of link based on meta-tag match
  Assign weight to link
END-FOR
Sort links based on calculated weight
```

Fig.1: SIR algorithm

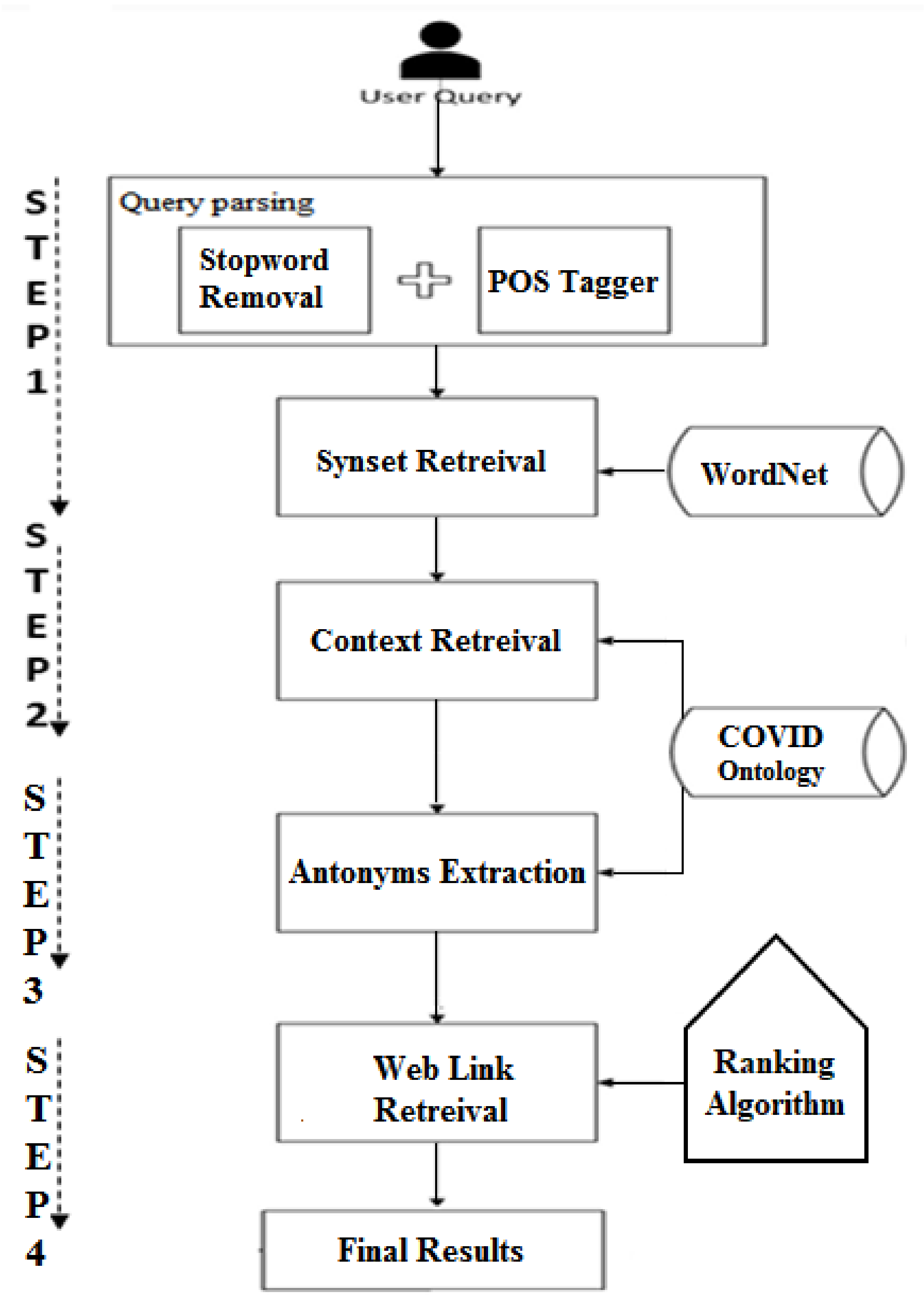

Fig. 2: Architecture of SIR

4.2. Semantics Extractor (Step 2)

This step involves extracting the context of the user query through three modules.

i. COVID-19 Ontology: A specialized ontology is created using COVID-19-related concepts and keywords to facilitate effective semantic information retrieval. The information is sourced from various health-focused platforms, hospital case studies, and WordNet. The extracted keywords and concepts are systematically organized into a hierarchical structure, comprising classes, subclasses, superclasses, and instances. These elements are interconnected through specific properties that define their relationships. In addition, an antonym relationship is added to represent the antonyms of each ontology element. The COVID-19 ontology is developed using Protégé software, with version 4.1 being utilized for this research. This ontology consists of 33 concepts and five relations. Fig. 3 shows a graphviz snapshot of the ontology depicting COVID-19 concepts and relationships.

ii. Context Retrieval: The module involves extracting the semantic context of the query by mapping its keywords to concepts, properties, and relationships within the ontology. This process identifies partial matches between user keywords and ontology concepts and retrieves associated instances, enriching the domain-specific keywords. For example, if a query keyword matches a concept such as "virus," related instances like "COVID-19" and its associated properties are retrieved. The Jena API is utilized to navigate the ontology, ensuring that the expanded context is comprehensive and accurately reflects the domain knowledge. This step is critical for bridging the gap between the user's intent and the structured knowledge contained within the ontology.

iii. Query Expansion: By using an enriched set of ontology-derived keywords, an expanded query is constructed to improve the relevance of search results. Boolean operator + is applied to combine keywords in a way that highlights their semantic relationships. For instance, if the original query was ambiguous, the expanded query would focus on domain-specific terms ensuring an optimal balance between query specificity and user intention.

4.3. Antonyms Extractor (Step 3)

The user query is further refined by appending antonyms in the query. The antonym extractor comprises two modules.

i. Antonym Retrieval: The module involves extracting the antonyms of the query antonym relationship within the ontology. This process identifies exact matches between expanded query keywords and ontology concepts. The Jena API is utilized to navigate the antonym relation with the ontology.

ii. Query Formulation: Using the antonyms, a refined query is constructed to improve the relevance of search results. Additional operator + is applied to combine antonyms with the expanded query (i.e., achieved in step 2). This process ensures that the final refined query aligns with the intended meaning of the user’s input, effectively prioritizing relevant results while filtering out irrelevant ones. For instance, if the original query was ambiguous, the refined query would focus on domain-specific terms and relationships, reducing noise in the results. The refined query is then submitted to the search engine to provide users with targeted and meaningful outputs. This iterative refinement process is central to ensuring an optimal balance between query specificity and relevance.

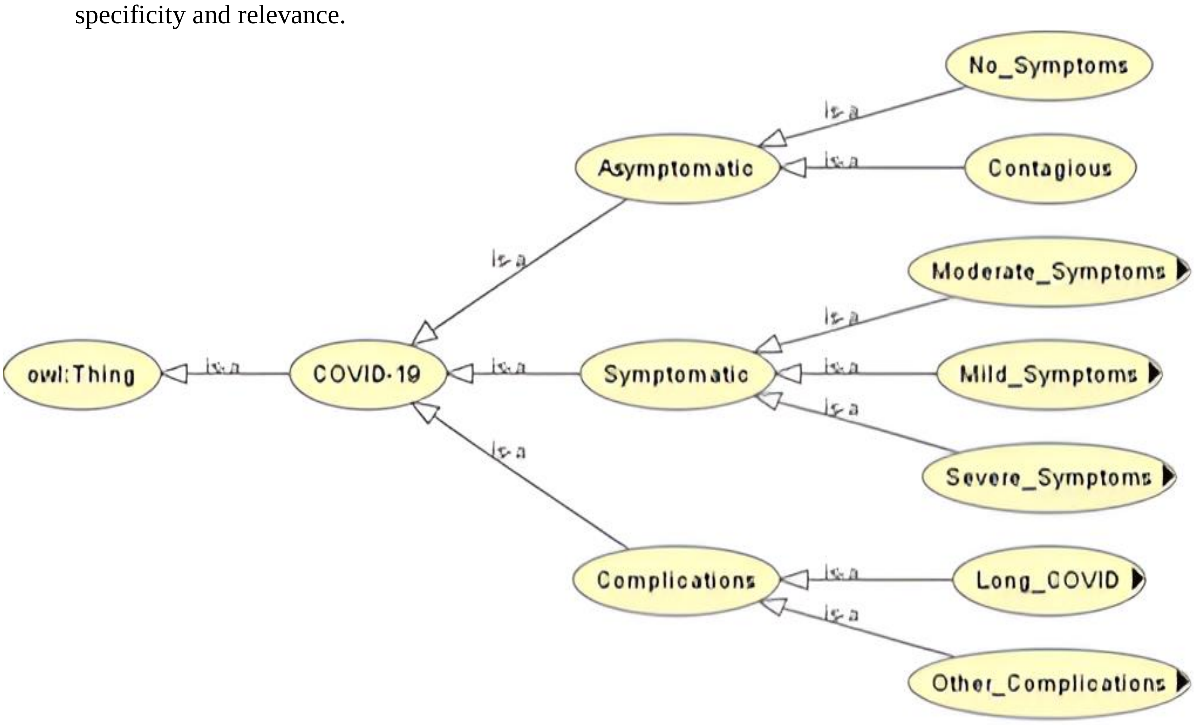

Fig. 3: Graphviz snapshot of the ontology up to three levels

4.4 Web Link Ranking (Step 4)

The process of web link ranking involves retrieving links using refined queries and ranking them based on their relevance to the user's intent. The Google Search API is used to fetch these links, and a weighting system evaluates their relevance by comparing meta tags and ontology keywords.

i. Retrieval of Web Links: Refined queries, along with the original user query, are submitted to the Google search engine to fetch the most relevant web links. The system retrieves the top-ranking links that best align with the search criteria.

ii. Ranking of Retrieved Links: Once the web links are retrieved, they are ranked based on their relevance to the user query. This ranking is achieved by assigning weights to each link, determined through a comparison of meta tags with the keywords extracted from the query and the ontology. Links that exhibit a higher degree of alignment with the query semantics are ranked higher, ensuring that users receive the most appropriate results.

5.0 SIR PERFORMANCE

A. Performance Evaluation Metrics

This research work computes recall and precision values to evaluate the SIR performance.

i. Recall*:* Measures the coverage of relevant information extracted by the SIR system (see (1)).

$$\text{Recall} = \frac{\text{\# of relevant retrieved links}}{\text{Total \# of relevant links}} \quad (1)$$

ii. Precision*:*Measures the retrieval of relevant information by the SIR system (see (2)).

$$\text{Precision} = \frac{\text{\# of relevant retrieved links}}{\text{Total \# of retrieved links}} \quad (2)$$

B. Dataset Selection

We have carefully selected a COVID dataset that includes a wide range of concepts and related documents [13]. We have used queries from this dataset for testing purposes as the dataset is prepared by the experts. Table 1 shows the ten sample queries from this dataset.

Table 1: Set of Sample Queries [13]

No.	Sample Query
1	COVID-19 testing type
2	What are the vaccines for COVID-19
3	What causes death from COVID-19
4	Looking for studies identifying ways to diagnose COVID-19 more rapidly
5	Alpha and Beta, had emerged and started to spread throughout the world
6	What types of masks should or should not be used to prevent infection by COVID-19
7	Studies assessing chemicals and their concentrations needed to destroy the COVID-19 virus
8	Studies of patients and the first clinical manifestations they develop upon active infection
9	Studies of people who are known to be infected with COVID-19 but show no symptoms
10	Seeking specific information on clinical outcomes in COVID-19 patients treated with remdesivir

C. Result and Discussion

This research utilizes two variants of SIR: (i) basic SIR system variant that does not add antonyms in the process of query refinement, and (ii) SIR system incorporating antonyms (namely Enhanced SIR). We compared our proposed SIR variants with a baseline technique (i.e., Google search with initial user query). We input two queries: the initial user query and the corresponding refined query via SIR, in the Google IR system.

We have used Google search results (top 100 links) for each type of input query. The basic SIR shows marked improvement in performance for most queries due to semantic analysis. Sample queries demonstrate higher precision and recall values for basic SIR compared to Google. Table 2 shows the precision and recall values of ten sample queries over Google and basic SIR.

Table 2: Precision and recall of queries

Queries	Google		Basic SIR	
	Precision	**Recall**	**Precision**	**Recall**
Q1	0.88	0.64	1.0	0.70
Q2	0.82	0.61	1.0	0.80
Q3	0.88	0.70	0.98	0.74
Q4	0.76	0.63	0.97	0.77
Q5	0.95	0.66	1.0	0.76
Q6	0.73	0.51	0.97	0.76
Q7	0.90	0.65	1.0	0.75
Q8	0.86	0.72	0.93	0.80
Q9	0.76	0.71	0.88	0.74
Q10	0.80	0.75	0.87	0.75

The SIR approach achieves 100% precision for four queries and for the remaining six queries it shows higher precision than Google. Similarly, SIR achieves greater than 70% recall for all ten sample queries, while a maximum of 80% recall is calculated.

A chart (as shown in Fig. 4) comparing precision and recall values for SIR and Google also shows SIR's superior performance in retrieving semantically relevant links. It can be observed that for all ten sample queries, the SIR approach achieved much higher precision and recall as compared to the Google results.

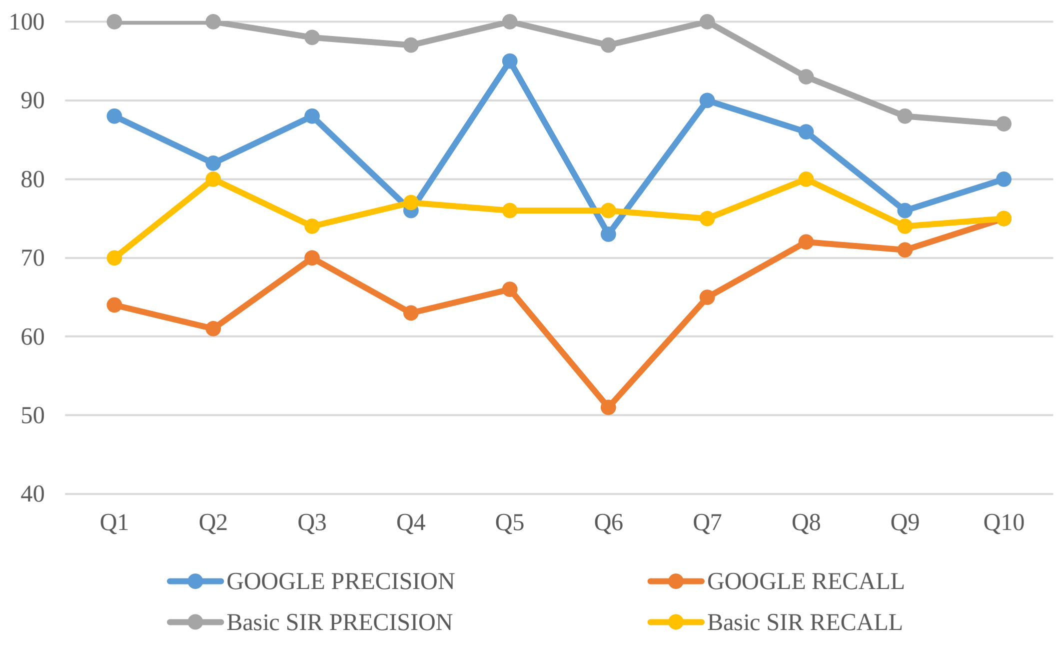

Fig. 4: Sample queries precision and recall graph

Further verifying the accuracy of the proposed system, we have compared the performance of two variants of SIR (i.e., basic SIR and enhanced SIR). This verification is important to show that the proposed system performs even better if antonyms are considered for query refinement. Table 3 shows the precision and recall values of ten sample queries over basic SIR and enhanced SIR. The enhanced SIR achieves 100% precision for six queries which is 20% better than basic SIR results. Considering precision and recall results, our proposed SIR, which incorporates antonyms (i.e., enhanced SIR), outperforms the basic SIR system.

Table 3: Result of queries for two SIR variants

Queries	Basic SIR		Enhanced SIR	
	Precision	**Recall**	**Precision**	**Recall**
Q1	1.0	0.70	1.0	0.80
Q2	1.0	0.80	1.0	0.90
Q3	0.98	0.74	1.0	0.90
Q4	0.97	0.77	1.0	0.88
Q5	1.0	0.76	1.0	0.85
Q6	0.97	0.76	0.99	0.90
Q7	1.0	0.75	1.0	0.88
Q8	0.93	0.80	0.96	0.86
Q9	0.88	0.74	0.92	0.80
Q10	0.77	0.75	0.90	0.79

Fig. 5 shows a graphical view of the precision and recall percentages of queries for SIR variant systems. The enhanced SIR achieved 60% better results in terms of precision than the basic SIR. This improvement suggests that the enhanced SIR method, which expands queries with semantically related concepts and refines queries with antonyms better meets the needs of user search. Considering the recall result of all 10 queries, the enhanced SIR outperforms the basic SIR search system. Thus, the proposed SIR system retrieves far better relevant results than the basic SIR.

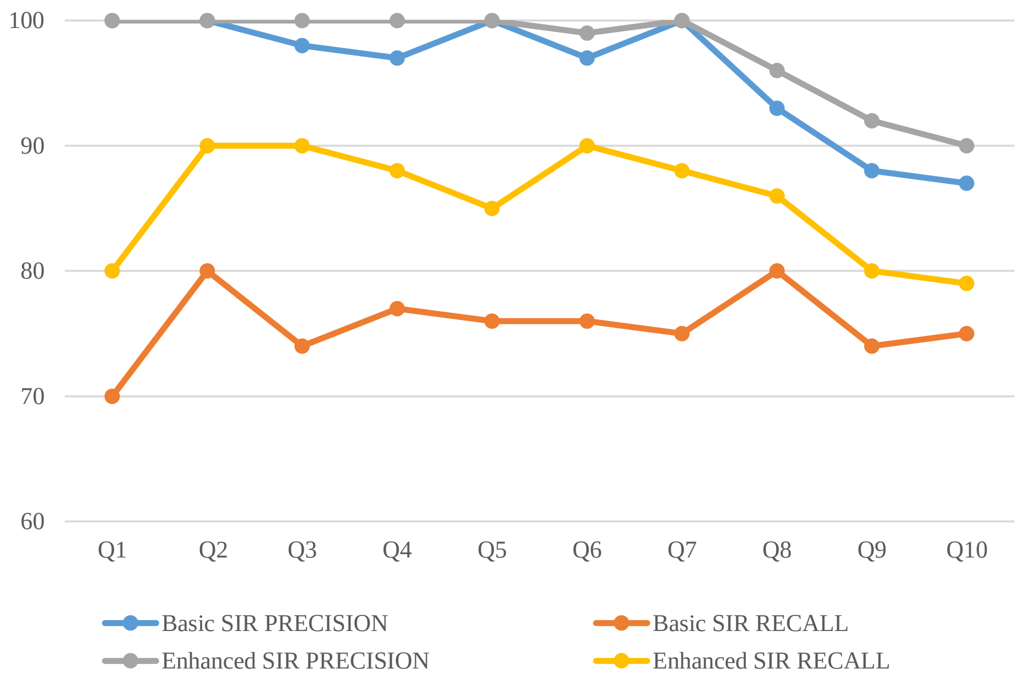

Fig. 5: Precision and Recall Results of SIR Variant Systems

To measure the overall performance of proposed SIR variants and baseline technique, we computed the average precision and recall of the queries. The average precision and recall ratios of the three systems are depicted in Fig. 6. The baseline technique (i.e., Google search) is 83.4% precision, while the proposed SIR (i.e., enhanced SIR) achieved 97.7%. This shows that SIR retrieved results are far more precise than the base system. On the other hand, the enhanced SIR achieved best 85.6% recall than the other two systems. The +19.8% and +9.9% improvements in recall compared to the base system and basic SIR, respectively, clearly represent that enhanced SIR focusing on antonym retrieved increased number of user-relevant results than the other systems.

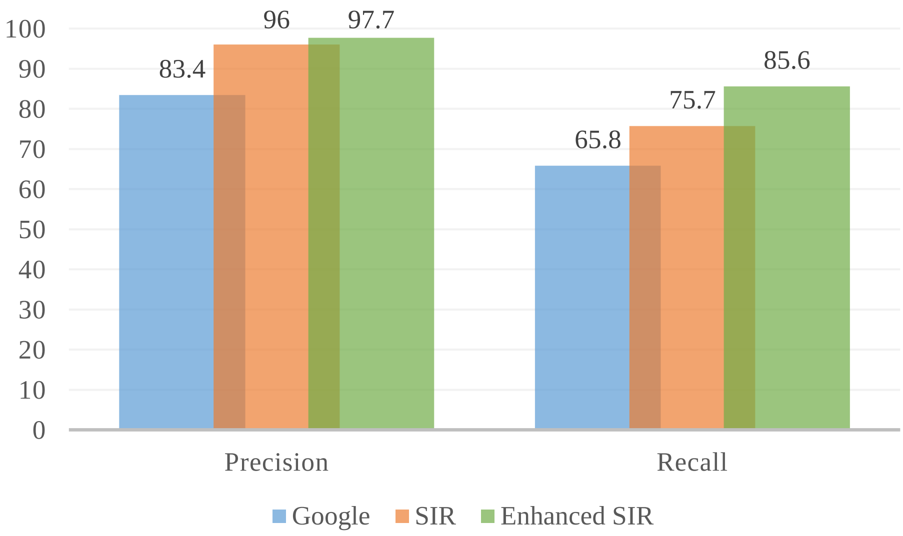

Fig. 6: Overall performance of three systems

6.0 CONCLUSION

The developed SIR algorithm effectively retrieves accurate results for COVID-19-related queries by leveraging advanced semantic techniques (i.e., ontology and antonym retrieval). The COVID-19 ontology, when combined with the antonym relation, provides a robust framework for query refinement. Through a detailed syntactic, semantic and antonym analysis, user queries are enriched with contextual information extracted from the ontology knowledgebase. The refined queries are then submitted to the search engine, and a ranking module organizes the retrieved results based on their relevance. The refined query and ranking process demonstrate a marked improvement in the performance and accuracy of search outcomes. By employing this ontology-driven approach, SIR enables semantic searches that outperform Google searches, providing more precise and contextually relevant results. The precision and recall metrics were used to compare SIR results against those of the Google search engine. To evaluate the effectiveness of SIR, the basic SIR (i.e., without antonym addition) results are compared with enhanced SIR (i.e., with antonym) results. The higher precision and recall scores achieved by enhanced SIR highlight the significant potential of Semantic Web technologies in enhancing search accuracy and efficiency. This advancement underscores the value of integrating ontological knowledge into information retrieval and paves the way for more intelligent and antonym-aware search systems in the future. The SIR key contributions in the field of the semantic web include: (1) Semantic information retrieval for COVID-19 research, (2) More precise results for the searched query using antonym analysis, and (3) The extracted web results are ranked according to the user queries, ensuring relevant and satisfactory results.

In future, the deep learning models (such as CNN) can be merged with the SIR approach. The CNN model can be helpful in the successful prediction of query senses as well as the context of the query. This blend of semantics and deep learning models can be effective in achieving precise information. Furthermore, the SIR can be applied to other health domains for its detailed analysis and authenticity.

ACKNOWLEDGEMENT

This research was funded under the IIUM-UMP Sustainable Research Grant (University Reference RDU223217) and Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan, Pahang.

REFERENCES

[1] V. M. Ngo and T. H. Cao, "Ontology-based query expansion with latently related named entities for semantic text search," *arXiv preprint*, arXiv:1807.05579, 2018.

[2] Hadzic, M., Wongthongtham, P., Dillon, T., Chang, E., Hadzic, M., Wongthongtham, P., & Chang, E., “Introduction to ontology,” Ontology-based multi-agent systems, 37-60, 2009.

[3] Mahalakshmi, P., & Fathima, N. S., “An Art of Review on Conceptual based Information Retrieval,” Webology, 18(1), 2021.

[4] Jain, S., Seeja, K. R., & Jindal, R., “A fuzzy ontology framework in information retrieval using semantic query expansion,” International Journal of Information Management Data Insights, 1(1), 100009, 2021.

[5] Elayeb, B., & Ayed, R., “Socio-Semantic Information Retrieval of Structured Arabic Texts,” In 2022 IEEE/ACS 19th International Conference on Computer Systems and Applications (AICCSA) (pp. 1-8). IEEE, 2022.

[6] Raza, M. A., Mokhtar, R., Ahmad, N., Pasha, M., & Pasha, U., “A taxonomy and survey of semantic approaches for query expansion,” IEEE Access, 7, 17823-17833, 2019.

[7] Esteva, A., Kale, A., Paulus, R., Hashimoto, K., Yin, W., Radev, D., & Socher, R., “COVID-19 information retrieval with deep-learning based semantic search, question answering, and abstractive summarization,” NPJ digital medicine, 4(1), 68, 2021.

[8] Lin, J., “A proposed conceptual framework for a representational approach to information retrieval,” In ACM SIGIR Forum Vol. 55, No. 2, pp. 1-29, 2022, New York, NY, USA: ACM.

[9] Wang, C., & Chen, K., “Retracted Analysis of Network Information Retrieval Method Based on Metadata Ontology,” International Journal of Antennas and Propagation, 2022 (1), 2664639, 2022.

[10] Guo, J., Cai, Y., Fan, Y., Sun, F., Zhang, R., & Cheng, X.,.”Semantic models for the first-stage retrieval: A comprehensive review,” ACM Transactions on Information Systems (TOIS), 40(4), 1-42, 2022.

[11] Zhu, J., Wu, J., Luo, X. et al. “Semantic matching based legal information retrieval system for COVID-19 pandemic,” Artif Intell Law 32, 397–426, 2024. https://doi.org/10.1007/s10506-023-09354-x.

[12] Qinjun Q, Miao T, Liufeng T, Zhong X, Kai M, “Semantic information extraction and search of mineral exploration data using text mining and deep learning methods,” Ore Geology Reviews, 165, 2024. https://doi.org/10.1016/j.oregeorev.2023.105863.

[13] Nist., "COVID19,", Retrieved October 20, 2024, from https://covid19-data.nist.gov.

[14] Pan M, Li T, Liu Y, Pei Q, Huang EA, Huang JX., “A semantically enhanced text retrieval framework with abstractive summarization,” Computational Intelligence. 2024; 40(1):e12603. doi: 10.1111/coin.12603.