

---

Articles

---

2021-11

## RCES: Rapid Cues Exploratory Search Using Taxonomies For COVID-19

Wei Li

*Dublin City University, wei.li@adaptcentre.ie*

Rishi Choudhary

*Technological University Dublin, rishi.choudhry@tudublin.ie*

Arjumand Younus

*University College Dublin, arjumand.younus@ucd.ie*

*See next page for additional authors*

Follow this and additional works at: <https://arrow.tudublin.ie/creaart>



Part of the [Computational Engineering Commons](#)

---

### Recommended Citation

Wei Li, Rishi Choudhary, Arjumand Younus, Bruno Ohana, Nicole Baker, Brendan Leen, and M. Atif Qureshi. 2021. RCES: Rapid Cues Exploratory Search Using Taxonomies For COVID-19. In Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM '21), November 1–5, 2021, Virtual Event, QLD, Australia. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3459637.3481990>

This Conference Paper is brought to you for free and open access by ARROW@TU Dublin. It has been accepted for inclusion in Articles by an authorized administrator of ARROW@TU Dublin. For more information, please contact [arrow.admin@tudublin.ie](mailto:arrow.admin@tudublin.ie), [aisling.coyne@tudublin.ie](mailto:aisling.coyne@tudublin.ie), [gerard.connolly@tudublin.ie](mailto:gerard.connolly@tudublin.ie).



This work is licensed under a [Creative Commons Attribution-NonCommercial-Share Alike 4.0 License](#)  
Funder: Science Foundation Ireland

---

**Authors**

Wei Li, Rishi Choudhary, Arjumand Younus, Bruno Ohana, Nicole Baker, Brendan Leen, and Muhammad Atif Qureshi

# RCES: Rapid Cues Exploratory Search Using Taxonomies For COVID-19

Wei Li  
wei.li@adaptcentre.ie  
Dublin City University

Rishi Choudhary  
rishi.choudhry@tudublin.ie  
Technological University Dublin

Arjumand Younus  
arjumand.younus@ucd.ie  
University College Dublin

Bruno Ohana  
bruno.ohana@adaptcentre.ie  
Trinity College Dublin

Nicole Baker  
nicole.baker@adaptcentre.ie  
Trinity College Dublin

Brendan Leen  
brendan.leen@hse.ie  
Health Services Executive, Ireland

M. Atif Qureshi  
muhammatatif.qureshi@tudublin.ie  
ADAPT Centre, Technological  
University Dublin

## ABSTRACT

To assist the COVID-19 focused researchers in life science and healthcare in understanding the pandemic, we present an exploratory information retrieval system called RCES. The system employs a previously developed EVE (Explainable Vector-based Embedding) model using DBpedia and an adopted model using MeSH taxonomies to exploit concept relations related to COVID-19. Various expansion methods are also developed, along with explanations and facets that collectively form rapid cues for a valuable navigational and informed user experience.

## CCS CONCEPTS

• **Information systems** → **Search interfaces**; • **Applied computing** → **Document searching**.

## KEYWORDS

Exploratory Search, Query Expansion, Facets, Explanations, Rapid Cues, COVID-19

### ACM Reference Format:

Wei Li, Rishi Choudhary, Arjumand Younus, Bruno Ohana, Nicole Baker, Brendan Leen, and M. Atif Qureshi. 2021. RCES: Rapid Cues Exploratory Search Using Taxonomies For COVID-19. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM '21), November 1–5, 2021, Virtual Event, QLD, Australia*. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3459637.3481990>

## 1 INTRODUCTION

The COVID-19 pandemic threw the scientific community into an emergency response mode leading to a plethora of scientific literature on the topic within a brief time frame [6]. Keeping up with the massive amount of ever-increasing literature generated at an

exponential rate [7] is challenging. There is an imperative demand for tools to facilitate navigation of scientific publications related to COVID-19 [2]. Some search and recommendation tools were developed by various research groups to address this demand [3, 9]. However, to the best of our knowledge, systems allowing deep and insightful exploration into COVID-19 concepts are not available; and neither do they provide associated explanations for the retrieved results.

Pursuing a study of the scientific literature is a complex phenomenon, and besides an effective information retrieval (IR) model, there is an additional aspect of exploratory search that needs to be incorporated [8]. The nature of COVID-19 as a virus where researchers, clinicians, virologists, epidemiologists, and health policymakers know very little make the information need concerning the virus very unclear and non-specific [4]. In such a scenario, an exploratory search that supports navigation through facets [10] helps with the organisation of results in a structure while allowing techniques that expand a query, thereby increasing the recall or the search space. Overall, both navigation and explainable query expansion techniques form the basis of what we call rapid cues in this work.

Essentially, on top of rapid cues we propose a human-centric, exploratory semantic retrieval system (via browser), Rapid Cues Exploratory Search System (RCES). In RCES, the user begins the exploration process by defining a query concerning the information need concepts. The system retrieves a connection between the queried concepts and COVID-19 related topics, supporting visual exploratory search by generating a list of suggested relationships and associations useful for researchers in life sciences and healthcare areas. To achieve this, we take advantage of taxonomies in medical science and previous research on explainable embeddings using our model [5]. The main idea of using these resources is to increase the retrieval effectiveness by employing concept relations within our exploratory search process. These relations are extracted from an existing linked data source such as DBpedia<sup>1</sup> and MeSH<sup>2</sup>. In particular, for a given query, rich association concepts related to COVID-19 are extracted from the initial query by using mapping



This work is licensed under a Creative Commons Attribution International 4.0 License.

CIKM '21, November 1–5, 2021, Virtual Event, Australia.

© 2021 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-8446-9/21/11.

<https://doi.org/10.1145/3459637.3481990>

<sup>1</sup><https://www.dbpedia.org/>

<sup>2</sup><https://www.ncbi.nlm.nih.gov/mesh/>

functions over taxonomies (explained in more detail in Section 3). Furthermore, facets are extracted using retrieved results to navigate the scientific publications related to the query topic effectively.

Lastly, and most importantly, this is a first of its kind COVID19 literature search project in which an official health and personal social services body of Ireland is directly involved. Their input also informed some of the case-studies and expert assessment efforts around those case-studies.

## 2 ARCHITECTURE AND METHODOLOGY

This section introduces the architecture of RCES and the methodologies behind its various functionalities. Figure 1 shows that the exploratory search system consists of four components:

### 2.1 Data Preparation

This component is responsible for crawling various COVID-19 related articles from websites periodically, such as PubMed<sup>3</sup>, medRxid<sup>4</sup>, Elsevier<sup>5</sup> and arXiv<sup>6</sup> to prepare the corpus for the system. Duplication between articles is also removed at this stage, and this is done by means of checking the authors and titles of the articles.

### 2.2 Indexing

Within this system component, we build the index. This index is updated periodically in a streaming fashion to cope efficiently with the increasing volume of COVID-19 related articles. The dataset gathered is dated from Dec 2019 to date. Moreover, we also index the data from taxonomies at this stage for efficiency purposes, i.e., MeSH and DBpedia. The crawled articles are indexed in Elasticsearch<sup>7</sup>. The use of Elasticsearch helps incorporate custom query expansion methods and different filters within the search interface in an easy-to-use and intuitive manner. Furthermore, it allows a full-text search over a collection of documents, thereby enabling near real-time search.

Finally, post-processing is applied to the crawled articles at index time to enable named entity recognition and other value-added information.

### 2.3 Query Expansion

This component is a critical part of RCES. It is the engine behind the generation of expansion terms and their associated explanations for exploratory search. It allows to increase the recall when the search space is limited to address the information need. As previously mentioned in Section 1, RCES supports rapid cues for ease of exploration into various aspects of COVID-19 related topics. In the following, we present the explanation of taxonomies and the associated methods we apply on top of those taxonomies for the suggestion of candidate query expansion terms and their associated explanations:

- MeSH (Medical Subject Headings): This refers to the NLM<sup>8</sup> controlled vocabulary thesaurus used for indexing articles

<sup>3</sup><https://pubmed.ncbi.nlm.nih.gov/>

<sup>4</sup><https://www.medrxiv.org/>

<sup>5</sup><https://www.elsevier.com/>

<sup>6</sup><https://arxiv.org/>

<sup>7</sup><https://www.elastic.co/elasticsearch>

<sup>8</sup>National Library of Medicine in United States

for PubMed. For query expansion within our system, we perform a matching of the query terms against MeSH taxonomy elements to extract the top five relevant entity terms from MeSH. These MeSH entities are then used in a pseudo-relevance feedback [1] mechanism but with a slight modification that the four most frequent terms within each MeSH entity are used as pseudo-feedback documents. Finally, all frequent terms representing the top entities are aggregated and scored according to frequency while expanding a query. Also, these terms are then presented as explanations.

- MeSH Synonyms: MeSH is essentially a collection of medical subject headings, along with synonyms (called entity terms) within the taxonomy. We use those indicated synonyms against an exact matching entity with the query to retrieve synonyms as expanded terms.
- DBpedia: DBpedia is a large multi-domain ontology derived from Wikipedia. We exploit DBpedia structure through our previously proposed model [5], which essentially produces explainable vector embeddings using mappings between Wikipedia concept (article) and the corresponding Wikipedia category structure. As Figure 2 shows, the query is first mapped to a Wikipedia concept and then an explainable concept vector<sup>9</sup> is derived through its associativity with Wikipedia categories. Finally, using the vector's associativity scores, top-10 Wikipedia categories that match most strongly to the queried terms are retrieved as expansion terms and explained.
- DBpedia Synonyms: Like MeSH headings and sub-headings, DBpedia includes redirects of Wikipedia articles that we treat as synonyms for the input query. We use the matching of the query against a Wikipedia article to expand terms using Wikipedia redirects.
- Rule-based approach: This approach basically combines custom rules that can be specified by medical practitioners and healthcare researchers (who are users of the system) for their unique information needs. Here, the user uploads custom-defined taxonomy that RCES then uses to expand terms.

### 2.4 Retrieval Rank and Presentation

This component is responsible for ranking and presenting the documents retrieved in response to a query. Once users input a query, the most relevant documents are retrieved by standard *BM25* retrieval model, where  $k_1=1.2$  and  $b=0.75$ , with the search output including the following information (as shown in Figure 3):

- Expansion Terms' Explanation: associated explanations for the expansion terms corresponding to the selected query together with the scores assigned to them via our pseudo-relevance feedback mechanisms.
- Expansion Taxonomy: a choice of taxonomies to choose from amongst those mentioned previously in the Query Expansion component. The selected taxonomy is then used to select expansion terms, and it is possible to select more than one taxonomy at once.

<sup>9</sup>Explainable because each vector dimension is associated with a label derived from a Wikipedia category

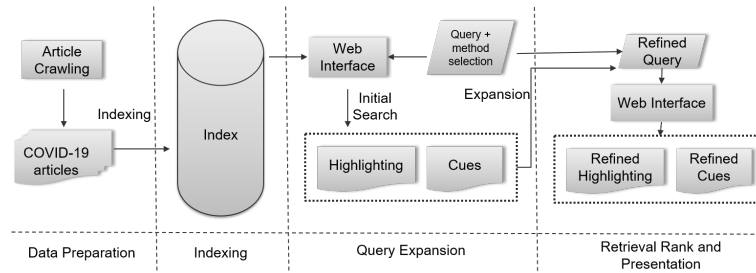


Figure 1: RCES system Architecture diagram

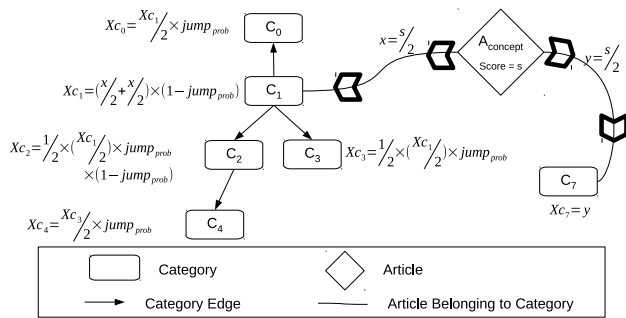


Figure 2: Mapping of concept (article) to Wikipedia categories.

- **Filters:** These are features of a particular granularity that can be applied over the retrieved set of articles. Since RCES searches over a corpus of medical articles<sup>10</sup> which fall into trials or reviews and can be related to dosage, drugs, duration, form, route, or strength and hence, the filters help in narrowing down the search. Other filters that users can choose are specific expansion methods, i.e. MeSH, MeSH synonyms, DBpedia, DBpedia synonyms and a rule-based approach. Furthermore, users can also filter search results based on the time period in advance of search to reduce retrieval output size by essentially limiting the publications during a certain period.
- **Facets:** these are essentially the extracted topical terms representing information extracted from the retrieved result set to bring a structure to a retrieved set of articles (and this is done by adopting biased PageRank model for extraction of topics from retrieved result set). Also, de-duplication is performed in the facet terminology (such as flu and influenza are treated as one) by exploiting semantic synonyms extracted through available taxonomies in the system (i.e., DBpedia, MeSH, and user-defined). Furthermore, using these facets, the users can re-rank the search results by selecting different facets, and this will then enable them to explore specific aspects of the input query.
- **Articles:** A list of articles retrieved in response to a query. Note that this list is a ranked list presented in the same manner as in traditional search engines. Moreover, when

the facets are used to re-rank the articles in our interface, we highlight those parts of article titles where these facets specifically appear.

### 3 DEMONSTRATION

A video demonstration of RCES is available at <https://youtu.be/67yCmwWAjN8> and below we present a brief overview of various features.

#### 3.1 Case Study: Unique Aspects of COVID-19

To illustrate the use of RCES, we present a few examples of concept discovery concerning COVID-19. We enter some unique query concepts obtained from the Health Service Executive in Ireland<sup>11</sup>.

The first of these is depicted in Figure 3, where the query “INFLUENZA VACCINE AND COVID-19” is entered. As per RCES’s support of a rule-based approach, the logical AND operator provides users with additional control over the exploratory search. As shown in Figure 3, the search output includes both the candidate expansion terms’ explanation, facets, and a list of retrieved articles. The expansion terms that led to a larger search space via query expansion component highlight useful relationships between influenza vaccine and COVID-19; particularly, the concepts such as vaccine derived from MeSH and animal virology derived from DBpedia explain the interplay of virus-related concepts. Furthermore, the facets are dynamically extracted from the top 200 retrieved result set. The various synonyms of the facets are combined to reduce any redundancy in facets, thereby depicting the same topic and combined through extracted synonyms from taxonomies. The extracted facets effectively highlight different aspects of the influenza vaccine’s effect concerning COVID-19, and in this case, these terms are “infection”, “disease”, “severity”, and “protein”. As shown in Figure 4, after selecting facet “disease” from the output in Figure 3, the top-ranked articles change accompanied with this selection.

#### 3.2 Expert Assessment

The following list shows some queries and information needs that were defined and used to design the system; we also mention the top two expansion terms and selected facets corresponding to each query. Note that the choice of queries reflects an aspect of COVID-19 necessary for healthcare practitioners to explore in order to have a deeper understanding of the virus, and a senior healthcare

<sup>10</sup>In this case medical articles specifically related to COVID-19 form our corpus

<sup>11</sup><https://www.hse.ie/>

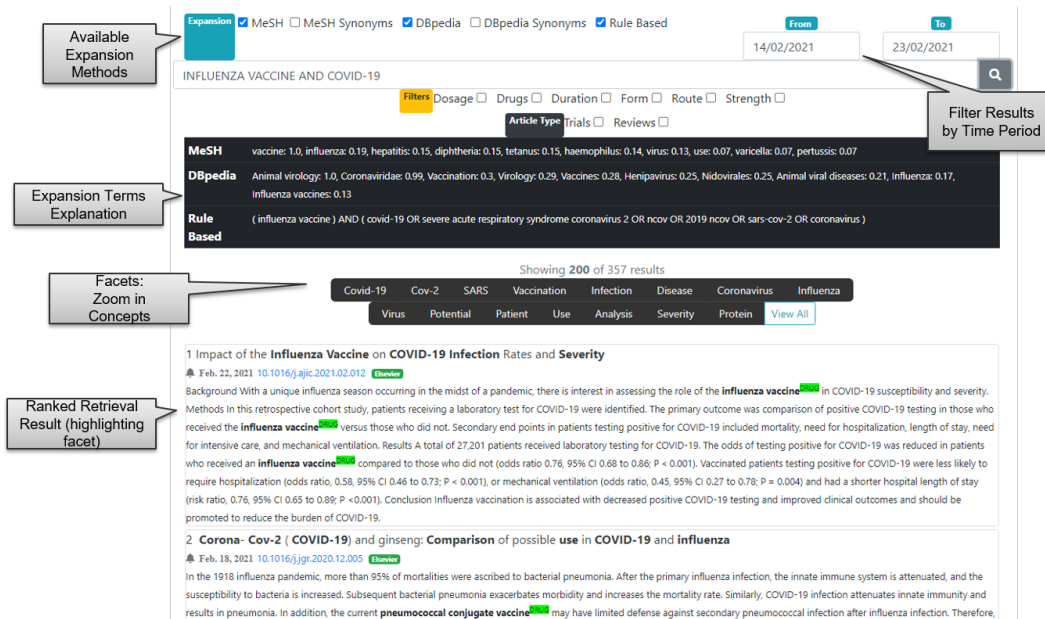


Figure 3: Example of presentation for query "INFLUENZA VACCINE AND COVID-19"

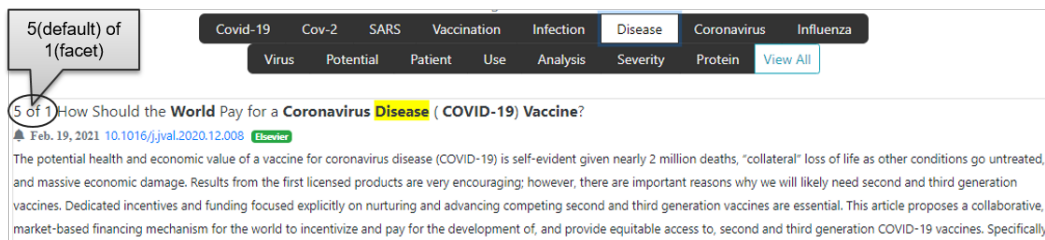


Figure 4: Example of re-ordering of retrieval results by selecting facet

practitioner served as an expert assessor for classifying results as relevant or irrelevant with and without the:

- SARS-CoV-2 + OPEN WATER SYSTEMS: Expansion terms are water and information from MeSH while systems and animal virology from DBpedia. Facets are Angiotensin, Water and System. Precision@5 with expansion methods is 0.8 as opposed to 0.6 without the methods.
- COVID-19 + TRANSMISSION + HEALTH WORKER TO PATIENT: Expansion terms are transmission and patient from MeSH while workforce and animal virology from DBpedia. Facets are Care, Syndrome and Acute. Precision@5 with expansion methods is 0.6 as opposed to 0.2 without the methods.
- SEQUELAE + COVID-19 + CRITICALLY ILL: Expansion terms are coronavirus and infection from MeSH while medical terminology and animal virology from DBpedia. Facets are Hydroxychloroquine, Medicine and Treatment. Precision@5 with expansion methods is 0.8 as opposed to 0.6 without the methods.

Each of the query, their associated expansion terms and the corresponding facets, as mentioned above, complement the exploration of search results with aspects related to query topic and COVID-19, and hence, allows access to deeper and varied aspects of retrieval.

## 4 CONCLUSION

In this paper, we present an exploratory semantic retrieval system. The up to date COVID-19 related articles are employed in the search engine to help researchers understand the pandemic better for medicinal needs. Various taxonomic resources and methods defined on top of those resources are applied to explore-exploit the search space. The pseudo-relevance feedback mechanisms over MeSH and the operations performed over embeddings derived from Wikipedia categories help improve the search effectiveness while also providing an enhanced user experience.

## ACKNOWLEDGMENTS

This publication has emanated from research conducted with the support of Science Foundation Ireland (SFI) under Grant Type COVID-RRC and ID: 20/COV/0218.

## REFERENCES

- [1] Efthimis N Efthimiadis. 1996. Query Expansion. *Annual review of information science and technology (ARIST)* 31 (1996), 121–87.
- [2] Sohrab Ferdowsi, Nikolay Borissov, Elaham Kashani, David Vicente Alvarez, Jenny Copara, Racha Gouareb, Poorya Amini, and Douglas Teodoro. 2021. Information retrieval in an infodemic: the case of COVID-19 publications. *bioRxiv* (2021).
- [3] Sean MacAvaney, Arman Cohan, and Nazli Goharian. 2020. SLEDGE-Z: A Zero-Shot Baseline for COVID-19 Literature Search. *arXiv preprint arXiv:2010.05987* (2020).
- [4] Gary Marchionini. 2006. Exploratory Search: From Finding to Understanding. *Commun. ACM* 49, 4 (April 2006), 41–46. <https://doi.org/10.1145/1121949.1121979>
- [5] M Atif Qureshi and Derek Greene. 2019. EVE: explainable vector based embedding technique using Wikipedia. *Journal of Intelligent Information Systems* 53, 1 (2019), 137–165.
- [6] Kirk Roberts, Tasmee Alam, Steven Bedrick, Dina Demner-Fushman, Kyle Lo, Ian Soboroff, Ellen Voorhees, Lucy Lu Wang, and William R Hersh. 2020. TREC-COVID: rationale and structure of an information retrieval shared task for COVID-19. *Journal of the American Medical Informatics Association* 27, 9 (2020), 1431–1436.
- [7] Lucy Lu Wang, Kyle Lo, Yoganand Chandrasekhar, Russell Reas, Jiangjiang Yang, Darrin Eide, Kathryn Funk, Rodney Kinney, Ziyang Liu, William Merrill, et al. 2020. Cord-19: The covid-19 open research dataset. *ArXiv* (2020).
- [8] Ryen W White, Bill Kules, Steven M Drucker, et al. 2006. Supporting exploratory search, introduction, special issue, communications of the ACM. *Commun. ACM* 49, 4 (2006), 36–39.
- [9] Edwin Zhang, Nikhil Gupta, Raphael Tang, Xiao Han, Ronak Pradeep, Kuang Lu, Yue Zhang, Rodrigo Nogueira, Kyunghyun Cho, Hui Fang, et al. 2020. Covidex: Neural ranking models and keyword search infrastructure for the covid-19 open research dataset. *arXiv preprint arXiv:2007.07846* (2020).
- [10] Bweijunl Zheng, Wei Zhang, and Xiaoyu Fu Boqin Feng. 2013. A survey of faceted search. *Journal of Web engineering* 12, 1&2 (2013), 041–064.