# Mental Health Resource Retrieval using Semantic Similarity and Knowledge Graphs

Varshini Prakash[1,3][0009-0000-2331-0552], Mohamad Ali Gharaat[1][0000-0003-3342-0549], Alex Lambe Foster[1,3][0000-0002-8424-9457], Dylan Merrick[4][0000-0001-7882-3478], Emilie Desnoyers[2][0009-0005-7211-4626], Jasmine M. Noble[2][0000-0001-5979-136X], Ken Porter[3][0000-0001-9802-3021], Andrew J. Greenshaw[2][0000-0002-9097-900X], and Osmar R. Zaiane[1][0000-0002-0060-5988]

[1] Department of Computing Science, University of Alberta
[2] Department of Psychiatry, University of Alberta
[3] Mood Disorders Society of Canada
[4] Department of Indigenous Studies, University of Saskatchewan

**Abstract.** As unmet mental health needs continue to rise, challenges in accessing and navigating resources and services remain a significant barrier to finding effective care. To address this challenge, MIRA, an AI-driven mental health virtual assistant, was developed [1], along with the MIRA library—a curated database of vetted mental health resources. This paper details the methods for designing an intelligent resource retrieval system that processes queries from a chatbot to identify relevant resources from our library. The proposed method combines term frequency and sentence embeddings to create vector representations of queries to perform similarity search for resource ranking. In addition, we incorporate query relaxation to leverage ontology-based hierarchical relationships. This ensures broader, yet contextually relevant search results. The proposed retrieval method outperforms traditional keyword-based and BERT-based methods in precision, as shown through our evaluation. The design and implementation of the ranking algorithm are detailed, including the construction of a knowledge graph and the integration of TF-IDF (Term Frequency-Inverse Document Frequency) based scoring. MIRA leverages the proposed retrieval method to rank resources effectively, recognizing user needs and connecting them to the most relevant support. By doing so, MIRA aims to bridge the gaps in traditional mental health care and improve accessibility.

**Keywords:** Retrieval · Ontology · Semantic Search.

## 1 Introduction

The need for mental health support continues to rise in Canada, with resource navigation being a significant barrier to accessing support [5]. The complex and fragmented mental health landscape presents significant challenges in connecting individuals to mental health resources through traditional methods. In a survey conducted by the Canadian Institute for Health Information (CIHI) in 2022, 86% of children and youth reported feeling overwhelmed, uncomfortable, or not knowing how to navigate the system as the primary barrier to seeking mental health support [5]. Even when individuals seek help, extended wait times make timely access to a professional difficult. Furthermore, among those who secure care, only 43% of Canadians reported receiving support from a professional in navigating mental health and substance use services in a survey conducted in 2025 [4].

This highlights a persistent gap in accessibility despite increased efforts. The adoption of virtual mental health care has shown promise, with 11% of Canadians reporting use of these services in 2022, higher than the Commonwealth Fund average of 7% [6]. However, virtual mental health care remains costly. For instance, 57% of young adults aged 18-24 cited cost as a major barrier to accessing private mental health services [5]. These findings underscore the ongoing need for innovative and affordable solutions to improve mental health resource navigation and accessibility across Canada.

While accessibility remains a challenge, chatbots can be a great solution for more effective resource navigation by providing instant, round-the-clock support. MIRA, an AI-driven mental health virtual assistant, was launched to revolutionize mental health resource navigation [2]. Originally developed for healthcare workers, MIRA has since been improved and expanded to serve a wider audience, including the general public, children and youth, French-speaking Canadians, veterans, and Indigenous communities [7].

MIRA does not attempt to provide a diagnosis or identify specific problems or conditions. Neither does it claim to suggest treatments to cure an ailment or reduce the effect of a disorder. Its only aim is to become aware of the information needs of an interlocutor and provide the relevant and trusted resources for the identified mental health topic, assisting the interlocutor in navigating the complex and disparate compilations of mental health material.

This paper focuses on the chatbot's core functionality: to search and retrieve the most relevant resources from an extensive database curated by experts. We use a semantic-based search to interpret the underlying context of user queries. A key component of our system is the knowledge graph, which structures domain-specific information into a formal ontology and allows us to map the intricate relationship of mental health entities. The knowledge graph was handcrafted by experts in psychiatry. The constructed ontology structure is used by the search algorithm to identify hierarchical and lateral associations within the domain. This ontology-driven search expands queries to retrieve resources that may differ lexically but remain semantically aligned to the user's queries. As a result, MIRA offers a more context-aware search experience for resource navigation.

## 2 Background

Information Retrieval (IR) is a key component in navigating large knowledge bases. While numerous algorithms have been developed to efficiently search and retrieve relevant information from large datasets, many of them are complex.

Traditional retrieval techniques can be categorized into term-based or semantic-based approaches [9]. The foundation of information retrieval has traditionally relied on simple algorithms such as ElasticSearch, edit distance, and keyword matching [8]. These techniques have been widely adopted across various applications, from library catalogues to web search engines, enabling efficient retrieval of relevant information.

A widely used IR technique is the Vector Space Model (VSM), which represents documents and queries as vectors in a multi-dimensional space [10]. Within this framework, TF-IDF weighting and cosine similarity have been effective tools to measure the relevance of documents to a given query [11]. Incorporating positional information such as term proximity and dependencies refined retrieval further [12, 13]. With the advent of deep learning, solutions that pre-trained on query and document embeddings gained popularity [14, 15]. Deep learning methods were broadly interaction-based [14] or representation-based models [15]. Interaction-based models, such as cross-attention models in neural retrieval, require both the document and query to be present simultaneously to compute relevance [14], unlike representation-based models, which encode documents independently and store them in an index for fast retrieval [15]. Interaction-based models perform computations dynamically for each new query, making them computationally expensive since each query requires complete pairwise comparisons with documents rather than leveraging precomputed representations [16].

Recent advancements in IR have seen the integration of knowledge graphs to enhance query understanding and relaxation [17]. Query relaxation is when requirements for a document to match the query are eased by removing or altering terms or tags. Knowledge graphs provide a structured representation of domain knowledge, enabling more sophisticated query processing [19]. Query relaxation techniques, such as those based on ontologies, allow for the expansion or modification of queries to improve recall while maintaining precision [18].

In the context of mental health resource navigation, the construction of domain-specific knowledge graphs has shown promise [23]. This process often involves mapping resources to hierarchical ontologies and collaborating with domain experts for manual curation. Such approaches can significantly improve the accuracy and relevance of search results in specialized fields [24].

State-of-the-art IR approaches increasingly leverage generative artificial intelligence (AI). Advanced chatbots such as ChatGPT [20], DeepSeek [21], and Perplexity [22] have become widely popular for their ability to have human-like conversations. These chatbots are Large Language Models (LLMs) that employ deep learning techniques, transformer architectures, natural language understanding (NLU), and contextual awareness to generate more accurate and relevant responses in real time. The use of deep learning-based retrieval models is not practical for our use case, given their requirement for large datasets for adequate performance, as well as their high computational cost and latency.

### 2.1 MIRA Resource Library

When a user interacts with MIRA, the resources recommended to users through the search mechanism are sourced from the MIRA Resource Library — a comprehensive database of over 1,500 resources catalogued by volunteers from the Mood Disorders Society of Canada (MDSC). The resources undergo a rigorous vetting process by a team of subject matter experts to ensure their quality and reliability. This evaluation is based on a hybrid assessment tool developed from validated instruments [2]. The review process, as outlined in Fig 1 mirrors academic peer review standards, with each resource being assessed by at least two reviewers [2]. If consensus is not reached, a third reviewer is consulted to make the final decision.


**Fig. 1.** This figure illustrates the resource gathering and review process. Research associates and volunteers categorize resources using appropriate tags, which are then reviewed by experts.

In addition to vetting the resources to be added to the MIRA library, subject matter experts also assign tags to each document from a set of predefined tags. These are later used by the retrieval system to improve the relevance of the search.

## 3 Methods

Our retrieval system combines keyword search, edit distance search, and sentence transformer embedding similarity search to identify relevant resources. Once matching keywords are found, the system retrieves all associated resources and ranks them accordingly. For instance, if a user tells the chatbot, *"I am looking for grief counselling resources in Vancouver"*, as user input, key entities like location and mental health concern are identified and extracted, such as *"Vancouver"* and *"Grief Counseling"*. The Dual Intent and Entity Transformer (DIET) architecture is used to capture the intent and entity from the user simultaneously [3]. The captured entities are then sent to the search engine within the MIRA resource library as a query. A basic keyword search in resource titles or descriptions would not suffice, therefore a semantic search engine that leverages an ontology of key mental health entities is employed to enhance query interpretation and resource retrieval.

### 3.1 Knowledge Graphs

To incorporate semantic understanding, the ontology is extended to include mental health entities and concepts, enabling the system to interpret hierarchical relationships and perform query relaxation to retrieve similar documents if an exact match is unavailable. Another advantage of this semantic search engine is its ability to recognize synonymous terms. For example, consider the phrases *"substance abuse recovery"* and *"addiction rehabilitation"*, which could both refer to strategies or treatments to overcome addiction, even though the terminology differs. Semantic search should recognize that the two phrases are linked in a mental health context. This process requires in-depth domain knowledge and expertise to effectively construct knowledge graphs and define relationships between entities. Understanding the nuances of specific fields, such as mental health or addiction recovery, is essential for accurately identifying how different terms, concepts, and treatments are interrelated. Therefore, we rely on the expertise of a multidisciplinary team of subject matter experts, including individuals with lived experience, Indigenous community members, clinicians, and psychiatrists, to construct these knowledge graphs.

We construct a domain-specific knowledge graph that organizes mental health topics into hierarchical categories, as depicted in Fig 2. The MIRA Resource Library is organized into seven categories of predefined tags, along with numerous additional tags that can be assigned to each resource during the submission process. Each tag in the database is mapped to an ontology entity, and parent-child relationships are established to enable query relaxation. For example, the tag "Cognitive Behavioral Therapy (CBT)" refers to a specific type of therapy, which falls under the broader category of "Psychological Treatment" in the ontology. This process requires input from experts to ensure that resource categorization aligns with real-world mental health concerns.

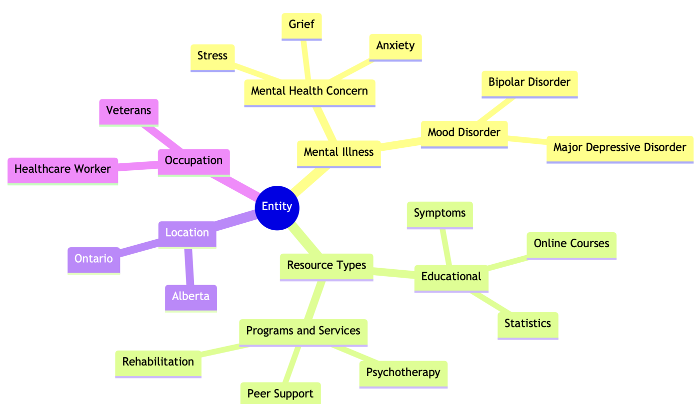


**Fig. 2.** This figure presents a simplified version of the MIRA ontology, showing the main superclasses and their child categories to provide an overview of its hierarchical structure.

### 3.2 Search Algorithm

The high-level architecture of the search mechanism and the algorithm for MIRA are illustrated in Fig. 3 and Algorithm 1, respectively. The architecture diagram in Fig. 3 outlines the key components and their interactions within the system, providing an overview of the search process.

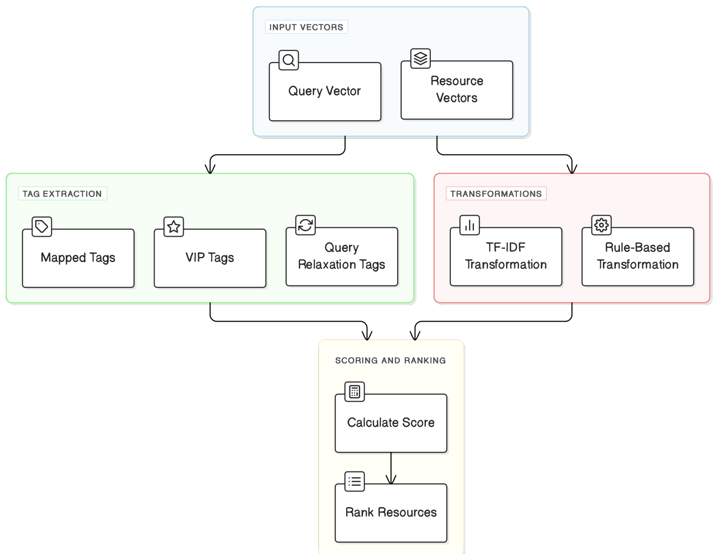


**Fig. 3.** Architecture of the MIRA search mechanism

The algorithm below presents a step-by-step breakdown of the MIRA search mechanism, detailing how data is processed to deliver accurate results. To increase the precision of our ranking algorithm, we transform the resources in our database into a multidimensional vector representation. This representation takes the form of a seven-dimensional vector, with each dimension corresponding to one of our predefined tag categories. Each vector component quantifies the resource's relevance with the specific categories.

We also transform the user queries into a similar representation. Through this approach, we aim to perform a more granular and context-aware comparison between resources and queries. Resources are ranked based on the inner product between their vector representation and the query vector. The similarity score for a resource within a tag category is determined using the following rules:

1. The similarity between a resource and a tag (ontology entity) is computed using a TF-IDF model trained on resource text data.
2. We identify hierarchical relationships between tags by computing the similarity between a tag and its parent entity is assigned a fixed hyperparameter, to ensure that parent-tagged resources receive slightly lower scores than directly matching ones.

**Algorithm 1** MIRA Ranking Function Algorithm

```
Input:
query_vector
library_resources - Set of resources from MIRA library with associated tags.
knowledge_graph - Ontology of predefined tags and hierarchical relationships.
Initialize:
Map query keywords to tags mapped_tags.
Retrieve parent tags for mapped_tags from the MIRA knowledge graph into
  query_relaxation_tags.
Identify "mental health" tags in mapped_tags as VIP_tags.
Resource Retrieval: Retrieve resources from library_resources with associa-
  tions to mapped_tags, query_relaxation_tags, and VIP_tags.
Scoring:
for each resource r in Resources do
    Compute TF-IDF for mapped_tags and query_relaxation_tags.
    Store TF-IDF scores in a 7-dimensional embedding.
    Adjust resource scores based on predefined ranking rules.
end for
Final Ranking:
Compute the query embedding from query_vector.
Calculate dot product similarity with each resource.
Sort by similarity.
return sorted list of resources.
```

3. Tags defined as synonyms in the ontology have a similarity score of 1.0, in order to treat the two as exact matches.
4. Resources from preferred organizations receive a small additional score to slightly boost their ranking while maintaining overall fairness.

To improve recall, we implement query relaxation based on ontology relationships. When an exact match for a query tag is unavailable, we retrieve resources linked to parent entities, applying a weighted relevance decay factor $\lambda$, typically set to 0.85, to lower their ranking compared to direct matches.

## 4 Results and Discussion

We employ two additional methods to transform both resources and queries into vector representations, evaluating their effectiveness using cosine similarity. This allows us to compare the performance between our model, TF-IDF and a Bidirectional Encoder Representations from Transformers (BERT) model. BERT-base is a 12-layer pre-trained language model that generates embeddings that capture the contextual meaning of words [25]. The TF-IDF and BERT-based retrieval methods encode queries and resources into vectors, measuring similarity with cosine distance.

To assess accuracy, we constructed 50 test cases, each containing an input query and three highly relevant resources identified by domain experts as the ground truth. A test is considered successful if at least one target resource appears in the top five ranked results. Fig. 4 presents our evaluation findings. The BERT-based model struggled to differentiate between similar resources accurately, while TF-IDF, being highly sensitive to exact keywords, performed the worst overall but had the fastest runtime. Given the limitations in computational efficiency and processing speed of purely deep learning models like BERT, we chose to combine different approaches for more balanced performance. The trade-off between using embeddings from advanced models like BERT, which offer strong language understanding capabilities, and the processing latency in computing these embeddings becomes especially significant in real-time applications. We explored integrating both TF-IDF and BERT, utilizing BERT selectively when TF-IDF is insufficient. This hybrid approach allows us to improve accuracy, maintain lower compute times and reduce overall latency compared to a fully BERT-based model.

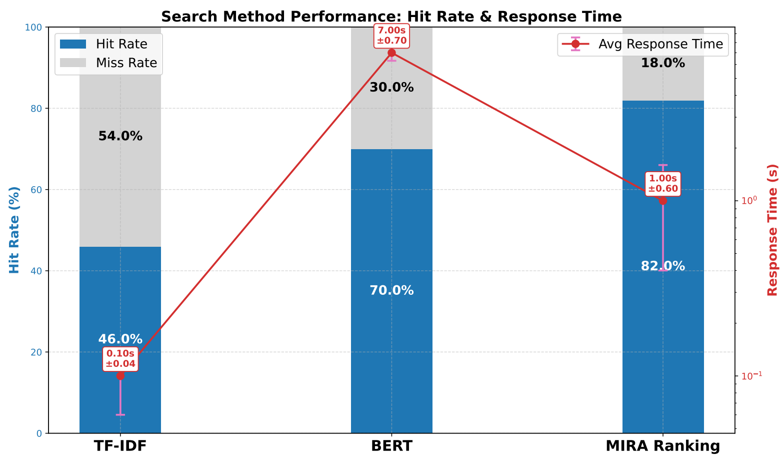


**Fig. 4.** Performance comparison of three search methods (TF-IDF, BERT, and MIRA Ranking) based on hit rate and response time. The stacked bar chart represents the hit rate (blue) and miss rate (grey) for each method. The response time is plotted on a logarithmic scale, with red markers representing the average response time (in seconds) and error bars indicating the standard deviation. MIRA Ranking achieves the highest hit rate (82%) with a moderate response time, while BERT has a higher hit rate than TF-IDF but a significantly longer response time.

## 5 Conclusion

Our ranking function improves resource navigation for mental health resources by integrating semantic search, domain-specific knowledge graphs, and query relaxation mechanisms. We use a vector embedding-based approach to capture query-resource relationships, to retrieve contextually aware results aligned to the user's needs.

A key advantage of our approach is its ability to interpret user queries more contextually, ensuring that resources are retrieved based on semantic relevance rather than just keyword matching. The incorporation of ontology-driven query relaxation further refines search results by broadening queries when exact matches are unavailable, thereby enhancing accessibility to critical mental health resources. While our search mechanism improves resource retrieval, our approach is limited in handling ambiguous or intricate queries and may not fully address the broad spectrum of user needs in mental health. Another challenge arises when the desired resource is missing in our curated library, as query relaxation may return seemingly irrelevant results, especially if the user is not informed about the query relaxation process. Additionally, with an increasing volume of resources, maintaining effective query relaxation and ranking becomes increasingly difficult.

For future directions, we aim to refine our system by incorporating human-annotated metadata tags to further enhance ranking accuracy and contextual understanding by building an LLM agent. We plan to fine-tune the agent on our curated domain-specific data using the retrieval-augmented generation (RAG) approach. This system would facilitate guided conversations with users, dynamically interpret their needs, answer queries with grounded knowledge from the curated database, and provide highly precise resource recommendations. By improving MIRA's conversational capability and retrieval accuracy, we plan to create a more intelligent, user-centric mental health support system that reduces barriers to care and improves accessibility for diverse populations.

## References


1. MIRA - The Mental Health Virtual Assistanthttps://www.mymira.ca
2. Noble, J., Zamani, A., Gharaat, M., Merrick, D., Maeda, N., Foster, A., Nikolaidis, I., Goud, R., Stroulia, E., Agyapong, V. & Others Developing, implementing, and evaluating an artificial intelligence–guided mental health resource navigation chatbot for health care workers and their families during and following the COVID-19 pandemic: protocol for a cross-sectional study. *JMIR Research Protocols.* **11**, e33717 (2022)
3. Zamani, A., Reeson, M., Marshall, T., Gharaat, M., Foster, A., Noble, J. & Zaiane, O. Intent and Entity Detection with Data Augmentation for a Mental Health Virtual Assistant Chatbot. *Proceedings Of The 23rd ACM International Conference On Intelligent Virtual Agents.* pp. 1-4 (2023)
4. Canadian Institute for Health Information. Navigation of Mental Health and Substance Use Services Indicator. https://www.cihi.ca/en/indicators/navigation-of-mental-health-and-substance-use-services-indicator, accessed (2025).
5. Canadian Institute for Health Information Common Challenges, Shared Priorities - Volume 4. (2022), https://www.cihi.ca/sites/default/files/document/common-challenges-shared-priorities-vol-4-report-en.pdf, Accessed: 2025
6. Canadian Institute for Health Information Canadians report increasing need for mental health care alongside barriers to access. (2023), https://www.cihi.ca/en/canadians-report-increasing-need-for-mental-health-care-alongside-barriers-to-access, Accessed: 2025

7. Prakash, V., Foster, A., Noble, J. & Zaiane, O. Integrating Conversational Pathways with a Chatbot Builder Platform. *International Conference On Information Integration And Web Intelligence*. pp. 184-189 (2024)
8. Gormley, C. & Tong, Z. Elasticsearch: the definitive guide: a distributed real-time search and analytics engine. (" O'Reilly Media, Inc.",2015)
9. Guo, T., Guo, J., Fan, Y., Lan, Y., Xu, J. & Cheng, X. A comparison between term-based and embedding-based methods for initial retrieval. *Information Retrieval: 24th China Conference, CCIR 2018, Guilin, China, September 27–29, 2018, Proceedings 24*. pp. 28-40 (2018)
10. Raghavan, V. & Wong, S. A critical analysis of vector space model for information retrieval. *Journal Of The American Society For Information Science*. **37**, 279-287 (1986)
11. Tata, S. & Patel, J. Estimating the selectivity of tf-idf based cosine similarity predicates. *ACM Sigmod Record*. **36**, 7-12 (2007)
12. Tao, T. & Zhai, C. An exploration of proximity measures in information retrieval. *Proceedings Of The 30th Annual International ACM SIGIR Conference On Research And Development In Information Retrieval*. pp. 295-302 (2007)
13. Huston, S. & Croft, W. A comparison of retrieval models using term dependencies. *Proceedings Of The 23rd ACM International Conference On Conference On Information And Knowledge Management*. pp. 111-120 (2014)
14. Lu, Z. & Li, H. A deep architecture for matching short texts. *Advances In Neural Information Processing Systems*. **26** (2013)
15. Palangi, H., Deng, L., Shen, Y., Gao, J., He, X., Chen, J., Song, X. & Ward, R. Deep sentence embedding using long short-term memory networks: Analysis and application to information retrieval. *IEEE/ACM Transactions On Audio, Speech, And Language Processing*. **24**, 694-707 (2016)
16. Qin, J., Zhang, W., Su, R., Liu, Z., Liu, W., Tang, R., He, X. & Yu, Y. Retrieval interaction machine for tabular data prediction. *Proceedings Of The 27th ACM SIGKDD Conference On Knowledge Discovery  Data Mining*. pp. 1379-1389 (2021)
17. Mai, G., Yan, B., Janowicz, K. & Zhu, R. Relaxing unanswerable geographic questions using a spatially explicit knowledge graph embedding model. *Geospatial Technologies For Local And Regional Development: Proceedings Of The 22nd AGILE Conference On Geographic Information Science 22*. pp. 21-39 (2020)
18. Elbassuoni, S., Ramanath, M. & Weikum, G. Query relaxation for entity-relationship search. *The Semanic Web: Research And Applications: 8th Extended Semantic Web Conference, ESWC 2011, Heraklion, Crete, Greece, May 29–June 2, 2011, Proceedings, Part II 8*. pp. 62-76 (2011)
19. Hogan, A., Blomqvist, E., Cochez, M., D'Amato, C., Melo, G., Gutierrez, C., Kirrane, S., Gayo, J., Navigli, R., Neumaier, S. & Others Knowledge graphs. *ACM Computing Surveys (Csur)*. **54**, 1-37 (2021)
20. ChatGPT by OpenAIhttps://chat.openai.com
21. DeepSeekhttps://deepseek.com
22. Perplexity AIhttps://www.perplexity.ai
23. Park, C., Lee, H. & Jeong, O. Leveraging medical knowledge graphs and large language models for enhanced mental disorder information extraction. *Future Internet*. **16**, 260 (2024)
24. Abu-Salih, B. Domain-specific knowledge graphs: A survey. *Journal Of Network And Computer Applications*. **185** pp. 103076 (2021)
25. Kenton, J. & Toutanova, L. Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings Of NaacL-HLT*. **1** (2019)