HEILBRONN UNIVERSITY
OF APPLIED SCIENCES

Smarter Literature Reviews: Modernizing Scientific Knowledge with Ai Based Automation

Priya Bhayani (225017)
Pooja Kuntinamadunagaraju (225270)

HEILBRONN UNIVERSITY
OF APPLIED SCIENCES

Agenda

1. Introduction
2. Program Workflow
3. Evaluation & Results
4. Conclusion

Overview

This project develops an AI-powered pipeline to automate essential steps in scientific literature reviews. Unlike traditional approaches based on TF-IDF, it focuses on enhancing the critical filtering and clustering phase through semantic embeddings.

The goal is to improve the relevance and coherence when grouping research papers.

Specifically, we aim to:

- Replace outdated keyword-based filtering with deep learning-driven semantic understanding.
- Enable meaningful clustering of related research works based on context, not just word overlap.
- Develop a scalable toolchain that can be adapted to a variety of research fields.

HEILBRONN UNIVERSITY
OF APPLIED SCIENCES

Introduction

What is the Current Problem

- Information Overload: The sheer volume of new research papers is overwhelming, making it hard to stay current.

- Outdated Methods: Traditional literature reviews are slow and manual, taking too much time.

- Semantic Gap: Old methods like TF-IDF don't understand the true meaning of words, leading to:

    - Poorly organized topics.
    - Missed connections between related papers.
    - A lot of extra manual work for researchers.

HEILBRONN UNIVERSITY OF APPLIED SCIENCES

Introduction

Our Solution: Smart Automation with Embeddings

- Core Idea: We replace old methods with advanced AI models called "transformer-based sentence encoders.“

Key Embedding Models:

- E5-base: Great at understanding the full meaning and context of sentences.
- All-MiniLM-L6-v2: Powerful and efficient, providing deep understanding in a compact form.

Benefits:

- Generates "semantic embeddings" that truly capture meaning.
- Allows for precise filtering and intelligent grouping of papers.

Program Workflow

HEILBRONN UNIVERSITY
OF APPLIED SCIENCES

Workflow: An End-to-End Automated Pipeline :

Step 1. Data Collection — Findpapers (PubMed, Scopus, arXiv)

Step 2. Converting data from JSON to Excel

Step 3. Clean and Preprocess Abstract Texts

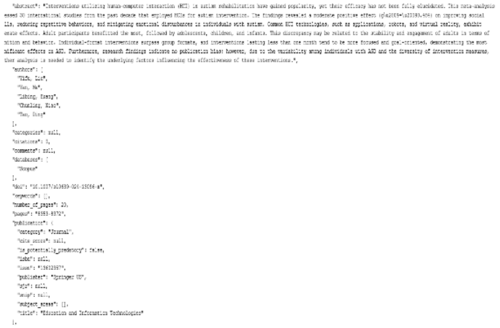
Step 4. Generate Sentence Embeddings (E5 & All-MiniLM-L6-v2)

Step 5. Filter via Cosine Similarity (≥ 0.80)

Program Workflow

Step 1. Data Collection (Findpapers Library):

- **Findpapers** is a Python library that allows you to automatically search academic papers from multiple databases.
- enter a search query (e.g. "VR AND Mental Health"), specify the date range, and select databases.
- It fetches paper details like **title, authors, publication date, abstract, DOI, journal name** and The results are saved in a **JSON file.**



Program Workflow

Step 2. JSON to Excel Conversion & Data Structuring

- Parses JSON to extract title, publication date, abstract, authors, database, publisher, DOI, citation count.
- Converts this data into a **Pandas DataFrame** (a tabular format) .
- Identifies and removes near-duplicate entries based on titles.
- Saves it as an **Excel file** for easy viewing and next processing steps.


```
  "abstract": "Interventions utilizing human-computer interaction (HCI) in autism rehabilitation have gained popularity, yet their efficacy has not been fully elucidated. This meta-analysis
essed 30 international studies from the past decade that employed HCIs for autism intervention. The findings revealed a moderate positive effect (g\u2009=\u20090.484) on improving social
lls, reducing repetitive behaviors, and mitigating emotional disturbances in individuals with autism. Common HCI technologies, such as applications, robots, and virtual reality, exhibit
erate effects. Adult participants benefitted the most, followed by adolescents, children, and infants. This discrepancy may be related to the stability and engagement of adults in terms of
nition and behavior. Individual-format interventions surpass group formats, and interventions lasting less than one month tend to be more focused and goal-oriented, demonstrating the most
nificant effects on ASD. Furthermore, research findings indicate no publication bias; however, due to the variability among individuals with ASD and the diversity of intervention measures,
ther analysis is needed to identify the underlying factors influencing the effectiveness of these interventions.",
  "authors": [
    "Yifu, Liu",
    "Yan, Ma",
    "Libing, Huang",
    "Chunling, Xiao",
    "Tao, Ding"
  ],
  "categories": null,
  "citations": 0,
  "comments": null,
  "databases": [
    "Scopus"
  ],
  "doi": "10.1007/s10639-024-13096-x",
  "keywords": [],
  "number_of_pages": 20,
  "pages": "8353-8372",
  "publication": {
    "category": "Journal",
    "cite_score": null,
    "is_potentially_predatory": false,
    "isbn": null,
    "issn": "13602357",
    "publisher": "Springer US",
    "sjr": null,
    "snip": null,
    "subject_areas": [],
    "title": "Education and Information Technologies"
  },
```


	A	B	C	D	E	F	G	H	I	J
1	Title	Year	Abstract	Authors	Databases	Publisher	Journal	Keywords	DOI	Citations
2	Current Tr	2025-05-01		Galieri G.	Scopus		Journal of Clinical M		10.3390/jc	0
3	The effect	2025-05-0	Interventi	Yifu, Liu/ Y	Scopus	Springer U	Education and Inforr		10.1007/s	0
4	Examining	2025-05-01		Kuleli D.	Scopus	Elsevier M	Journal of Behaviora		10.1016/j.	0
5	The influe	2025-05-01		Zhao M.	Scopus	Elsevier B.	Acta Psychologica		10.1016/j.	0
6	Examining	2025-05-01		Belhan Z.	Scopus	Tissue Via	Journal of Tissue Via		10.1016/j.	0
7	Underlyin	2025-05-01		Schevenee	Scopus	Elsevier In	Behavior Therapy		10.1016/j.	1
8	The Use o	2025-05-01		Laker C.	Scopus	MDPI AG	Nursing Reports		10.3390/n	0
9	The effect	2025-05-01		Ke Z.	Scopus	Elsevier In	General Hospital Psy		10.1016/j.	2
10	Retail valu	2025-05-01		Rumokoy	Scopus	Elsevier Lt	Journal of Retailing a		10.1016/j.	4
11	Personaliz	2025-05-01		Namgung	Scopus		JAMA Network Open		10.1001/ja	0
12	Supportin	2025-05-0	PURPOSE (	Hannah E	PubMed		Current ps	N Anxiety disorders/		0
13	A new tre	2025-05-0	INTRODUC	Nicholas D	PubMed	Frontiers I	Frontiers i	N emotions/ N treatr		0
14	Pathway t	2025-05-01		van 't Wou	Scopus		Brain Stimulation		10.1016/j.	0
15	Feasibility	2025-05-01		Kang B.R.	Scopus	SAGE Publ	Journal of Alzheimer		10.1177/1	0
16	Using Virtu	2025-05-01		Garcia-Gu	Scopus		Journal of Clinical Ps		10.1002/jc	2
17	Effects of	2025-05-01		Şimşekli D	Scopus	Elsevier In	Heart and Lung		10.1016/j.	0
18	Use of Tec	2025-05-01		Ferreira A	Scopus		Creative Nursing		10.1177/1	0
19	Assessing	2025-05-01		Sepanloo	Scopus		Sensors		10.3390/s	0
20	Research (	2025-05-01		Zhang B.	Scopus		Computer Animatior		10.1002/c	0

HEILBRONN UNIVERSITY OF APPLIED SCIENCES

Program Workflow

Step 3. Data Cleaning and Preprocessing:

- Raw abstracts often have noise: HTML tags, special characters, non-English text, empty values.
- Handles missing values, removes HTML tags and placeholders.
- Detects and discards non-English texts.
- Strips predefined patterns (e.g., "Objective:"), eliminates unwanted characters.
- Converts all text to lowercase for uniformity.

Program Workflow

Step 4. Generate Sentence Embeddings (E5 & All-MiniLM-L6-v2)

- Utilizes pre-trained language models from the sentence-transformers library.
- Converts cleaned abstract texts into dense numerical vectors (embeddings).
- **Two models used:**
  - E5-base → accurate, high precision
  - All-MiniLM-L6-v2 → fast and lightweight
- These embeddings are added as new columns in the DataFrame.

Program Workflow

HEILBRONN UNIVERSITY
OF APPLIED SCIENCES

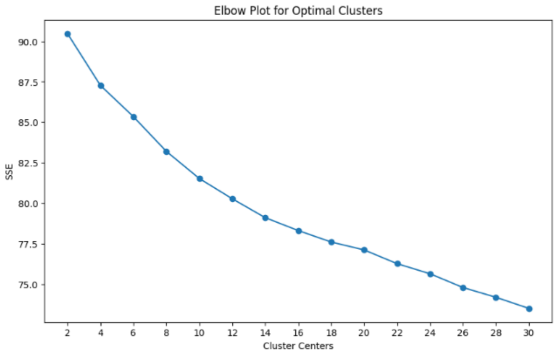
Step 5. Cosine Similarity Calculation:

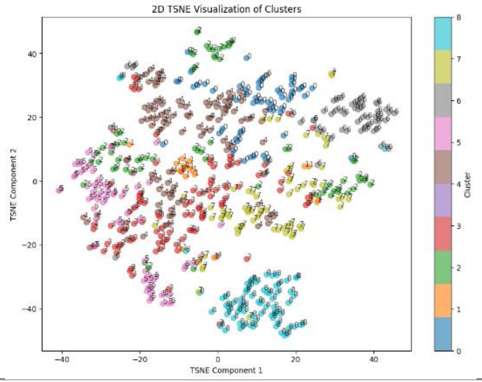
- Calculates semantic closeness between document embeddings and a user-defined retrieval query.
- A custom query like "VR AND Mental Health" is converted into an embedding.
- Calculate **cosine similarity** between the query embedding and each paper's embedding.
- Applies a stringent similarity threshold (e.g., ≥0.80) to retain only the most relevant papers.

Program Workflow

Step 5. Clustering (MiniBatchKMeans & Elbow Method, t-SNE Visualization):

- Employs MiniBatchKMeans for thematic grouping.
- Uses the Elbow Method to determine the optimal number of clusters (e.g., k=9 for E5).
- Assigns cluster labels to each paper.
- Visualizes high-dimensional embeddings in 2D using t-SNE (with PCA pre-reduction) for intuitive thematic inspection.





Evaluation & Results

HEILBRONN UNIVERSITY OF APPLIED SCIENCES

Search Query Definition and Data Collection Summary

To validate the utility of this automated system, we implemented the following criteria:

- **Query :** '([Virtual Reality] OR [Augmented Reality] OR [Extended Reality] OR [Mixed Reality]) AND ([Therapy] OR [Treatment] OR [Intervention] OR [Rehabilitation] OR [Outcome]) AND ([Mental] OR [Psychological] OR [Emotional] OR [Addiction] OR [Alcohol] OR [Exposure] OR [Behavior] OR [Disorder]) AND NOT [Physical]'

- **Period:** 2013 to 2023.
- **Selected databases**: PubMed, arXiv, and Scopus

    1. PubMed : 1000
    2. arXiv : 80
    3. Scopus : 1000

- The system successfully collected 2080 records.

Evaluation & Results

HEILBRONN UNIVERSITY OF APPLIED SCIENCES

1. Cosine Similarity-Based Relevance Filtering

- After converting abstracts into embeddings using E5-base and All-MiniLM-L6-v2:
- Applied a cosine similarity threshold of 0.80.
- From an initial ~2000 papers, the pipeline filtered down to 590 highly relevant papers.
- Demonstrated strong ability to isolate papers semantically aligned with the query (e.g., *VR and mental health*).

```
Saved 613 papers with similarity ≥ 0.8 to:
  /content/drive/MyDrive/DL/most_similar_paperse5.xlsx
                                              Title        Year Databases  \
0   Adverse Effects of Virtual and Augmented Reali...  2023-05-05    PubMed
1   Physiological Factors Based Depression Assessm...  2025-01-01    Scopus
2   Virtual Reality Interventions and Chronic Pain...  2025-01-01    Scopus
3   Virtual Reality as a Supplement to Traditional...  2025-01-01    Scopus
4   The use of virtual reality and augmented reali...  2022-12-14    PubMed
5   Virtual reality and artificial intelligence: t...  2025-03-01    Scopus
6   Virtual reality in the diagnostic and therapy ...  2022-10-30    PubMed
7   Efficacy of virtual reality-based training pro...  2024-12-01    Scopus
8   Extended Reality for Mental Health Evaluation:...  2024-07-24    PubMed
9   The Efficacy of Virtual Reality on the Rehabil...  2025-04-25    PubMed
10   Virtual Reality Interventions for Mental Health.  2023-01-01    PubMed
11  Reversing the Ruin: Rehabilitation, Recovery, ...  2022-10-01    PubMed
12  Advances in the use of virtual reality to trea...  2024-08-01    Scopus
13  Virtual reality interventions for the treatmen...  2023-02-25    PubMed
14  Innovative Approaches for the Mental Developme...  2025-01-01    Scopus
15  Mental health providers are inexperienced but ...  2024-07-03    PubMed
16  Immersive virtual reality in the treatment of ...  2024-03-01    PubMed
17  Virtual Reality Mental Health Interventions in...  2024-06-17    PubMed
18  Review on the Role of Virtual Reality in Reduc...  2024-07-08     arXiv
19  Virtual, mixed, and augmented realities: A com...  2024-07-08    PubMed
20  Immersive Technologies for Depression Care: Sc...  2024-04-25    PubMed
21  The Use of Virtual Reality Interventions to Pr...  2023-07-06    PubMed
22  Barriers to adopting therapeutic virtual reali...  2025-03-18    PubMed
23  Digital travel using virtual reality in inpati...  2025-04-28    PubMed
24  Use of Virtual Reality and Augmented Reality T...  2024-07-08    PubMed
25  [The application of virtual reality in the tre...  2022-09-02    PubMed
26  Mixed Reality Technology to Deliver Psychologi...  2023-07-26    PubMed
27  The Use of Virtual Reality in the Rehabilitati...  2023-09-01    PubMed
28  Effectiveness and safety of virtual reality re...  2023-09-14    PubMed
29  'Being there together for health': A Systemati...  2024-12-06     arXiv

    Similarity
0     0.854708
1     0.853047
2     0.852649
3     0.851417
4     0.846814
5     0.845775
6     0.845193
7     0.844850
8     0.844224
9     0.843551
10    0.843280
11    0.843059
12    0.842136
13    0.842011
14    0.842006
```

Evaluation & Results

HEILBRONN UNIVERSITY OF APPLIED SCIENCES

2. Clustering Analysis:

- **Methodology:** Cluster quality was assessed through the Elbow Method for optimal cluster count, analysis of cluster size distributions, manual purity labelling, and 2D semantic visualization via t-SNE.
- **Optimal Cluster Count:** The Elbow Method, plotting Sum of Squared Errors (SSE) against k (number of clusters), indicated "soft elbows" around k=9 for E5 and k=10 for All-Mini-LM6.
- **Cluster Purity:** Manual labelling revealed a mean purity of 70% for E5 clusters and 59% for All-Mini-LM6, with some clusters showing higher (e.g., E5 Cluster 7 at 83%) and lower purity.
- **2D Semantic Visualization (t-SNE):** This visualization, following PCA for dimensionality reduction, effectively illustrated well-isolated thematic clusters (e.g., autism-spectrum therapy) and areas of overlap reflecting semantic relatedness (e.g., chronic pain vs. palliative care).

Evaluation & Results

HEILBRONN UNIVERSITY
OF APPLIED SCIENCES

3. **Retrieval Effectiveness:**

- **Metrics:** Precision@K, Recall@K, Mean Average Precision (MAP), Mean Reciprocal Rank (MRR), and nDCG@10 were computed for E5, All-MiniLM, and TF-IDF models using manually labeled relevance judgments.
- **Key Findings:**

Model	P@10	P@20	P@30	R@20	R@30	MAP	MRR	nDCG@10
E5	0.900	0.850	0.767	0.739	1.000	1.000	1.000	0.927
All-miniLM	0.800	0.750	0.633	0.789	1.000	1.000	1.000	0.852
TF–IDF	0.700	0.650	0.567	0.765	1.000	1.000	0.333	0.700

Evaluation & Results

HEILBRONN UNIVERSITY OF APPLIED SCIENCES

4. Clustering Validity Indices:

- **Metrics:** Three standard validity metrics were applied: Silhouette Score (higher is better), Calinski-Harabasz Index (higher is better), and Davies-Bouldin Index (lower is better).
- **Findings:**
    - All-Mini-LM6 achieved the highest Calinski-Harabasz score and the lowest Davies-Bouldin index, signifying the most compact and well-separated clusters.
    - E5 followed closely with moderate cohesion.
    - TF-IDF consistently yielded the weakest cluster structure across all validity metrics.

Model	Silhouette	Calinski–Harabasz	Davies–Bouldin
E5	0.0258	10.6174	4.4173
All-miniLM	0.0221	18.0491	3.4929
TF–IDF	0.0105	5.8812	6.8061

Evaluation & Results

HEILBRONN UNIVERSITY
OF APPLIED SCIENCES

5. **Comparison with Automated AI Tools:**

- **Approach:** The pipeline's performance was benchmarked against external AI tools (Elicit, Litmaps, Semantic Scholar) using the same specialized Boolean-style query.

- **Results:** Despite indexing a wider range of sources, these tools returned only a handful of truly relevant results for our niche VR-mental health query.
- In stark contrast, our embedding-based pipeline's focused top 30 retrieval demonstrated substantially higher precision, significantly reducing the manual triage burden for researchers.

Conclusion

- We've successfully created an automated pipeline that transforms how literature reviews are done.
- The Power of Semantic Embeddings: By using E5-base and All-MiniLM-L6-v2, we've moved beyond the limitations of older methods like TF-IDF, truly understanding the meaning behind the text.
- **Through this approach, we achieved:**
  - High retrieval precision using cosine similarity filtering.
  - Accurate thematic clustering with MiniBatchKMeans, guided by the Elbow Method.
  - Clear 2D visualizations of high-dimensional data using PCA and t-SNE.
  - A systematic, structured review process aligned with the PRISMA framework.
- **The results showed that:**
  - E5-base delivered superior retrieval accuracy and early precision.
  - All-MiniLM-L6-v2 produced more cohesive and well-defined clusters.
  - Both models significantly outperformed traditional TF-IDF and common AI tools in precision, clustering quality, and efficiency.

References

- **Key References:**

- AI-Paper Miner — Open-source Project by Tamago55 (GitHub)
- Research Project by Alexander Bazhanov on automated literature review pipelines
- Bachelor Thesis by Rajakaruna on embedding-based literature review automation
- Sentence-Transformers Library — Pre-trained embedding models available via HuggingFace
- MTEB Benchmark Leaderboard — Evaluating embedding Model Performance Arcos Tasks
- PRISMA Guidelines — Standard Framework for Systematik literature Reviews and meta-Analyses
- APIs and Data Sources: Semantic Scholar API

THANK YOU!