

Notice – Article in Press

This article has been accepted for publication in *Rudarsko-geološko-naftni zbornik* and is currently available as an **Article in Press**. This version is an uncorrected proof and has not yet been copyedited or formatted according to the journal's editorial standards. As a result, the content may be subject to change prior to final publication.

Although Articles in Press do not yet have complete bibliographic details, they may be cited using the year of online publication, volume, issue, and DOI, which will remain unchanged in the final version of record.

Correct citation format:

Adjiski, V., Šancer, J., Schwendner, P., Crnogorac, L., Zeqiri, K., Zubíček, V., & Moellerherm, S. 2027. A hybrid knowledge system for mine rehabilitation decision support using relational filtering and semantic retrieval in a community-curated case repository. *Rudarsko-geološko-naftni zbornik* 42(1), Article in Press.

<https://doi.org/10.17794/rgn.2027.1.1>

Uncorrected proof

A hybrid knowledge system for mine rehabilitation decision support using relational filtering and semantic retrieval in a community-curated case repository

Rudarsko-geološko-naftni zbornik
(The Mining-Geology-Petroleum Engineering Bulletin)
DOI: 10.17794/rgn.2027.1.1

Original scientific paper



Vancho Adjiski^{1*}, Jindřich Šancer², Peter Schwendner³,
, Luka Crnogorac⁴, Kemajl Zeqiri⁵, Václav Zubiček²,
Stefan Moellerherm⁶

¹ Faculty of Natural and Technical Sciences, Goce Delcev University, Stip, North Macedonia.

² Department of Mining engineering and Safety, VSB – TUO, Ostrava Poruba, Czech Republic.

³ Zurich University of Applied Sciences, School of Management and Law, Switzerland.

⁴ University of Belgrade - Faculty of Mining and Geology, Belgrade, Serbia.

⁵ Faculty of Geosciences, University "Isa Boletini" Mitrovice – Kosovo.

⁶ Research Center of Post-Mining, Technical Hochschule Georg Agricola University of Applied Science 44787 Bochum Germany.

Abstract

Mine closure and post-mining repurposing require tools that connect practitioners with relevant rehabilitation examples across fragmented case evidence. This study transforms the Recovery of Mining District Network (REMINDNET) Work Group 3 (WG3) best-practices spreadsheet into a hybrid knowledge system combining structured relational filtering with lexical and semantic retrieval. From 149 submissions, the workflow produced a curated dataset of 130 unique cases from 21 countries, with validated WGS84 coordinates, standardised metadata, a controlled 12-category multi-label vocabulary, and reproducible 'case_card' and 'retrieval_text' representations. The system was evaluated using 25 natural-language benchmark queries and a relevance definition based on category overlap within structured constraints. The dataset is dominated by socially oriented and relatively low-impact reuse pathways, especially Community & culture, Non-intensive recreation, Conversion & ecosystem services, and Education & research. Co-occurrence and subgroup analyses show that mine type, operational status, ore type, and geography shape thematic case profiles, supporting their use as decision-support filters. In retrieval evaluation, both term frequency-inverse document frequency (TF-IDF) and embedding-based ranking improved over a filters-only random baseline, confirming that ranked search adds value after structured filtering. Filtered embeddings achieved the highest macro-average early-ranking scores, including Precision at 5 ($P@5 = 0.768$) and normalized Discounted Cumulative Gain at 10 ($nDCG@10 = 0.810$), but paired query-level tests did not show a statistically reliable advantage over filtered TF-IDF at the current benchmark scale. The strongest supported gain was the benefit of structured filtering for the embedding workflow. Overall, a relational-plus-semantic architecture can make community-curated knowledge on mine rehabilitation easier to search, compare, and interpret.

Keywords: mine rehabilitation, post-mining land use, decision-support system, semantic retrieval, community-curated database

1. Introduction

Mine closure, rehabilitation, and post-mining repurposing are now widely treated as life-cycle decisions rather than final compliance steps, because post-mining outcomes influence environmental conditions, community wellbeing, and long-term liabilities. Recent research now views mining as a temporary land use and post-mining transitions as an integral part of planning and governance (Keenan & Holcombe, 2021; Measham et al., 2024). Yet decisions about post-mining land use still depend on multiple criteria and site-specific context, including ecological conditions, stakeholder expectations, and changing regulatory and performance standards (Morrison-Saunders et al., 2016; Pagouni et al., 2024). In practice, the challenge is to identify rehabilitation approaches that can be adapted across sites with similar constraints and intended land-use outcomes (Measham et al., 2024).

In mining-engineering practice, closure and rehabilitation planning must convert broad post-mining land-use objectives into options that are technically feasible, environmentally stable, legally compliant, economically realistic, and socially acceptable. Comparable case evidence can support this process by showing how similar sites have addressed land-use conversion, risk reduction, environmental recovery, public access, cultural reuse, or socio-economic transition.

Cross-site learning is limited not so much by a lack of experience, but by evidence that is fragmented and not equally accessible. Lessons from mine closure are spread across different disciplines, jurisdictions, and project-specific documents, and even widely cited "success" cases are often difficult to generalise because social and governance factors are important but not reported consistently (Bainton & Holcombe, 2018; Measham et al., 2024). Where land-use outcomes are negotiated among stakeholders, the decision process itself becomes part of the knowledge that needs to be preserved and accessed (Crofts & Phelan, 2025; Worden et al., 2024). This creates a need for case-oriented repositories

*Corresponding author: Vancho Adjiski

e-mail address: vanco.adjiski@ugd.edu.mk

Received: 15 May 2026. Accepted: 26 June 2026.

that preserve site context as explicit metadata rather than leaving applicability conditions buried in narrative text (Pagouni et al., 2024). In practice, many environmental and engineering repositories begin as community-curated spreadsheets. This helps broaden coverage, but it also leads to uneven data completeness, different data structures, and inconsistent descriptions of similar practices (Goodchild, 2007; Senaratne et al., 2017). Controlled vocabularies improve comparability by making meaning more consistent, supporting reliable tagging, and helping records work together across systems, but labels alone rarely capture the full complexity of rehabilitation narratives or mixed and evolving post-mining land-use options (Ma et al., 2010; Pagouni et al., 2024; Y. Zhang et al., 2015). Research on geospatial and environmental decision-support systems shows how repositories can support the exploration of options and trade-offs through structured data, models, and interactive interfaces (Keenan & Jankowski, 2019; McIntosh et al., 2011). Web-based geographic information systems (Web GIS) are particularly relevant because map-based exploration and attribute filtering help users narrow relevant cases by geography and structured site attributes, while mining-oriented GIS applications demonstrate the value of structured geodatabases for reclamation planning and abandoned-mine management (Kim et al., 2012; Maina et al., 2014; Stanković et al., 2016; Suh et al., 2017). However, when the main evidence is contained in narrative text, discovery remains limited by heterogeneous metadata and simple keyword-only search (Athanasios et al., 2009; Lacasta et al., 2007). Faceted browsing, implemented as relational filtering over controlled attributes, supports exploratory search and step-by-step narrowing of candidate cases, but it cannot rank similar cases when exact matches are absent or when users seek comparable rather than identical examples (Vandic et al., 2024; Villa et al., 2008; Yee et al., 2003; Zhang & Zhang, 2010). TF-IDF remains a transparent lexical baseline, while dense sentence embeddings can support similarity search when the same idea is expressed using different terminology (Devlin et al., 2018; Reimers & Gurevych, 2019). Previous retrieval studies show that dense retrieval can outperform lexical methods in some settings, but also that sparse and dense signals are often complementary and that ranking robustness can benefit from combining retrieval signals (Arabzadeh et al., 2021; Cormack et al., 2009; Karpukhin et al., 2020; Luan et al., 2021). Evaluation research further shows why early ranking quality matters in search-based decision support, because users often focus on the first page of results and metrics such as nDCG emphasise correct ordering near the top of the ranked list (Järvelin & Kekäläinen, 2002; Voorhees & Harman, 2000). Even with this foundation, peer-reviewed evidence remains limited for domain-specific mine-rehabilitation repositories that combine structured filtering with ranked semantic matching and evaluate retrieval performance in realistic filter-then-search workflows.

From a mine-rehabilitation perspective, the value of such a system is not that it automatically selects the optimal post-mining land use, but that it improves access to comparable cases during early option screening. Practitioners can use structured filters to reflect known site constraints, such as mine type, status, ore type, or geography, and then use ranked retrieval to identify cases with similar rehabilitation goals. In this way, the system supports evidence-informed comparison, stakeholder communication, and the preparation of rehabilitation alternatives before detailed site-specific engineering, environmental, regulatory, and socio-economic assessment.

This study addresses that gap by converting the REMINDNET WG3 spreadsheet-based collection into a hybrid knowledge system that combines relational filtering with lexical and semantic ranking. Specifically, the study has three objectives: (1) to transform the spreadsheet-based repository into a relationally structured case dataset with a controlled multi-label category vocabulary, (2) to benchmark TF-IDF and embedding-based semantic ranking under identical relational filters, and (3) to examine dataset structure relevant to decision support, including category prevalence, co-occurrence, subgroup patterns, and spatial coverage. The goal is to assess whether semantic ranking helps users find comparable rehabilitation cases more effectively within a filter-then-search workflow, while also identifying coverage limitations that should guide interface design and the interpretation of results.

2. Materials and Methods

2.1. Data source and study dataset

Data were collected through the REMINDNET WG3 best-practices survey, a structured Google Form used to document rehabilitation and post-mining reuse cases across Europe. The survey recorded country, site name, mine type, operational status, ore deposit type, coordinates, current and former land use, a best-practice description, and optional media or links. The raw export contained 149 submissions.

After standardisation and screening for data completeness and site-level uniqueness, 130 cases were retained for analysis. Duplicate entries for the same site were removed, records without enough information for classification were excluded, and inconsistent coordinate formats were standardised to WGS84 decimal degrees. A reproducible 'site_key' was created from country, site name, and coordinates to support deduplication and consistent joins across processing steps. The final dataset covers 21 countries and includes closed, active, abandoned, and other sites from both underground and surface operations.

To support reproducibility and hybrid retrieval, the cleaned dataset was organised into two relational tables:

- The case table ('cases.csv') contains 130 rows and stores standardised metadata fields, numeric latitude and longitude (decimal degrees, WGS84), the best-practices narrative, a constructed 'case_card' string used to present retrieved results, and binary indicators for each category (Table 1). The cleaned fields needed to generate 'retrieval_text' are stored in this table, while 'retrieval_text' itself is generated at indexing time for the ranking experiments.

- The case-category table ('case_categories.csv') stores the many-to-many relationship between cases and controlled best-practice categories, with one row for each case-category assignment. This structure supports analyses of category prevalence, co-occurrence, subgroup patterns, and automated relevance assessment in retrieval evaluation.

The curated schema retains both the structured attributes needed for relational filtering and the narrative fields needed for lexical and semantic retrieval. **Table 1** links the original form fields to the analytical schema through a concise data dictionary, while **Table 2** summarises the text representations used for ranking and result display. Categories were harmonised through an explicit rule-based mapping from the original form land-use selections and activity descriptions to the 12 controlled labels.

Table 1. Analytical schema summary of the REMINDNET WG3 best-practices dataset (main tables, key field groups, and role in the hybrid system)

Component	Variables (examples)	Type / format	Role in system
Identifiers	'case_id', 'site_key'	strings	Stable referencing, deduplication, relational joins
Provenance	'submission_timestamp'	string	Original form export timestamp for audit trail and version traceability
Location	latitude, longitude	numeric, WGS84 decimal degrees	Spatial summaries, mapping, optional geographic constraints
Site descriptors	country, 'mine_type', status, 'type_of_ore_deposits' (+ 'mine_name')	categorical + string	Relational filtering and stratified analysis
Best-practices narrative	'best_practices' (category selection + optional "Other" text)	semi-structured text	Primary content signal for retrieval
Retrieval representation	'case_card', 'retrieval_text'	derived text (case_card: human-readable summary; 'retrieval_text': fixed ranking input)	'case_card' supports result presentation; 'retrieval_text' is used for TF-IDF and embeddings ranking
Category annotation	cat_* (12 binary flags); 'case_categories' (category)	multi-label (0/1 flags + link table)	Category statistics, co-occurrence, automated relevance labelling

Note (controlled vocabularies):

- 'mine_type': Surface mine, Underground mine.
- 'status': Active, Closed, Abandoned, Other.
- 'type_of_ore_deposits': Metallic, Non-metallic.
- Category vocabulary (12): Community & culture; Non-intensive recreation; Conversion & ecosystem services; Education & research; Renewable Energy; Forestry; Agriculture; Lake or pool; Intensive recreation; Light industrial; Adventure activities; Unclassified (The form includes an 'Other' option, entries are mapped to Unclassified when not assignable.)

2.2. Data preparation

Raw submissions were organised into a standard analytical schema to support relational filtering and retrieval experiments. Records were screened for data completeness and duplicate site entries, then standardised by aligning categorical fields (mine type, status, ore deposit type) with controlled vocabularies and by normalising text formatting. This made Web GIS facets and relational filters more reliable, because each filter corresponded to a single controlled label rather than to inconsistent spellings or near-synonymous terms. Coordinates were converted to WGS84 decimal degrees and checked against valid numeric ranges. To support reproducible deduplication and joins, each retained record was assigned a unique 'case_id' and a reproducible 'site_key' based on the cleaned country and mine name, together with latitude and longitude rounded to four decimal places. Best-practice entries, recorded through predefined land-use categories and activities, including an "Other" option, were harmonised into the 12-category analytical vocabulary used in this study. Categories were harmonised using an explicit rule-based mapping from the original form land-use selections and activity descriptions to the 12 controlled labels. When a submission included multiple land-use entries, all corresponding labels were retained as a multi-label assignment. Ambiguous cases were resolved by referring to the activity description, and any remaining 'Other' entries that could not be assigned were mapped to Unclassified. Consistency checks were then used to confirm that each case's category flags matched the case-category link table and that all labels belonged to the controlled vocabulary.

2.3. Case-card and retrieval-text construction

To support the hybrid retrieval workflow, two deterministic text representations were generated from each cleaned case record: 'case_card' and 'retrieval_text'. The 'case_card' is a user-facing summary used to display retrieved results, whereas 'retrieval_text' is the indexed input used for TF-IDF and embedding-based ranking. This separation keeps the presentation and retrieval layers functionally distinct while ensuring that both are generated reproducibly from the same cleaned case table.

The `case\_card` combines the key information needed for interpretation by users: `case\_id`, country, `mine\_name`, `type\_of\_ore\_deposits`, `mine\_type`, status, latitude, longitude, and the best-practices narrative. It is formatted using a fixed “field: value” template with a constant field order and pipe delimiters. Coordinates are expressed in WGS84 decimal degrees and rounded to four decimal places. The resulting `case\_card` is stored in the case table and used to present retrieved cases.

The `retrieval\_text` representation is generated programmatically at indexing time from country, `mine\_type`, status, `type\_of\_ore\_deposits`, and the best-practices narrative. It excludes identifiers, coordinates, category flags, and case-category links. This ensures that TF-IDF and embedding-based ranking are driven by contributor-provided narrative content and basic site metadata, rather than by direct matches to the controlled category labels used later for relevance assessment. **Table 2** summarises the field composition, fixed ordering, and storage approach for the two representations.

Table 2. Case-card and retrieval_text template specification used for display and ranking

Representation	Purpose in system	Fields included (fixed order)	Formatting rules	Stored in case table
`case_card`	User-facing display of retrieved results (case summary)	`case_id`, country, `mine_name`, `type_of_ore_deposits`, `mine_type`, status, latitude, longitude, `best_practices`	Deterministic “Field: value” template, constant field order, pipe delimiter ()	Yes, stored as single string field (`case_card`)
retrieval_text	Indexed input for ranking (TF-IDF and embeddings)	country, `mine_type`, status, `type_of_ore_deposits`, `best_practices`	Programmatically generated fixed template at indexing time. Excludes identifiers and coordinates. Category flags and case-category links excluded.	No, generated at indexing time (no separate column needed).

Figure 1 summarises how each cleaned relational record is transformed into two text representations: `case\_card` for display and `retrieval\_text` for ranking. It also shows the evaluation rule used throughout the study: controlled category labels are used only for filtering and relevance labelling and are excluded from the ranking input.

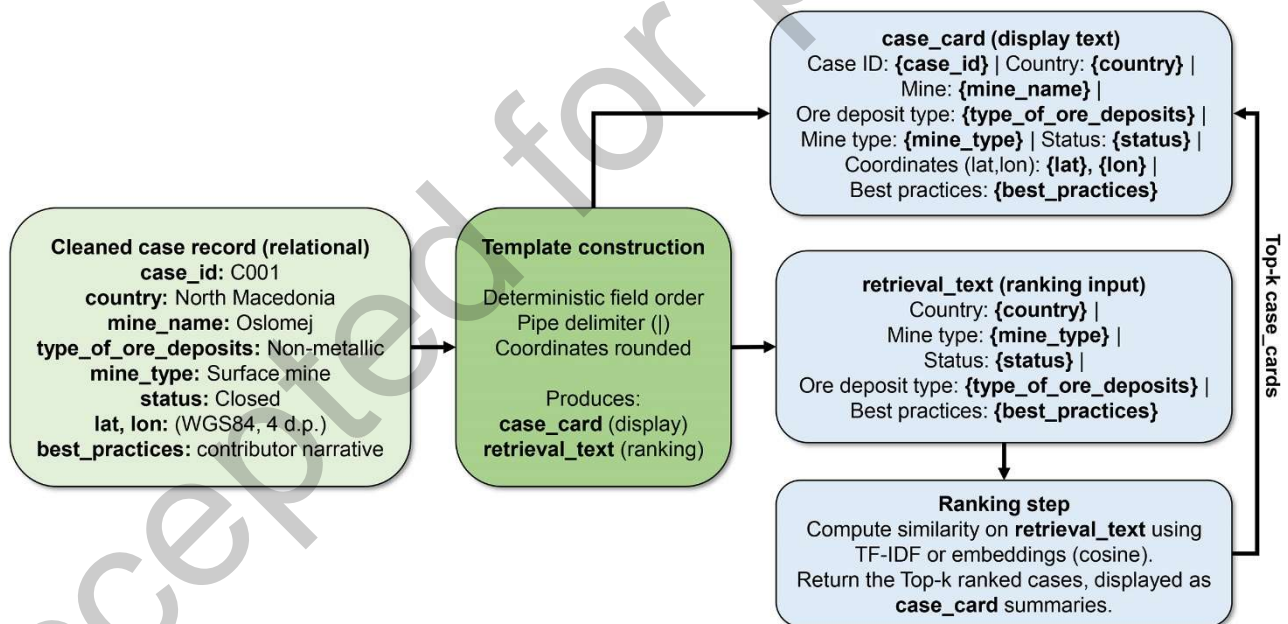


Figure 1. Deterministic construction of `case\_card` and `retrieval\_text` for hybrid retrieval. `case\_card` is used for result display, while `retrieval\_text` is used for TF-IDF and embedding-based ranking; category labels are excluded from the ranking input.

2.4. Hybrid retrieval workflow and ranking models

User-facing retrieval was implemented as a two-stage filter-then-rank process. First, structured metadata fields are used to narrow the candidate set through relational filtering. Second, the remaining cases are ranked according to the similarity between the user query and each case's `retrieval\_text` representation. This design reflects the intended use of the REMINDNET WG3 collection as a decision-support resource: users can apply explicit site constraints, such as mine type, status, ore type, or geography, and then use lexical or semantic ranking to identify the most comparable cases within that filtered subset. **Figure 2** presents the overall architecture of the proposed hybrid knowledge system. The architecture

links the data-ingestion and preparation workflow to the relational database structure, retrieval indexing, user query interface, Web-GIS component, and decision-support outputs. It shows how community-curated records are transformed into a searchable system in which structured filtering, GIS-based exploration, and ranked retrieval operate together.

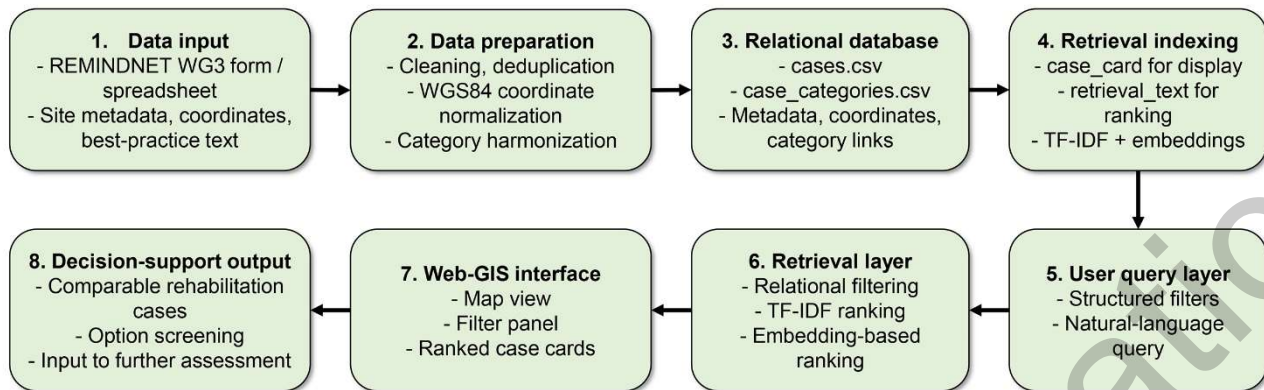


Figure 2. System architecture of the proposed hybrid mine-rehabilitation knowledge system. The figure links data input, preparation, relational storage, retrieval indexing, user query handling, Web-GIS interaction, and decision-support output.

Figure 3 illustrates this workflow: structured fields define the candidate set, similarity ranking is then computed on 'retrieval_text', and the returned items are displayed as 'case_card' summaries. **Figure 3** also shows that all ranking models operate on the same candidate set once filters have been applied, so any performance differences reflect the ranking function rather than differences in search scope.

Relational filtering is applied to the cleaned case table by exact matching on the standardised categorical fields country, 'mine_type', status, and 'type_of_ore_deposits'. Filters are applied only when specified by the user or by an evaluation query. The resulting candidate set is then defined as

$$C(F) = \{ d_i \in D \mid d_i \text{ satisfies all filter constraints } F \} \quad (1)$$

where D denotes the full case corpus and F is the set of structured constraints. If no structured filters are specified (that is, $F = \emptyset$), the candidate set defaults to the full corpus ($C = D$) and ranking is performed over all cases.

Keyword ranking was implemented with term frequency-inverse document frequency (TF-IDF) over the 'retrieval_text' representation. The vectoriser applied lowercasing, English stop-word removal, word n-grams of 1 and 2, and 'min_df = 1'. Cases and queries were represented in the same TF-IDF space, and candidate cases were ranked by cosine similarity,

$$s(q, d_i) = \frac{v(q) \cdot v(d_i)}{\|v(q)\| \|v(d_i)\|} \quad (2)$$

where $v(\cdot)$ denotes the TF-IDF vector. This provides a transparent lexical baseline consistent with classical information retrieval for small-to-medium corpora (Salton & Buckley, 1988).

Semantic ranking was implemented using dense sentence embeddings derived from the same 'retrieval_text' representation (Reimers & Gurevych, 2019). Case embeddings were generated in advance with the Sentence-Transformers implementation of all-MiniLM-L6-v2, producing a 384-dimensional vector for each case. During retrieval, the query text was embedded with the same model, and candidate cases were ranked by cosine similarity between the query embedding and each case embedding (with $v(\cdot)$ in equation 2). Using the same 'retrieval_text' for both TF-IDF and embeddings ensured that differences in performance reflected the ranking model rather than the underlying text representation.

We evaluated both unfiltered and filtered variants for TF-IDF and embeddings. In the unfiltered setting, ranking was performed across all cases ($C = D$), whereas in the filtered setting, relational constraints were applied first and ranking was then performed within the filtered subset ($C = C(F)$), following the intended hybrid workflow. To keep the comparison consistent across methods, each run returned the top- k ranked cases, with $k = 10$ in the evaluation. Results were ordered by descending similarity; exact ties were rare and were resolved using a fixed index order for identical inputs. Given the small corpus size (130 cases), similarity search was computed directly rather than using approximate nearest-neighbour indexing.

All retrieval pipelines and evaluation experiments were implemented in Python, using scikit-learn for TF-IDF vectorisation and Sentence-Transformers for embedding generation, while ranking was based on cosine similarity.

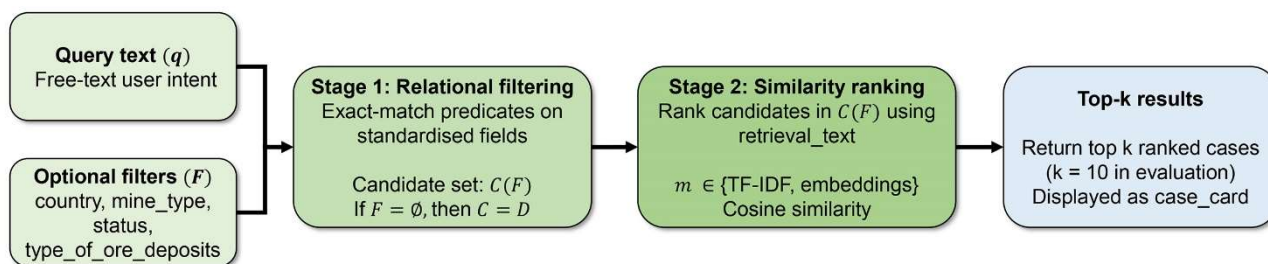


Figure 3. Hybrid retrieval pipeline with relational filtering and keyword or semantic ranking. Structured filters define the candidate set, after which TF-IDF or embedding-based ranking is applied to ‘retrieval_text’ and results are displayed as ‘case_card’ summaries.

2.5. Evaluation protocol

Retrieval performance was evaluated using a benchmark set of 25 natural-language queries (**Table S2**) chosen to reflect realistic information needs in post-mining rehabilitation and repurposing. Each query includes a ‘query_text’, target categories, and optional structured constraints aligned with the relational schema. For this benchmark, constraints were limited to ‘status’, ‘mine_type’, and ‘type_of_ore_deposits’ (**Table S2**). This reflects the intended user workflow, in which structured filters are applied when known and ranking is then performed on the remaining candidate set (Section 2.4).

To enable reproducible evaluation without manual relevance judgements, relevance was defined deterministically from the category annotations in the analytical dataset. For a query with filters F and target category set T_q , a case d_i was labelled relevant if it satisfied all specified structured constraints and contained all target categories. In formal terms, this meant $d_i \in C(F)$ and $T_q \subseteq \text{CatSet}(d_i)$, where $\text{CatSet}(d_i)$ is the set of categories assigned to case d_i from the binary flags and the case-category link table. This automated criterion keeps the evaluation focused on the task of retrieving cases that match both the contextual constraints and the intended post-mining land use. We use strict category overlap as a reproducible proxy for similar rehabilitation goals, because the controlled category vocabulary provides the most consistent thematic descriptor across all contributed records. In a community-curated dataset, narrative descriptions vary in length, terminology, and level of technical detail, whereas the category labels provide a common analytical layer across the full corpus. The strict “all target categories” rule was therefore selected to create a transparent and repeatable benchmark in which all retrieval methods are evaluated against the same relevance definition.

At the same time, this relevance definition should be interpreted as a conservative approximation of practical rehabilitation similarity, not as a complete measure of usefulness to practitioners. Category assignments may contain subjectivity, and semantically relevant cases may be counted as non-relevant when annotations are incomplete or when practical similarity is expressed in the narrative text but not captured by the assigned labels. Consequently, the reported retrieval metrics may underestimate practical relevance for some queries. The automated relevance definition was nevertheless retained because it enables reproducible, transparent, and internally consistent comparison across methods within the current dataset. This makes it possible to benchmark ranking methods reproducibly without additional manual annotation, while keeping the evaluation consistent with the database schema used for faceted browsing.

Five retrieval configurations were compared under the same query conditions:

- filters-only random baseline (expected performance of an unranked result list);
- TF-IDF unfiltered;
- TF-IDF filtered;
- embeddings unfiltered;
- embeddings filtered (hybrid).

Because the filters-only configuration does not produce a ranked result list, its performance was estimated as the mean over 200 random orderings of the filtered candidate set, using a fixed random seed. Filtered variants apply structured constraints before ranking, whereas unfiltered variants rank across the full corpus. In the unfiltered variants, ranking was computed over all cases, but relevance was still defined by the query’s filter constraints and target categories. All ranked methods returned the top- k cases, with $k = 10$, ordered by cosine similarity. Exact ties were rare and were resolved using a fixed index order for identical inputs. Ranking quality was assessed using metrics that reflect practical search behaviour, since users usually focus on the first page of results. We therefore report Precision@5 (P@5), Recall@10 (R@10), Mean Reciprocal Rank@10 (MRR@10), and nDCG@10 under binary relevance. P@5 measures the proportion of relevant cases in the top five results, while R@10 measures the proportion of all relevant cases retrieved within the top ten. MRR@10 captures how early the first relevant result appears, up to rank 10, and nDCG@10 is a position-weighted metric that places greater emphasis on correct ordering near the top of the list (Järvelin & Kekäläinen, 2002). Metrics were calculated for each query and then summarised as macro-averages across queries. Uncertainty in the macro-averaged metrics was estimated using paired non-parametric bootstrap resampling over queries (10 000 resamples), and 95% percentile confidence intervals are reported for each method (Efron, 1979). In addition to per-method confidence intervals, we report paired tests on per-query metric differences between competing methods. For each method pair (A,

B) and each metric, the per-query difference $\Delta_q = \text{metric}_A(q) - \text{metric}_B(q)$ was computed across queries q , and the mean Δ was estimated together with a 95% percentile bootstrap CI from 10 000 paired resamples of query indices, preserving the pairing between methods. A Wilcoxon signed-rank test on the Δ vector was used as a non-parametric complement to the bootstrap CI, with zero-difference pairs handled consistently across all paired comparisons (Wilcoxon, 1945). Two-sided p-values are reported by default; one-sided p-values testing "A greater than B" are reported where directional claims are made. Paired tests are reported for the principal method comparisons (dense versus lexical retrieval under identical filtering, ranked methods versus the random filters-only baseline, and filtered versus unfiltered variants for each retriever) over the same three subsets used for the macro averages.

All retrieval and evaluation experiments were run in Python. For each method, filters were applied when specified, ranked lists were generated, and query-level and aggregated metrics were computed at the stated cut-offs. The full 25-query benchmark is reported in **Table S2** (Supplementary Information).

2.6. Descriptive analyses for corpus characterisation

We carried out descriptive analyses to characterise dataset coverage and inform facet design. Category prevalence was calculated as the proportion of cases assigned to each controlled category under multi-label coding. Co-occurrence was examined for the subset of multi-label cases (assigned ≥ 2 categories) using pairwise category counts. Weighted degree centrality was then calculated as the sum of co-occurrence counts connected to each category (Opsahl et al., 2010). Stratified prevalence by mine type, status, and ore deposit type was calculated as within-group category shares, while spatial coverage was summarised with point maps of WGS84 coordinates coloured by mine status.

3. Results and discussion

The results and discussion are organised around two objectives relevant to the development of a decision-support knowledge system for mine rehabilitation. First, we describe the curated dataset in terms of topic prevalence, co-occurrence structure, patterns across key fields, and spatial coverage, to show how the database can be presented through a web interface and faceted filters. Second, we evaluate a realistic filter-then-search workflow to assess how lexical and semantic ranking perform when users seek comparable cases within selected constraints.

3.1. Database coverage relevant to decision support

The curated REMINDNET WG3 best-practices dataset includes 130 unique mine-rehabilitation cases (C001-C130). All retained records contain complete core metadata, namely country, mine name, mine type, operational status, and ore type, together with validated georeferences stored as WGS84 latitude and longitude in decimal degrees. Coverage spans 21 countries, and the dataset includes 65 surface and 65 underground cases. This balance reflects the current contributed sample after curation. Mine status is reported using a controlled vocabulary, with Closed ($n = 63$) and Active ($n = 38$) sites forming the largest groups, followed by Abandoned ($n = 20$) and Other ($n = 9$). Ore deposit type is also standardised, with 70 Non-metallic and 60 Metallic cases. This structured metadata supports reproducible relational filtering in the hybrid retrieval workflow, for example by restricting the search to a given mine type or status before applying text-based ranking. **Figure 4** maps the spatial distribution of cases across Europe, with points coloured by mine status to show the coverage available for decision support.

Each case was assigned to one or more of 12 controlled best-practice categories derived from the "Best Practices" narrative. Across the dataset, there are 239 case-category links, giving a mean of 1.84 categories per case (median 1, range 1-8). Most cases describe a single dominant reuse pathway (78/130, 60%), while a substantial minority combine multiple pathways, including 23.8% with three or more categories. This multi-label structure is important for two reasons: it supports analysis of co-occurring rehabilitation uses, and it provides an objective basis for retrieval evaluation through category overlap. Two coverage issues should be kept in mind when interpreting the following results.

First, country contributions are uneven, which is typical of a community-curated resource. For example, Bulgaria contributes 38 cases, the largest share, followed by Italy (17) and Serbia (12), while several countries are represented by a single case (**Table S1**, Supplementary materials). Many countries fall into the ≤ 2 -case group (**Table S1**), which limits country-specific interpretation and means that geographic patterns should be read as coverage rather than prevalence. Second, controlled vocabularies improve comparability but also reduce some detail. The 12-category scheme standardises heterogeneous descriptions across contributors, but it can under-represent emerging or highly specific practices that do not map clearly onto the predefined labels. These limitations do not reduce the value of the dataset for benchmarking hybrid retrieval, but they do mean that the results should be interpreted as evidence from the current compiled dataset rather than as a statistically representative sample of all European mine-reuse outcomes.

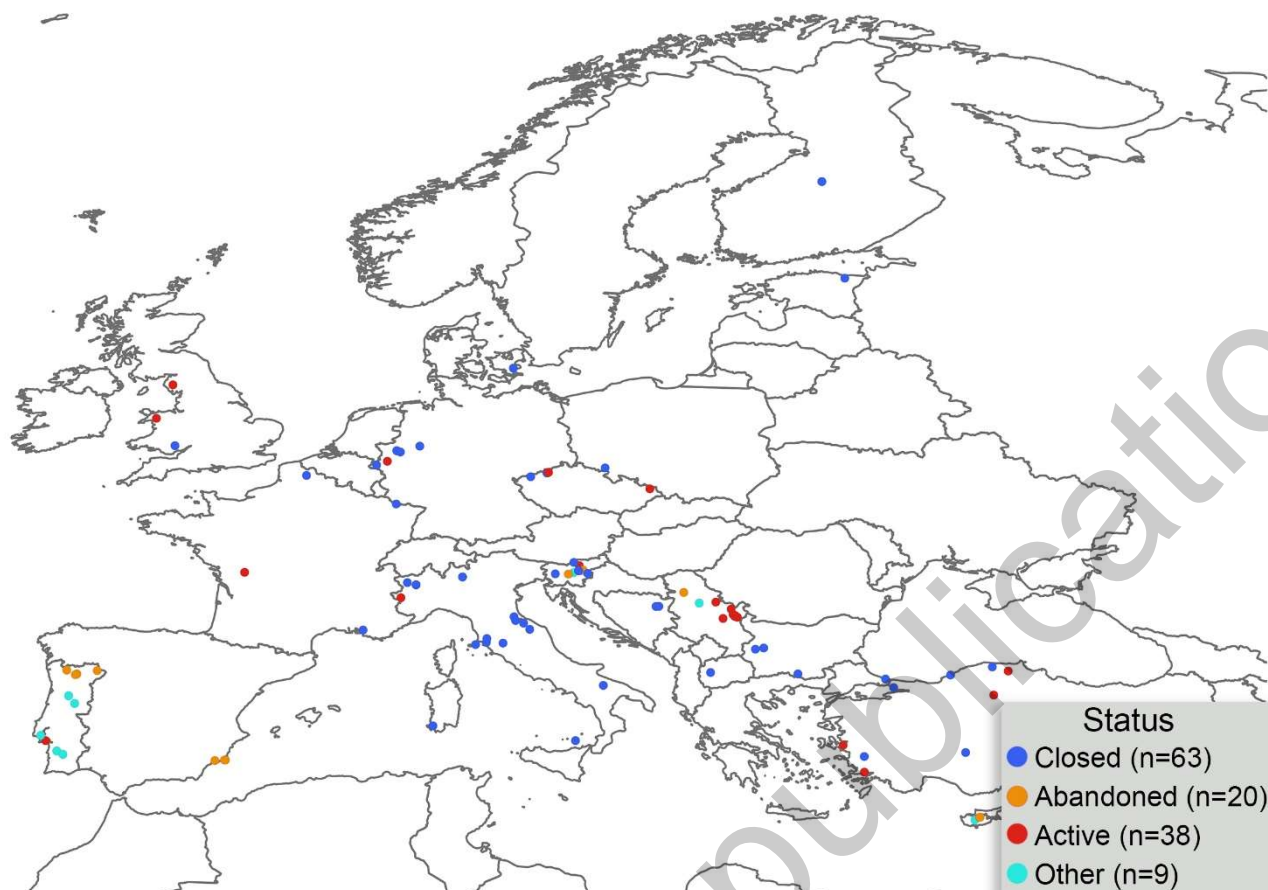


Figure 4. Spatial distribution of cases by mine status ($n = 130$). Points show cleaned WGS84 case locations across Europe, coloured by operational status; the map represents dataset coverage, not a representative European inventory.

3.2. Category prevalence and implications for faceted browsing

The 130 cases were assigned to a controlled set of 12 best-practice categories, producing 239 case-category links (mean 1.84 categories per case, median 1). Prevalence is reported as the proportion of cases assigned to each category under multi-label coding, so the percentages do not sum to 100%. As shown in **Figure 5**, category prevalence is uneven. Community & culture is the most frequent category (58 cases, 44.6%), followed by Non-intensive recreation (41, 31.5%), Conversion & ecosystem services (37, 28.5%), and Education & research (35, 26.9%). Together, these four categories make up a large share of the dataset, suggesting that the current dataset places greater emphasis on socially oriented reuse and relatively low-impact post-mining land uses. Mid-frequency categories include Forestry (16, 12.3%), Renewable Energy (14, 10.8%), and Lake or pool (13, 10.0%). Lower-frequency categories are Agriculture (10, 7.7%), Intensive recreation (9, 6.9%), and Light industrial (4, 3.1%), while Adventure activities and Unclassified occur only once each (1, 0.8% per category). This imbalance is important when interpreting both the descriptive analyses and the retrieval benchmarks, because rare categories naturally produce fewer relevant candidates and can increase variation in query-level metrics.

Most case descriptions correspond to a single dominant reuse pathway, but a substantial minority show multi-use portfolios. Overall, 78 cases (60.0%) are assigned one category, while 52 cases (40.0%) include two or more categories. Multi-label portfolios of three or more categories occur in 31 cases (23.8%), reflecting combined land uses such as recreation paired with ecological restoration or education-oriented reuse. A small number of cases are highly multi-functional, with up to eight categories assigned. These cases are especially informative for semantic retrieval because their narratives span multiple themes that may not be captured through keyword overlap alone. This class imbalance is relevant for the co-occurrence analysis in Section 3.3 and should also be taken into account when interpreting retrieval metrics for rare categories in Sections 3.6-3.7. It also has implications for interface design: high-frequency categories are natural candidates for default facets in the web interface, whereas less frequent categories are more likely to benefit from semantic retrieval, especially when keyword matching is limited by small candidate sets.

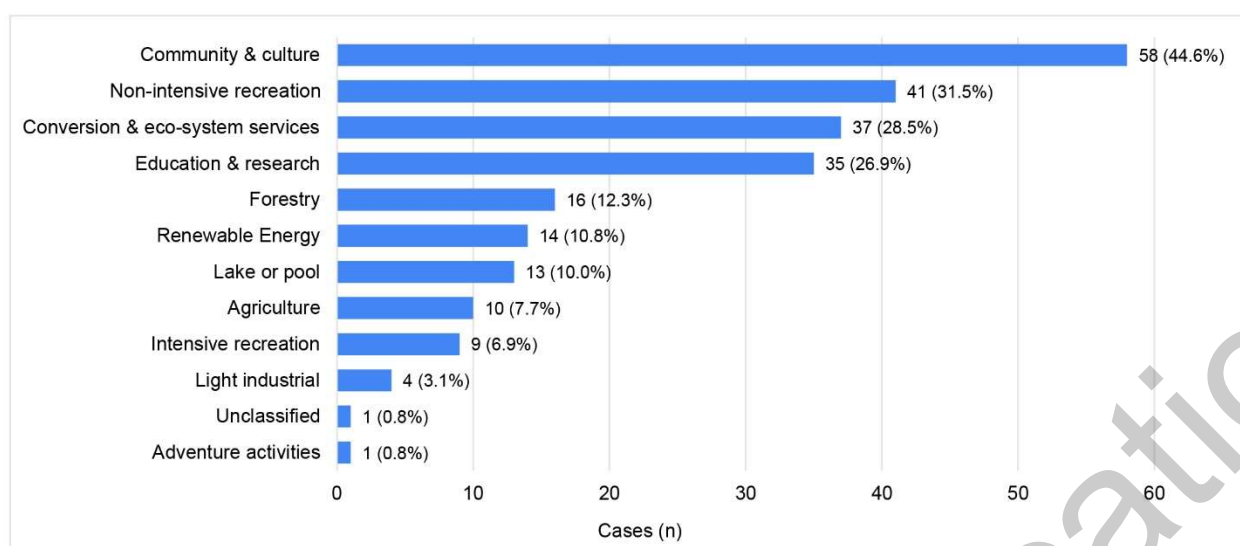


Figure 5. Category prevalence across controlled best-practice categories (n = 130). Bars show multi-label category counts and shares; percentages do not sum to 100% because cases may have multiple categories.

3.3. Co-occurrence structure and implications for query expansion

Co-occurrence was assessed on the subset of multi-label cases (assigned ≥ 2 categories, n = 52), where each pairwise count represents the number of cases tagged with both categories. Across the 12-category scheme, 43 of the 66 possible category pairs were observed at least once, showing that post-mining reuse in the compiled dataset often involves multiple uses rather than a single dominant pathway. At the same time, these co-occurrences are not random, but tend to cluster around a small set of socially and ecologically oriented themes. This pattern is shown by the most frequent co-occurring pairs (**Figure 6**) and the full category co-occurrence matrix (**Figure 7**). Pair counts are computed for the multi-label subset (n = 52), and each pair is also reported as a share of the full corpus (N = 130) to make the results easier to interpret.

The strongest co-occurrence patterns connect community-oriented reuse, low-intensity recreation, ecological conversion, and education or research. The most common category pair was Community & culture + Education & research, which occurred in 25 cases, representing 19.2% of all cases (N = 130) and 48.1% of multi-label cases. This was followed by Community & culture + Non-intensive recreation (24 cases, 18.5% of all cases) and Conversion & ecosystem services + Non-intensive recreation (21 cases, 16.2% of all cases). Other frequent combinations included Education & research + Non-intensive recreation (18 cases), Community & culture + Conversion & ecosystem services (17 cases), and Conversion & ecosystem services + Education & research (13 cases). Taken together, these patterns suggest a consistent picture of co-use in which sites are repurposed through combinations of public-oriented functions, such as cultural heritage, museums, and events, together with nature-based uses, accessible recreation, and, in many cases, educational or research activities.

This pattern is also reflected in the network centrality results. Weighted degree centrality was calculated as the sum of pairwise co-occurrence counts connected to each category node. Non-intensive recreation, Community & culture, and Conversion & ecosystem services emerge as the main hubs, because they co-occur with the widest range of other categories and have the highest weighted degree totals (87, 79, and 73, respectively). The central role of these categories is clear in the dense upper-left block of **Figure 7**. Categories are ordered by overall frequency along the diagonal, which makes the concentration among high-prevalence categories easier to see. Education & research has fewer distinct partners but a high co-occurrence weight (65), consistent with frequent pairing with the hub categories rather than broad mixing across all themes. Rare categories contribute little to the network. For example, Adventure activities appears only once and is involved in a single pairing, while Unclassified does not co-occur with any other category because it appears in a single-category case. These imbalances are also relevant for the retrieval evaluation discussed later, because queries targeting rare categories have smaller relevant pools and are therefore more sensitive to ranking errors. These co-occurrence patterns support interface features such as related-category suggestions and query expansion, for example by recommending themes that often appear together. They also help explain why semantic ranking is especially useful for queries that span several categories within a single narrative.

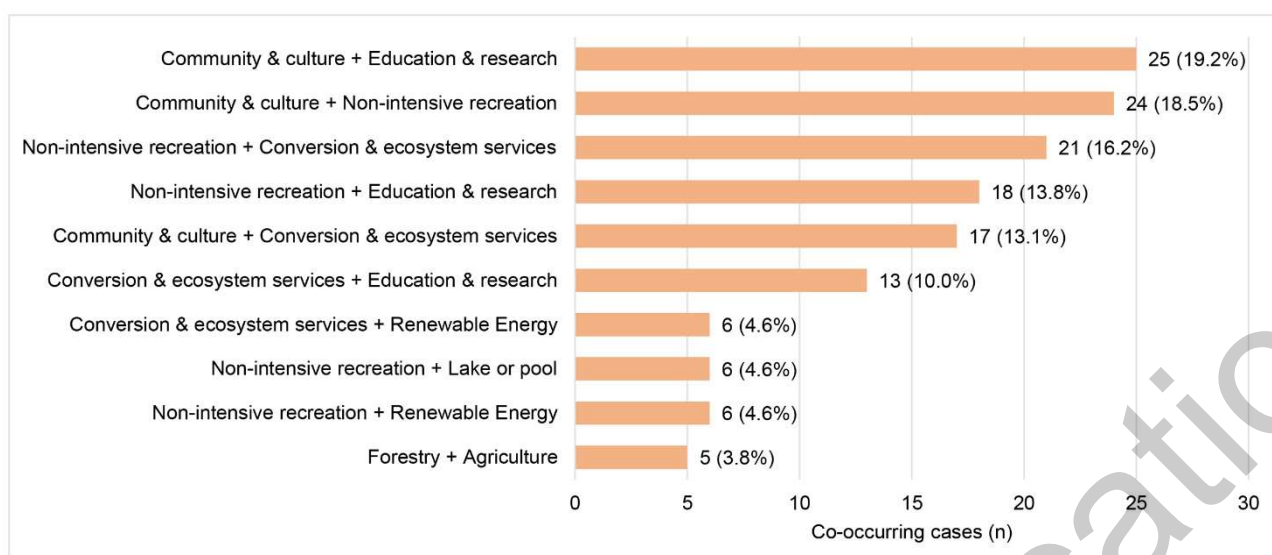


Figure 6. Top co-occurring category pairs in the multi-label subset ($n = 52$). Pair counts show how often two controlled categories occur together within the same case.

Community & culture	58	24	17	25	4	2	3	2	1	0	1	0
Non-intensive recreation	24	41	21	18	6	4	4	6	3	1	0	0
Conversion & ecosystem services	17	21	37	13	6	3	4	5	3	1	0	0
Education & research	25	18	13	35	3	2	2	2	0	0	0	0
Renewable Energy	4	6	6	3	14	3	3	1	1	1	0	0
Forestry	2	4	3	2	3	16	5	2	1	1	0	0
Agriculture	3	4	4	2	3	5	10	1	2	1	0	0
Lake or pool	2	6	5	2	1	2	1	13	2	1	0	0
Intensive recreation	1	3	3	0	1	1	2	2	9	1	0	0
Light industrial	0	1	1	0	1	1	1	1	1	4	0	0
Adventure activities	1	0	0	0	0	0	0	0	0	0	1	0
Unclassified	0	0	0	0	0	0	0	0	0	0	0	1

Figure 7. Co-occurrence matrix of controlled best-practice categories in the REMINDNET corpus. Off-diagonal cells show pairwise co-occurrence counts among multi-label cases ($n = 52$), while diagonal cells show total category frequency in the full dataset ($N = 130$).

3.4. Stratified patterns across key filters (country, mine type, status, ore type)

To test whether structured metadata supports useful relational filtering, category prevalence was stratified by mine type, status, ore type, and country. Within each group g , the prevalence of category c was calculated as the within-group share of cases assigned to that category,

$$p_{g,c} = \frac{n_{g,c}}{n_g} \quad (3)$$

where n_g is the number of cases in group g and $n_{g,c}$ is the number of cases in g assigned to category c . Because categories are multi-label, prevalence values can sum to more than 100% within a group. These comparisons show that the same best-practice categories play different roles across mine contexts, which supports the two-stage workflow of filtering first and then ranking by text similarity. In practice, they also indicate which structured facets should be more prominent in the user interface, namely mine type, status, ore type, and country, because these fields change the expected category profile and therefore the set of comparable cases.

3.4.1. Mine type $\times$ category (surface vs underground)

Category portfolios differ clearly between surface and underground operations. Underground cases are disproportionately associated with public-oriented reuse and knowledge-related functions. Community & culture is present in 67.7% of underground cases (44/65), compared with 21.5% of surface cases (14/65), while Education & research accounts for 38.5% (25/65) versus 15.4% (10/65). Non-intensive recreation is also more common in underground settings (43.1% vs 20.0%). In contrast, surface mines more often exhibit landscape-scale restoration and land conversion pathways, including Forestry (23.1% vs 1.5%), Lake or pool (15.4% vs 4.6%), Agriculture (12.3% vs 3.1%), and Intensive recreation (12.3% vs 1.5%). Together, these differences show that mine type is a useful first-stage filter in the hybrid workflow, because it strongly conditions the category profile likely to be retrieved before text ranking is applied (**Figure 8a**).

3.4.2. Status $\times$ category (closed vs active vs abandoned vs other)

Reuse patterns also vary by operational status. Active mines show relatively low representation of post-closure leisure and ecological transition pathways, including Non-intensive recreation (10.5%, 4/38) and Conversion & ecosystem services (13.2%, 5/38), while maintaining moderate representation of production-compatible uses such as Forestry (21.1%, 8/38) and Agriculture (13.2%, 5/38). In comparison, abandoned sites are more often associated with restorative and public-oriented portfolios, particularly Conversion & ecosystem services (55.0%, 11/20), Non-intensive recreation (45.0%, 9/20), and Education & research (45.0%, 9/20). Closed sites occupy an intermediate position, with Community & culture (49.2%, 31/63) and Non-intensive recreation (36.5%, 23/63) as the most common categories (**Figure 8b**). The Other status group ($n = 9$) is reported for transparency, but percentages are unstable at this sample size and should be interpreted as descriptive only.

3.4.3. Ore type $\times$ category (metallic vs non-metallic)

Ore type also differentiates reuse patterns. Metallic sites are more frequently associated with cultural and knowledge-oriented reuse, with Community & culture present in 51.7% (31/60) and Education & research in 38.3% (23/60), compared with 38.6% (27/70) and 17.1% (12/70), respectively, in non-metallic sites. In comparison, waterbody formation and more intensive leisure uses are more strongly represented in non-metallic settings: Lake or pool occurs in 18.6% of non-metallic cases (13/70) and Intensive recreation in 12.9% (9/70), while neither category is observed among metallic cases in this dataset. Taken together, these differences support ore type as a practical filter for narrowing candidate sets in targeted queries (**Figure 8c**).

a)	Underground mine (n=65)	67.7	43.1	30.8	38.5	9.2	1.5	3.1	4.6	1.5	0	1.5	0
	Surface mine (n=65)	21.5	20	26.2	15.4	12.3	23.1	12.3	15.4	12.3	6.2	0	1.5
b)	Closed (n=63)	49.2	36.5	27	20.6	12.7	7.9	4.8	12.7	9.5	3.2	0	0
	Active (n=38)	31.6	10.5	13.2	15.8	5.3	21.1	13.2	5.3	5.3	2.6	2.6	2.6
	Abandoned (n=20)	40	45	55	45	15	15	10	10	0	5	0	0
	Other (n=9)	77.8	55.6	44.4	77.8	11.1	0	0	11.1	11.1	0	0	0
c)	Metallic (n=60)	51.7	35.0	35.0	38.3	10.0	13.3	6.7	0.0	0.0	5.0	0.0	1.7
	Non-metallic (n=70)	38.6	28.6	22.9	17.1	11.4	11.4	8.6	18.6	12.9	1.4	1.4	0.0
		Community & culture	Non-intensive recreation	Conversion & ecosystem services	Education & research	Renewable energy	Forestry	Agriculture	Lake or pool	Intensive recreation	Light industrial	Adventure activities	Unclassified

Figure 8. Land-use category prevalence by key filters in the REMINDNET rehabilitation cases (cell values: within-group %, multi-label coding; a) mine type $\times$ category, b) status $\times$ category, c) ore deposit type $\times$ category; group sizes n in row labels).

3.4.4. Country $\times$ category

Country-level contributions are uneven, and category profiles vary across the principal contributing countries, suggesting that geographic filtering may be useful in user-facing search. For example, Bulgaria ($n = 38$) shows greater representation of land-based conversion pathways such as Forestry (23.7%) and a relatively low share of Community & culture (10.5%), whereas Italy ($n = 17$) is characterized by a very high prevalence of Community & culture (94.1%), often in combination with education and recreation categories. Portugal ($n = 10$) exhibits a distinctly multi-functional profile, with high prevalence of Non-intensive recreation (70.0%), Education & research (70.0%), and Conversion & ecosystem

services (60.0%). These contrasts should be treated as indicative rather than definitive, since they may reflect both national practice and reporting patterns, and several countries are represented by small sample sizes. Accordingly, we summarize only the main contributing countries here. Full country-by-category cross-tabulations can be reproduced from the released dataset available on the REMINDNET project website.

3.5. Spatial distribution and transferability context

The cleaned georeferences provide a consistent spatial view of the best-practice cases in the dataset. Case locations span Europe, and the map offers an immediate overview of geographic coverage for users assessing transferability across similar regulatory and geographic settings. In the intended interface, selecting a country or region is likely to be an early step in the search process, which makes map views and country filtering important design components (**Figure 4**).

Spatial density is highest in the regions with the greatest number of contributions, particularly South-Eastern Europe and the Mediterranean, with clear clustering in Bulgaria ($n = 38$) and secondary clusters in Italy ($n = 17$), Serbia ($n = 12$), Slovenia ($n = 10$), Portugal ($n = 10$), and Turkey ($n = 10$) (**Figure 4**). This concentration helps identify areas where the database can already support more detailed country-specific retrieval, but it also suggests that the observed spatial pattern is shaped largely by reporting activity and network participation. We therefore interpret the map as a descriptive view of dataset coverage rather than as evidence for broader conclusions about the geographic prevalence of rehabilitation practices.

Dominant practice categories are distributed across multiple countries rather than concentrated in a single region. Community & culture is represented in 13 countries (58 cases), Non-intensive recreation in 12 countries (41 cases), and both Conversion & ecosystem services and Education & research in 13 countries (37 and 35 cases, respectively). By contrast, more specialised categories show a more limited geographic distribution. For example, Forestry appears in six countries (16 cases). Considered alongside the status-coded distribution in **Figure 4**, these patterns support the use of map-based summaries as a rapid overview of spatial coverage and help contextualise the later retrieval evaluation, where rare or geographically sparse categories generate smaller pools of relevant cases.

3.6. Retrieval evaluation under a filter-then-search workflow

Given the structured metadata and multi-label thematic patterns described above, we next assess whether a hybrid workflow, combining relational filtering with semantic ranking, improves the retrieval of relevant cases in practice. Retrieval performance was evaluated using 25 benchmark queries designed to reflect realistic user needs, including topic-based queries with optional structured filters (**Table S2**). Relevance was defined strictly using the controlled labels: a retrieved case was considered relevant only if it satisfied the query's relational constraints, where applicable, and contained all target categories in its multi-label assignment (Methods). We compared five methods: filters-only random baseline, TF-IDF (unfiltered, filtered), and embeddings (unfiltered, filtered), where the filtered variants implement the intended hybrid workflow (filter first, then rank).

3.6.1. Overall macro performance across queries

Across all 25 queries, the filtered embeddings method showed the highest macro-averaged early-ranking metrics, with $P@5 = 0.768$, $nDCG@10 = 0.810$, and $MRR@10 = 0.847$. Recall at 10 was comparable to the best TF-IDF configuration ($R@10 = 0.494$). By comparison, filtered TF-IDF reached $P@5 = 0.704$, $nDCG@10 = 0.762$, and $MRR@10 = 0.734$, with a similar $R@10$ of 0.496 (**Figure 9**). The filters-only random baseline performed much worse ($P@5 = 0.210$, $nDCG@10 = 0.213$, $R@10 = 0.109$), indicating that relational filtering alone, without semantic ranking, is not sufficient to return relevant items consistently on the first page of results. To assess uncertainty around the macro averages, we computed 95% confidence intervals using non-parametric bootstrap resampling over queries (10 000 resamples). The filtered embeddings method yielded $P@5 = 0.768$ (95% CI 0.648-0.880), $nDCG@10 = 0.810$ (95% CI 0.699-0.907), and $MRR@10 = 0.847$ (95% CI 0.727-0.953), whereas filtered TF-IDF yielded $P@5 = 0.704$ (95% CI 0.552-0.840), $nDCG@10 = 0.762$ (95% CI 0.624-0.882), and $MRR@10 = 0.734$ (95% CI 0.583-0.873).

Compared with filtered TF-IDF, the hybrid embeddings method showed mean per-query Δ of +0.064 on $P@5$, +0.048 on $nDCG@10$ and +0.112 on $MRR@10$, with $R@10$ effectively unchanged (−0.002). Paired analysis of these per-query differences did not, however, establish any of these gaps as statistically reliable: paired-bootstrap 95% CIs were [−0.064, +0.200] on $P@5$, [−0.076, +0.180] on $nDCG@10$, [−0.034, +0.258] on $MRR@10$ and [−0.083, +0.076] on $R@10$, all crossing zero; two-sided Wilcoxon signed-rank p-values were 0.327, 0.616, 0.161 and 0.910 respectively. The per-query sign distribution is consistent with this picture: on $P@5$ seven queries favoured embeddings and seven favoured TF-IDF with eleven ties, and on $nDCG@10$ the split was nine and nine with seven ties. Excluding the single query with zero relevant cases under the strict relevance definition gives the same paired-test conclusions, with slightly higher absolute macro values ($P@5 = 0.800$ and $nDCG@10 = 0.844$ for the hybrid embeddings method). At the current scale of 25 queries and 130 cases, the headline numerical advantage of dense over lexical retrieval is therefore not supported by paired tests; the substantive contribution of the hybrid design is characterised in Section 3.6.2.

To avoid overinterpretation, the retrieval results should be read at three levels. First, descriptively, the filtered embedding workflow produced the highest numerical macro-average scores for $P@5$, $nDCG@10$, and $MRR@10$. Second, statistically, the paired tests do not support a general conclusion that filtered embeddings are superior to filtered

TF-IDF in the current benchmark. The better supported findings are that ranked retrieval improves substantially over an unranked filters-only baseline, and that applying structured filters improves the embedding workflow. Third, practically, these results support the use of embeddings as part of a hybrid filter-then-rank decision-support workflow, but not as a replacement for lexical retrieval. TF-IDF remains a competitive and transparent baseline, while embeddings may be most useful where query intent is expressed through varied terminology or multi-functional rehabilitation narratives.

3.6.2. Effect of relational filtering

At the overall benchmark level, relational filtering improved both keyword-based and embedding-based retrieval, with embeddings showing the larger gains in early precision. In this benchmark, 7 of the 25 queries include a single structured constraint (status, mine type, or ore type; **Table S2**), whereas the remaining 18 are unfiltered thematic queries. As a result, the overall gains from filtering are likely conservative, because the filtered and unfiltered variants are identical for most queries. The aggregate improvements reported here should therefore be viewed as lower-bound estimates of the benefit of a filter-first interface in practice. Future benchmarks should include a larger share of constraint-rich queries, including country or regional constraints and multi-field filters (country, mine type, status, ore type), to better represent real decision-support use. For embeddings, applying filters increased $P@5$ from 0.720 to 0.768 (+0.048) and $nDCG@10$ from 0.769 to 0.810 (+0.041). For TF-IDF, the improvement in $P@5$ was smaller, rising from 0.696 to 0.704 (+0.008), while $nDCG@10$ increased from 0.737 to 0.762 (+0.025). Recall at 10 improved by a similar margin for both approaches (approximately +0.025), which suggests that filtering removes clearly irrelevant cases and directs the ranking toward a smaller, contextually consistent candidate set. Taken together, these results support the two-stage workflow described in Methods, with structured metadata narrowing the candidate space and ranked retrieval ordering cases more effectively than an unranked filtered list (**Figure 9**). The paired analysis further suggests that filtering benefits the embedding workflow more clearly than the TF-IDF workflow, although the dense and lexical filtered rankings are not statistically distinguishable at the current benchmark scale.

For the seven single-constraint queries (Q16-Q22; **Table S2**), relational filtering reduces the candidate pool from 130 cases to a mean of 54 cases (range 20-70), so ranking is performed on a smaller and more contextually coherent set. Within this subset, filtered TF-IDF achieves $P@5 = 0.714$ and $nDCG@10 = 0.717$, whereas filtered embeddings achieves $P@5 = 0.657$ and $nDCG@10 = 0.631$. However, the gain attributable to filtering, calculated relative to the unfiltered version of each method, is larger for embeddings (+0.171 $P@5$, +0.146 $nDCG@10$) than for TF-IDF (+0.029 $P@5$, +0.089 $nDCG@10$). The difference from the overall ranking is not unexpected, because this constrained subset is small ($n = 7$) and more strongly driven by explicit query terms, which keeps TF-IDF highly competitive. Paired analysis across all 25 queries supports this filtering gain for the dense retriever ($nDCG@10 +0.041$, paired-bootstrap CI [+0.007, +0.091], Wilcoxon $p = 0.043$) but not for the lexical retriever, whose corresponding gain (+0.025) does not reach conventional significance; the magnified effect on the constrained subset (+0.146 versus +0.089 on $nDCG@10$) is consistent with this pattern. The strongest supported interpretation is therefore not that embedding-based retrieval is generally superior to lexical retrieval, but that structured filtering and ranked retrieval play complementary roles. Filtering reduces the search space to contextually plausible cases, while ranking orders these cases according to textual relevance. In the present benchmark, this interaction is most clearly supported for the embedding workflow, although TF-IDF remains highly competitive for keyword-specific and low-frequency categories.

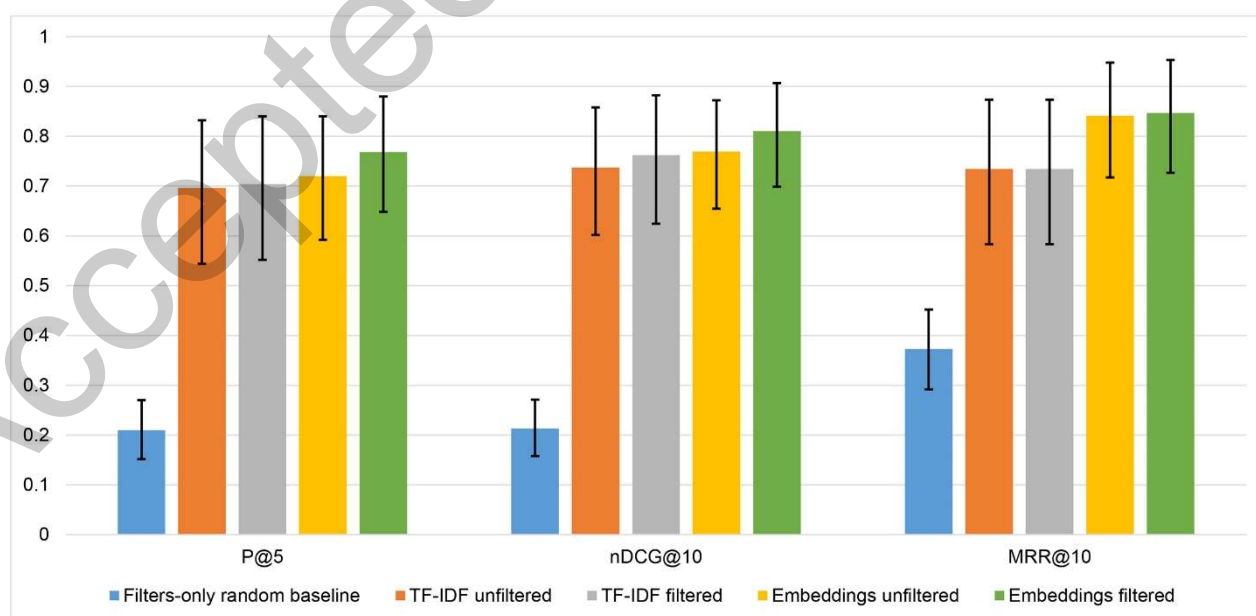


Figure 9. Retrieval performance across 25 benchmark queries. Bars show macro-averaged $P@5$, $R@10$, $nDCG@10$, and $MRR@10$ for the evaluated methods; error bars show 95% bootstrap confidence intervals.

3.6.3. Query-level behaviour and coverage limits

Query-level examples suggest that embeddings can be useful when relevant content is described in diverse language that is not consistently captured by lexical matching. The benchmark query set does not yet include country-constrained queries, because the current filters are limited to status, mine type, and ore type. This is an important limitation, since country filters are central to map-based browsing and to assessing transferability. Future retrieval benchmarks should therefore include country constraints and multi-constraint scenarios, alongside the stratified evidence reported in Section 3.4.4. Because relevance is defined through strict category overlap, the evaluation is objective and reproducible, but it remains conservative when category labels are incomplete, since semantically relevant cases that are not tagged with the target category, as illustrated in Section 3.7, are counted as non-relevant. For example, for Q13 (recreation combined with ecosystem restoration; **Table S2**), filtered TF-IDF returned no relevant results in the top 5 ($P@5 = 0.0$), whereas filtered embeddings achieved $P@5 = 1.0$ and a much higher nDCG (0.066 vs 0.931). Similar gains are observed for Q24 (reforestation and tree planting) and Q12 (cultural heritage combined with education/research), where embeddings substantially improved both first-page precision and rank positions.

In contrast, embeddings underperformed TF-IDF for a small set of highly keyword-specific or low-frequency categories, where exact lexical matches may be more important. Notable examples include Q10 (Light industrial) and Q08 (Agriculture), for which TF-IDF achieved near-perfect ranking, whereas embeddings retrieved semantically similar but label-mismatched cases. This pattern is not unexpected, given the rarity of these categories in the corpus (Section 3.2), which increases the proportional effect of ranking errors.

Finally, one benchmark query (Q18) had zero relevant cases under the strict relevance definition. This should be interpreted as a coverage gap in the current database for that particular combination of intent and constraints, rather than as evidence of model failure. We therefore report metrics both including and excluding this query, and provide full per-query results together with representative top-k outputs in the supplementary material.

3.7. Qualitative examples and failure modes

To illustrate query-level behaviour, we selected three benchmark queries for which embeddings with relational filtering produced the largest improvements in early ranking performance (Q13, Q12, Q24; **Table S2**). For each query, we summarise the top three retrieved cases for filtered TF-IDF and for the hybrid embeddings approach, where filtering is applied before semantic ranking. Case cards follow the standardised schema defined in Methods, and the assigned controlled categories are additionally reported to make the evaluation basis explicit.

Example 1, Q13: recreation combined with ecosystem restoration

Query: “Sites combining recreation with ecosystem restoration (green infrastructure)”

Target categories: Non-intensive recreation; Conversion & ecosystem services

Performance shift: $P@5 = 0.0 \rightarrow 1.0$, $nDCG@10 = 0.066 \rightarrow 0.931$, $MRR@10 = 0.111 \rightarrow 1.000$

TF-IDF filtered, top-3 (none satisfy both target categories):

- C098 (Bulgaria), “Pernik Mines”, Underground mine, Closed. Categories: Non-intensive recreation.
- C097 (Bulgaria), “Pernik Mines”, Surface mine, Closed. Categories: Non-intensive recreation.
- C123 (Bulgaria), “Plakalnitsa” tailings pond, “Plakalnitsa” mine, Underground mine, Abandoned. Categories: Non-intensive recreation.

Embeddings filtered, top-3 (all satisfy both target categories):

- C042 (Portugal), Estufa Fria – São Sebastião da Pedreira – Lisboa, Surface mine, Other. Categories: Conversion & ecosystem services; Non-intensive recreation; Intensive recreation.
- C068 (France), Nord-Pas-de-Calais coal mine, Underground mine, Closed. Categories: Conversion & ecosystem services; Non-intensive recreation; Intensive recreation; Community & culture; Agriculture.
- C045 (Portugal), Jales, Underground mine, Abandoned. Categories: Conversion & ecosystem services; Non-intensive recreation; Community & culture; Education & research.

Interpretation: TF-IDF appears to favour recreation-related lexical cues and therefore retrieves cases tagged only with Non-intensive recreation, whereas embeddings place multi-functional cases that combine recreation with ecological conversion at the top of the ranking.

Example 2, Q12: cultural heritage combined with education/research

Query: “Examples combining cultural heritage and education/research at the same site”

Target categories: Community & culture; Education & research

Performance shift: $P@5 = 0.4 \rightarrow 1.0$, $nDCG@10 = 0.453 \rightarrow 1.000$, $MRR@10 = 0.250 \rightarrow 1.000$

TF-IDF filtered, top-3 (education-only, missing the cultural component):

- C109 (Bulgaria), “Chelopech” mine – DPM Metals, Underground mine, Active. Categories: Education & research.
- C112 (Bulgaria), Eastern dump, “Elatsite MED” JSC, Surface mine, Active. Categories: Education & research.
- C081 (Czechia), Ostrava Karviná district, Underground mine, Active. Categories: Education & research.

Embeddings filtered, top-3 (all satisfy both target categories):

- C023 (Serbia), Crveni Breg, Underground mine, Other. Categories: Community & culture; Education & research.
- C046 (Portugal), Minas de Argemela e Recheira-Sempre, Underground mine, Other. Categories: Community & culture; Education & research; Non-intensive recreation.
- C038 (Portugal), Urgeiriça, Underground mine, Other. Categories: Community & culture; Education & research; Conversion & ecosystem services; Non-intensive recreation.

Interpretation: When the query requires a specific combination of themes, semantic ranking more reliably prioritises cases that combine both cultural and educational functions, rather than cases matching only one part of the request.

Example 3, Q24: reforestation on spoil heaps or dumps

Query: “Examples of reforestation and tree planting on spoil heaps or dumps”

Target category: Forestry

Performance shift: $P@5 = 0.0 \rightarrow 0.8$, $nDCG@10 = 0.000 \rightarrow 0.716$, $MRR@10 = 0.000 \rightarrow 0.500$

TF-IDF filtered, top-3 (no Forestry-labelled cases):

- C049 (Cyprus), Agrokippia Mine, Surface mine, Abandoned. Categories: Conversion & ecosystem services; Education & research.
- C001 (North Macedonia), Oslomej, Surface mine, Closed. Categories: Renewable Energy.
- C003 (Slovenia), Sitarjevec, Underground mine, Other. Categories: Community & culture; Education & research; Conversion & ecosystem services; Non-intensive recreation.

Embeddings filtered, top-3 (Forestry cases appear at ranks 2–3):

- C049 (Cyprus), Agrokippia Mine, Surface mine, Abandoned. Categories: Conversion & ecosystem services; Education & research.
- C070 (Spain), Minas de Mazarrón, Surface mine, Abandoned. Categories: Forestry; Conversion & ecosystem services; Non-intensive recreation; Agriculture; Renewable Energy.
- C071 (Spain), La Unión, Surface mine, Abandoned. Categories: Forestry; Community & culture; Education & research; Conversion & ecosystem services; Non-intensive recreation; Agriculture; Renewable Energy.

Interpretation: This example also highlights a limitation of strict label-based evaluation. Although embeddings retrieves two Forestry-labelled cases at ranks 2 and 3, the top-ranked case, C049, does not carry the Forestry label in the controlled vocabulary. This may reflect a labelling issue or partial mismatch between narrative description and assigned category, but that interpretation remains uncertain because the underlying narrative text is not reproduced here.

3.8. Practical contribution to mine rehabilitation decision-making

The practical value of the proposed system lies in supporting early-stage mine closure and rehabilitation planning, where practitioners must compare possible post-mining land-use options under site-specific constraints. The system is not intended to replace engineering design, regulatory assessment, environmental impact assessment, or stakeholder consultation. Instead, it provides a structured knowledge-support tool that helps users identify comparable rehabilitation examples more systematically than a spreadsheet, static map, or keyword-only search.

In a practical workflow, users can first define the site context through structured filters such as mine type, operational status, ore deposit type, country, or region. They can then submit a natural-language query, for example to identify closed underground mines reused for cultural or tourism functions, abandoned surface mines with ecosystem-restoration outcomes, or non-metallic surface mines converted into water-based recreation areas. The system ranks the filtered cases by textual similarity and presents them as compact case cards containing the site context and rehabilitation narrative. These examples can support the screening of post-mining land-use alternatives, comparison of transferable rehabilitation models, preparation of conceptual options, selection of examples for stakeholder discussions, and identification of cases requiring more detailed technical, regulatory, environmental, or socio-economic assessment.

From a mining-engineering perspective, the retrieved cases are most useful during the option-screening and conceptual-design stages of closure planning. They can help practitioners assess whether a proposed post-mining use is broadly compatible with the mine type, closure status, ore-deposit context, site morphology, environmental constraints, and likely regulatory requirements. For example, pit-lake cases can inform early questions on water quality, slope stability,

access control, and long-term monitoring; cultural or tourism reuse cases can highlight issues of underground safety, visitor access, heritage preservation, and emergency planning; and ecosystem-restoration cases can support discussion of landform design, soil reconstruction, vegetation establishment, habitat creation, and post-closure maintenance. The system does not provide engineering design parameters directly, but it helps practitioners identify relevant precedents that can guide the questions, constraints, and design alternatives considered in subsequent technical studies.

Retrieval quality is practically meaningful because mine rehabilitation planning often starts from a limited set of familiar or easily searchable examples. If relevant precedents are missed, the range of considered rehabilitation options may be unnecessarily narrow. Improved early ranking places more contextually relevant cases on the first results page, where users are most likely to inspect them. In this study, ranked retrieval clearly outperformed the filters-only random baseline, showing that filtering alone is insufficient when candidate sets remain heterogeneous. Although filtered embeddings achieved the highest numerical early-ranking scores, their advantage over filtered TF-IDF was not statistically reliable at the current benchmark scale. Therefore, the main practical contribution should be interpreted as the combination of transparent structured filtering with ranked retrieval, rather than as evidence that semantic retrieval alone produces better rehabilitation decisions.

The retrieved cases do not prescribe a final rehabilitation solution. They provide a structured evidence base for comparison and discussion. Final decisions still require site-specific geotechnical, hydrogeological, ecological, economic, social, legal, and safety assessments. The proposed hybrid system therefore improves the knowledge-discovery stage of mine closure planning by helping practitioners move from a broad and unevenly described case repository toward a smaller, better ordered set of potentially transferable examples that can inform professional judgement and subsequent detailed assessment.

3.9. Future implementation of the Web-GIS interface

The proposed hybrid knowledge system is intended to be implemented as an interactive Web-GIS interface linked to the REMINDNET WG3 case repository. In such an interface, users would first explore the spatial distribution of cases on a map and apply structured filters such as country, mine type, operational status, ore deposit type, and best-practice category. A natural-language search box would then allow users to retrieve ranked cases within the selected subset. Retrieved cases would be displayed as standardised 'case_card' summaries, including key site metadata, rehabilitation narrative, coordinates, and assigned categories.

This interface would support both map-based browsing and query-based case discovery. For example, a practitioner could zoom to a region, filter for closed surface mines or abandoned underground mines, and then search for cases related to ecosystem restoration, cultural reuse, pit-lake development, renewable energy, or recreation. The Web-GIS component would therefore connect geographic transferability, structured filtering, and semantic retrieval in a single decision-support environment.

Future development should prioritise transparent user controls, exportable case summaries, links to supporting documents or media, and clear warnings about corpus coverage and transferability limitations. The interface should also allow new records to be added through a controlled submission workflow, with validation of coordinates, metadata completeness, and category assignments before publication. This would make the system expandable while maintaining the data quality needed for reliable filtering and retrieval.

3.10. Discussion of implications and limitations

Taken together, the results show that the current REMINDNET WG3 corpus is characterised primarily by socially oriented and low-intensity reuse patterns, with frequent multi-use combinations involving Community & culture, Non-intensive recreation, Conversion & ecosystem services, and Education & research. Comparative analyses indicate that these patterns are not evenly distributed across mine contexts, reinforcing the value of structured metadata as a first-stage retrieval filter. In the retrieval evaluation, the main benefit of the hybrid workflow lies in improving the ordering of relevant cases on the first results page, rather than in increasing recall. Zero-relevant queries under the strict category-overlap definition, together with example-level label mismatches, indicate that current performance is limited not only by ranking quality but also by corpus coverage and label completeness. These findings support the proposed hybrid knowledge system as a practical improvement over spreadsheet-based search, while also highlighting targeted data enrichment and category refinement as the next priorities for improving system performance.

Corpus coverage is a central limitation of the present system. The dataset contains 130 curated cases and reflects the current contributions of the REMINDNET network rather than a statistically representative inventory of European mine rehabilitation practices. The strong contribution from a few countries, particularly Bulgaria, means that the spatial distribution of cases should be interpreted as dataset coverage, not as evidence that rehabilitation practices are more common or more advanced in those countries. Countries represented by only one or two cases cannot support reliable country-level interpretation, and underrepresented regions may contain rehabilitation approaches that are absent from the current corpus.

This uneven coverage also affects transferability and retrieval performance. Retrieved cases can support early comparison and option screening, but their applicability to a new site remains conditional on local geology, mine type, closure status, regulatory requirements, land ownership, environmental constraints, and community priorities. Some rehabilitation pathways are also underrepresented in the database, especially rare categories such as Adventure activities,

Light industrial, and Unclassified. As a result, queries targeting these pathways have smaller relevant pools and are more sensitive to individual ranking errors. Semantic retrieval can improve discovery when relevant cases are described using different terminology, but it cannot compensate for missing, incomplete, or geographically biased source records. Future development of the repository should therefore prioritise targeted data enrichment for underrepresented countries, mine types, ore types, site statuses, and rehabilitation categories, together with more complete narrative descriptions and category annotations.

The distinction between numerical differences and statistically supported effects is important for interpreting the retrieval results. The embedding-based method achieved the best descriptive macro-average scores, but the paired query-level tests show that this advantage over filtered TF-IDF is not statistically reliable at the present corpus and benchmark size. The practical implication is therefore not to replace lexical retrieval with semantic retrieval, but to design the system so that structured filters, interpretable keyword matching, and semantic similarity can be used together. This interpretation is consistent with the decision-support purpose of the system, where robustness, transparency, and the ability to retrieve semantically related cases are all important.

The present benchmark should nevertheless be understood as an offline evaluation of retrieval behaviour within a curated corpus, rather than as a direct measure of end-user decision quality (Turpin & Scholer, 2006). The results show that ranked retrieval, especially when combined with structured filtering, improves early ordering relative to a filters-only baseline under the relevance definition adopted in this study. However, they do not yet demonstrate whether users would make better rehabilitation decisions in a live interface, or whether the same pattern would persist under richer geographic constraints, multi-field filtering, or alternative relevance judgements. The current evidence therefore supports the feasibility of a prototype implementation, while also indicating that future validation should include user-centred testing, expanded query coverage, and further refinement of category assignments.

The relevance definition used in the retrieval evaluation also requires careful interpretation. The strict category-overlap rule makes the benchmark reproducible and internally consistent, but it cannot fully represent how practitioners may judge similarity between rehabilitation cases. A case may still be useful for decision support even if it lacks one of the target labels, particularly when the narrative description indicates a related land-use function or transferable rehabilitation concept. Conversely, a case may satisfy the category rule but be less transferable in practice because of differences in regulatory setting, geotechnical conditions, land ownership, community priorities, or environmental constraints. The reported retrieval metrics should therefore be interpreted as conservative estimates under a transparent offline benchmark, not as a complete validation of practical case usefulness. Future work should include independent expert relevance judgements and user-centred testing with mine closure practitioners to assess how well category-based relevance corresponds to professional decision-making needs.

4. Conclusions

This study addresses a practical and methodological gap in post-mining knowledge management by transforming a community-curated spreadsheet of best practices into a hybrid knowledge system for mine rehabilitation and repurposing. The main contribution is not the automatic selection of a final rehabilitation solution, but improved access to comparable cases that can support early-stage option screening, conceptual planning, stakeholder discussion, and preparation for detailed technical assessment.

The REMINDNET WG3 repository was converted from 149 raw submissions into 130 curated cases with standardised metadata, validated WGS84 coordinates, a controlled 12-category multi-label vocabulary, and reproducible 'case_card' and 'retrieval_text' representations. This structure enabled a two-stage decision-support workflow in which users can first narrow the candidate set through relational filtering and then rank the remaining cases using lexical or semantic retrieval. In practical terms, this allows users to move beyond static spreadsheet browsing and identify rehabilitation examples that are more relevant to a specific site context, such as mine type, status, ore type, geography, or intended post-mining land use.

The resulting corpus shows that the current database is characterised primarily by socially oriented and low-intensity post-mining reuse patterns. Community & culture, Non-intensive recreation, Conversion & ecosystem services, and Education & research were the most prevalent categories, and co-occurrence analysis showed that these themes also define the main multi-use combinations in the dataset. Stratified analyses further indicate that mine type, status, ore type, and geography shape the thematic profile of cases, reinforcing the practical value of structured metadata as first-stage filters in a user-facing system.

The retrieval experiments show that ranked retrieval improves the first-page ordering of relevant cases compared with filters-only browsing in the current corpus. Filtered embeddings produced the highest numerical macro-average P@5, nDCG@10, and MRR@10, but paired query-level tests did not establish a statistically reliable advantage over filtered TF-IDF at the current benchmark scale of 130 cases and 25 queries. The most robust interpretation is therefore that structured filters and ranked retrieval should be used together: filters make the candidate set contextually meaningful, while lexical and semantic ranking help users identify the most relevant examples within that subset. A fused lexical-plus-dense ranking over the filtered candidate set, together with an expanded benchmark containing more filter-bearing queries, represents a logical next step.

Overall, the proposed system can help mine-rehabilitation practitioners move from a broad and unevenly described case repository toward a smaller, better ordered set of potentially transferable examples. These examples can inform early questions about land-use compatibility, environmental recovery, cultural reuse, recreation, pit-lake development,

renewable energy, public access, and further technical assessment. Future work should expand corpus coverage, refine category annotations, include expert and user-centred relevance validation, and implement the proposed Web-GIS interface with transparent filters, natural-language search, ranked case cards, and controlled data-submission workflows.

5. References

- Arabzadeh, N., Yan, X., & Clarke, C. L. A. (2021). *Predicting Efficiency/Effectiveness Trade-offs for Dense vs. Sparse Retrieval Strategy Selection* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2109.10739>
- Athanasis, N., Kalabokidis, K., Vaitis, M., & Soulakellis, N. (2009). Towards a semantics-based approach in the development of geographic portals. *Computers & Geosciences*, 35(2), 301–308. <https://doi.org/10.1016/j.cageo.2008.01.014>
- Bainton, N., & Holcombe, S. (2018). A critical review of the social aspects of mine closure. *Resources Policy*, 59, 468–478. <https://doi.org/10.1016/j.resourpol.2018.08.020>
- Cormack, G. V., Clarke, C. L. A., & Buettcher, S. (2009). Reciprocal rank fusion outperforms condorcet and individual rank learning methods. *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 758–759. <https://doi.org/10.1145/1571941.1572114>
- Crofts, K., & Phelan, L. (2025). Delphi and Deliberation: Engaging community expertise to inform post-mining land use in the Hunter Valley, Australia. *Energy Research & Social Science*, 127, 104218. <https://doi.org/10.1016/j.erss.2025.104218>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.1810.04805>
- Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1). <https://doi.org/10.1214/aos/1176344552>
- Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69(4), 211–221. <https://doi.org/10.1007/s10708-007-9111-y>
- Järvelin, K., & Kekäläinen, J. (2002). Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems*, 20(4), 422–446. <https://doi.org/10.1145/582415.582418>
- Karpukhin, V., Oguz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W. (2020). Dense Passage Retrieval for Open-Domain Question Answering. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 6769–6781. <https://doi.org/10.18653/v1/2020.emnlp-main.550>
- Keenan, J., & Holcombe, S. (2021). Mining as a temporary land use: A global stocktake of post-mining transitions and repurposing. *The Extractive Industries and Society*, 8(3), 100924. <https://doi.org/10.1016/j.exis.2021.100924>
- Keenan, P. B., & Jankowski, P. (2019). Spatial Decision Support Systems: Three decades on. *Decision Support Systems*, 116, 64–76. <https://doi.org/10.1016/j.dss.2018.10.010>
- Kim, S.-M., Choi, Y., Suh, J., Oh, S., Park, H.-D., Yoon, S.-H., & Go, W.-R. (2012). ArcMine: A GIS extension to support mine reclamation planning. *Computers & Geosciences*, 46, 84–95. <https://doi.org/10.1016/j.cageo.2012.04.007>
- Lacasta, J., Noguera-Iso, J., Béjar, R., Muro-Medrano, P. R., & Zarazaga-Soria, F. J. (2007). A Web Ontology Service to facilitate interoperability within a Spatial Data Infrastructure: Applicability to discovery. *Data & Knowledge Engineering*, 63(3), 947–971. <https://doi.org/10.1016/j.datak.2007.06.002>
- Luan, Y., Eisenstein, J., Toutanova, K., & Collins, M. (2021). Sparse, Dense, and Attentional Representations for Text Retrieval. *Transactions of the Association for Computational Linguistics*, 9, 329–345. https://doi.org/10.1162/tacl_a_00369
- Ma, X., Wu, C., Carranza, E. J. M., Schetselaar, E. M., Van Der Meer, F. D., Liu, G., Wang, X., & Zhang, X. (2010). Development of a controlled vocabulary for semantic interoperability of mineral exploration geodata for mining projects. *Computers & Geosciences*, 36(12), 1512–1522. <https://doi.org/10.1016/j.cageo.2010.05.014>
- Maina, M. M., Amin, M. S. M., & Yazid, M. A. (2014). Web geographic information system decision support system for irrigation water management: A review. *Acta Agriculturae Scandinavica, Section B — Soil & Plant Science*, 64(4), 283–293. <https://doi.org/10.1080/09064710.2014.896935>
- McIntosh, B. S., Ascough, J. C., Twery, M., Chew, J., Elmahdi, A., Haase, D., Harou, J. J., Hepting, D., Cuddy, S., Jakeman, A. J., Chen, S., Kassahun, A., Lautenbach, S., Matthews, K., Merritt, W., Quinn, N. W. T., Rodriguez-Roda, I., Sieber, S., Stavenga, M., ... Voinov, A. (2011). Environmental decision support systems (EDSS) development – Challenges and best practices. *Environmental Modelling & Software*, 26(12), 1389–1402. <https://doi.org/10.1016/j.envsoft.2011.09.009>
- Measham, T., Walker, J., Haslam McKenzie, F., Kirby, J., Williams, C., D’Urso, J., Littleboy, A., Samper, A., Rey, R., Maybee, B., Brereton, D., & Boggs, G. (2024). Beyond closure: A literature review and research agenda for post-mining transitions. *Resources Policy*, 90, 104859. <https://doi.org/10.1016/j.resourpol.2024.104859>
- Morrison-Saunders, A., McHenry, M. P., Rita Sequeira, A., Gorey, P., Mtegha, H., & Doepel, D. (2016). Integrating mine closure planning with environmental impact assessment: Challenges and opportunities drawn from African and Australian practice. *Impact Assessment and Project Appraisal*, 34(2), 117–128. <https://doi.org/10.1080/14615517.2016.1176407>
- Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3), 245–251. <https://doi.org/10.1016/j.socnet.2010.03.006>

- Pagouni, C., Pavloudakis, F., Kapageridis, I., & Yiannakou, A. (2024). Transitional and Post-Mining Land Uses: A Global Review of Regulatory Frameworks, Decision-Making Criteria, and Methods. *Land*, 13(7), 1051. <https://doi.org/10.3390/land13071051>
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3980–3990. <https://doi.org/10.18653/v1/D19-1410>
- Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5), 513–523. [https://doi.org/10.1016/0306-4573\(88\)90021-0](https://doi.org/10.1016/0306-4573(88)90021-0)
- Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., & Haklay, M. (Muki) (2017). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, 31(1), 139–167. <https://doi.org/10.1080/13658816.2016.1189556>
- Stanković, R., Vulović, N., Lilić, N., Obradović, I., Tošović, R., & Pešić-Georgiadis, M. (2016). A WebGIS Decision Support System for Management of Abandoned Mines. *Energies*, 9(7), 567. <https://doi.org/10.3390/en9070567>
- Suh, J., Kim, S.-M., Yi, H., & Choi, Y. (2017). An Overview of GIS-Based Modeling and Assessment of Mining-Induced Hazards: Soil, Water, and Forest. *International Journal of Environmental Research and Public Health*, 14(12), 1463. <https://doi.org/10.3390/ijerph14121463>
- Turpin, A., & Scholer, F. (2006). User performance versus precision measures for simple search tasks. *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 11–18. <https://doi.org/10.1145/1148170.1148176>
- Vandic, D., Nederstigt, L. J., Frasincar, F., Kaymak, U., & Ido, E. (2024). A framework for approximate product search using faceted navigation and user preference ranking. *Data & Knowledge Engineering*, 149, 102241. <https://doi.org/10.1016/j.datak.2023.102241>
- Villa, R., Gildea, N., & Jose, J. M. (2008). FacetBrowser: A user interface for complex search tasks. *Proceedings of the 16th ACM International Conference on Multimedia*, 489–498. <https://doi.org/10.1145/1459359.1459424>
- Voorhees, E. M., & Harman, D. (2000). Overview of the Sixth Text REtrieval Conference (TREC-6). *Information Processing & Management*, 36(1), 3–35. [https://doi.org/10.1016/S0306-4573\(99\)00043-6](https://doi.org/10.1016/S0306-4573(99)00043-6)
- Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, 1(6), 80–83. <https://doi.org/10.2307/3001968>
- Worden, S., Svobodova, K., Côte, C., & Bolz, P. (2024). Regional post-mining land use assessment: An interdisciplinary and multi-stakeholder approach. *Resources Policy*, 89, 104680. <https://doi.org/10.1016/j.resourpol.2024.104680>
- Yee, K.-P., Swearingen, K., Li, K., & Hearst, M. (2003). Faceted metadata for image search and browsing. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 401–408. <https://doi.org/10.1145/642611.642681>
- Zhang, L., & Zhang, Y. (2010). Interactive retrieval based on faceted feedback. *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 363–370. <https://doi.org/10.1145/1835449.1835511>
- Zhang, Y., Ogletree, A., Greenberg, J., & Rowell, C. (2015). Controlled vocabularies for scientific data: Users and desired functionalities. *Proceedings of the Association for Information Science and Technology*, 52(1), 1–8. <https://doi.org/10.1002/pra2.2015.145052010054>

SAŽETAK

Hibridni sustav znanja za potporu odlučivanju u sanaciji rudarskih prostora primjenom relacijskog filtriranja i semantičkog pretraživanja u zajednički uređivanom repozitoriju studija slučajeva

Zatvaranje rudnika i prenamjena prostora nakon završetka eksploatacije zahtijevaju alate koji stručnjacima omogućuju pristup relevantnim primjerima sanacije rudarskih prostora unatoč fragmentiranosti dostupnih podataka i iskustava. U ovom radu tablica primjera dobre prakse Radne skupine 3 (WG3) mreže REMINDNET (Recovery of Mining District Network) transformirana je u hibridni sustav znanja koji objedinjuje strukturirano relacijsko filtriranje s leksičkim i semantičkim pretraživanjem. Od ukupno 149 dostavljenih primjera izrađen je uređeni skup podataka koji obuhvaća 130 jedinstvenih studija slučajeva iz 21 države, s validiranim koordinatama u sustavu WGS84, standardiziranim metapodacima, kontroliranim višestrukim rječnikom od 12 kategorija te ponovljivim prikazima u obliku kartica slučajeva i tekstova za pretraživanje. Sustav je vrednovan primjenom 25 referentnih upita na prirodnom jeziku i definicijom relevantnosti utemeljenom na preklapanju kategorija unutar zadanih strukturnih ograničenja. U skupu podataka prevladavaju društveno usmjerene i relativno niskoutjecajne mogućnosti prenamjene prostora, osobito kategorije Zajednica i kultura, Neintenzivna rekreacija, Prenamjena i usluge ekosustava te Obrazovanje i istraživanje. Analize istodobne pojavnosti i podskupina pokazuju da tip rudnika, operativni status, vrsta mineralne sirovine i geografski položaj oblikuju tematske profile studija slučajeva, što potvrđuje njihovu primjenjivost kao filtera u sustavima potpore odlučivanju. U vrednovanju učinkovitosti pretraživanja metode temeljene na frekvenciji pojavljivanja pojmova i njihovoj inverznoj frekvenciji u dokumentima (TF-IDF), kao i rangiranje temeljeno na semantičkim vektorskim reprezentacijama, ostvarile su bolje rezultate od slučajnog pristupa koji koristi isključivo filtriranje, potvrđujući da rangirano pretraživanje donosi dodatnu vrijednost nakon strukturiranog filtriranja. Pristup temeljen na filtriranim semantičkim vektorskim reprezentacijama

ostvario je najbolje prosječne rezultate ranog rangiranja, uključujući Precision at 5 ($P@5 = 0,768$) i normalizirani diskontirani kumulativni dobitak pri 10 rezultata ($nDCG@10 = 0,810$), no testovi na razini pojedinačnih upita nisu pokazali statistički pouzdanu prednost u odnosu na filtrirani TF-IDF pri trenutačnoj veličini referentnog skupa. Najizraženija potvrđena prednost odnosi se na korist strukturiranog filtriranja unutar postupka semantičkog pretraživanja. Ukupno gledano, relacijsko-semantička arhitektura omogućuje jednostavnije pretraživanje, usporedbu i tumačenje zajednički prikupljenih i uređenih znanja o sanaciji rudarskih prostora.

Ključne riječi: sanacija rudarskih prostora, prenamjena prostora nakon eksploatacije, sustav potpore odlučivanju, semantičko pretraživanje, zajednički uređivana baza podataka

Acknowledgment

The authors acknowledge COST Action CA22138, Recovery of Mining District Network (REMINDNET), for providing the collaborative framework and network support for the development, discussion, and refinement of the best-practices case repository used in this study. The authors also thank the REMINDNET members who contributed to the collection and sharing of mine rehabilitation case information.

Funding

This work was carried out as part of COST Action CA22138, Recovery of Mining District Network (REMINDNET), supported by COST, European Cooperation in Science and Technology. No additional external funding was received for this specific study.

Author contribution

Vancho Adjiski (Assoc. Prof. Dr.): conceptualization, methodology, data curation, formal analysis, software, investigation, visualization, writing, original draft preparation, review and editing, supervision and project coordination. **Jindřich Šancer** (Assoc. Prof. Dr.): conceptualization, mining engineering interpretation, validation, writing, review and editing. **Peter Schwendner** (Prof. Dr.): methodology, statistical validation, retrieval-evaluation interpretation, writing, review and editing. **Luka Crnogorac** (Assist. Prof. Dr.): mining rehabilitation interpretation, validation, writing, review and editing. **Kemajl Zeqiri** (Prof. Dr.): post-mining rehabilitation interpretation, regional case validation, writing, review and editing. **Václav Zubíček** (Assoc. Prof. Dr.): mining engineering expertise, methodological validation, writing, review and editing. **Stefan Moellerherm** (Prof. Dr.): post-mining knowledge-system interpretation, decision-support framing, validation, writing, review and editing.

All authors have read and agreed to the published version of the manuscript.