Entropy of Void at QuantumCLEF 2026: Hardware-Aware QUBO Decomposition via Quantum Annealing - Instance Selection and Document Clustering

Gaurav Jain[1,*,†], Hitesh Hitesh[1,*,†], Srishty Bedi[1,†] and Sparsh Sundriyal[1,†]

[1]*SBO Bangalore*

Abstract

We present team *Entropy of Void*'s submissions to two QuantumCLEF 2026 tracks: Task 2 (noise-robust instance selection for text classification) and Task 3A (document clustering for retrieval with quantum K-Medoids). Both tasks are modeled as combinatorial optimization under resource constraints and addressed with Quadratic Unconstrained Binary Optimization (QUBO) formulations solved using D-Wave quantum annealing, compared against simulated annealing (SA) baselines. For Task 2, we propose a multi-objective QUBO that combines class-aware diversity, boundary preservation, and kNN-based noise scores, plus a k-medoids decomposition to fit current QPU topology. On 5-fold NYT Vader sentiment data with 20% label noise, QA/SA retain about 40% of training instances while matching full-data validation performance (avg Macro-F1: 0.9484) on training data and (avg Macro F1: 63.9(std: 6.7)) in official evaluation. For Task 3A, we use a three-stage hierarchical pipeline (MiniBatchKMeans compression, QPU K-Medoids on representatives, full-corpus assignment) over the 6,486-document ANTIQUE corpus. Across 15 variants and $K \in \{10, 25, 50\}$, QPU outperforms SA in 5 variants officially, with strongest gains at $K = 10$ (best nDCG@10 = 0.6363, highest among participating teams at this K). Jointly, results indicate that current annealers are effective for medium-scale QUBO subproblems, with performance sensitivity to embedding density, chain breaks, and decomposition design.

Keywords

quantum annealing, QUBO, instance selection, K-Medoids, document clustering, information retrieval

1. Introduction

QuantumCLEF evaluates practical quantum optimization workflows for information retrieval and text analytics [1, 2]. This paper combines our two team submissions in a single working note, as required for final proceedings: Task 2 (instance selection under noisy labels) and Task 3A (quantum clustering for retrieval).

Both tasks reduce to NP-hard subset selection. In Task 2, selecting a compact and noise-robust training subset is a combinatorial problem akin to prototype selection and dataset reduction in supervised learning [3]. In Task 3A, K-Medoids is NP-hard and requires selecting K representative documents from large corpora [4, 5]. We adopt a unified strategy: formulate each task as QUBO [6, 7], solve via quantum annealing [8, 9], and benchmark against SA [10, 11].

Main outcomes. For Task 2, QA and SA match full-data validation Macro-F1 while reducing training size by approximately 60%. For Task 3A, QPU achieves best official nDCG@10 of 0.6363 at $K = 10$ and beats SA in 5/15 variants, while SA is more stable at larger K where embedding pressure and chain breaks increase.

CLEF 2026 Working Notes, 21 – 24 September 2026, Jena, Germany
*Corresponding author.
† All authors contributed equally.
g.jain3@shell.com (G. Jain); hitesh.hitesh@shell.com (H. Hitesh); Srishty.Bedi@shell.com (S. Bedi); Sparsh.Sundriyal@shell.com (S. Sundriyal)
0009-0006-9451-0061 (G. Jain); 0000-0001-9061-4388 (H. Hitesh); 0009-0005-8740-2911 (S. Bedi); 0009-0003-7983-5704 (S. Sundriyal)

2. Related Work

Instance selection and noisy labels. Classical approaches include edited nearest-neighbor [12], prototype selection taxonomies [3], and modern noisy label mitigation such as confident learning [13] and co-teaching [14]. Our Task 2 contribution differs by casting instance selection as a constrained QUBO with explicit diversity and boundary preservation terms.

QUBO and quantum optimization in ML. QUBO provides a common discrete optimization interface for Ising hardware [7, 6]. Prior work has explored QUBO in feature selection and clustering [15, 16]. Practical annealing workflows often require decomposition or hybridization on NISQ hardware [17, 18].

Clustering for retrieval. K-Medoids (PAM/FastPAM) remains a robust clustering baseline [4, 5], while MiniBatchKMeans provides scalable pre-clustering [19]. For retrieval evaluation we follow nDCG@10 [20] on ANTIQUE [21].

3. Background: QUBO and D-Wave Constraints

3.1. QUBO Formulation and Ising Equivalence

A binary QUBO objective seeks

$$\mathbf{x}^* = \arg\min_{\mathbf{x}\in\{0,1\}^n} E(\mathbf{x}), \quad E(\mathbf{x}) = \mathbf{x}^\top Q\mathbf{x} = \sum_i Q_{ii}x_i + \sum_{i<j} Q_{ij}x_i x_j. \tag{1}$$

The diagonal entries Q_{ii} act as linear biases on individual variables; the off-diagonal entries Q_{ij} encode pairwise interactions. D-Wave hardware solves the equivalent Ising problem

$$H = \sum_i h_i\,\sigma_i^z + \sum_{i<j} J_{ij}\,\sigma_i^z\sigma_j^z, \quad \sigma_i^z \in \{-1,+1\}, \tag{2}$$

obtained from Equation 1 via the substitution $x_i = (1-\sigma_i^z)/2$. Substituting into the diagonal terms gives $Q_{ii}x_i = Q_{ii}(1-\sigma_i^z)/2$, and into each off-diagonal term gives $Q_{ij}x_ix_j = Q_{ij}(1-\sigma_i^z-\sigma_j^z+\sigma_i^z\sigma_j^z)/4$. Because Q is upper-triangular, variable i enters the sum *twice*: as the first index of every pair (i,j) with $j>i$, and as the second index of every pair (j,i) with $j<i$. Collecting the σ_i^z coefficient from all three sources— diagonal ($-Q_{ii}/2$), row-i off-diagonal ($-Q_{ij}/4$ for $j>i$), and column-i off-diagonal ($-Q_{ji}/4$ for $j<i$)—yields:

$$h_i = -\frac{Q_{ii}}{2} - \frac{1}{4}\left(\sum_{j>i} Q_{ij} + \sum_{j<i} Q_{ji}\right), \qquad J_{ij} = \frac{Q_{ij}}{4}, \quad i<j. \tag{3}$$

This mapping is standard [7, 6]; the shorthand $-\frac{1}{4}\sum_{j\neq i} Q_{ij}$ found in some references assumes a *symmetric* Q, whereas for upper-triangular Q the column-i contributions Q_{ji} ($j<i$) must be counted separately as shown above. Note the negative sign on h_i: minimising $E(\mathbf{x})$ in QUBO space corresponds to minimising H in Ising space, since higher Q_{ii} raises the diagonal penalty and therefore decreases h_i (making that spin prefer $\sigma_i^z = +1$, i.e. $x_i = 0$, deselecting that variable). The quantum annealer is initialised in the ground state of a transverse-field Hamiltonian $H_{\text{init}} = -\Gamma\sum_i \sigma_i^x$ (a uniform superposition over all 2^n spin states) and interpolates to H_{final} over anneal time T_a. If T_a is long relative to the inverse spectral gap, the adiabatic theorem guarantees convergence to the ground state of H_{final}; in practice T_a is set to 20–2000 μs, providing a heuristic rather than provable advantage.

3.2. Hardware Constraints and Decomposition Necessity

D-Wave Advantage implements the Pegasus P16 topology: 5,627 physical qubits each connected to at most 15 neighbours [22, 23]. Because QUBO problems with fully connected interaction graphs

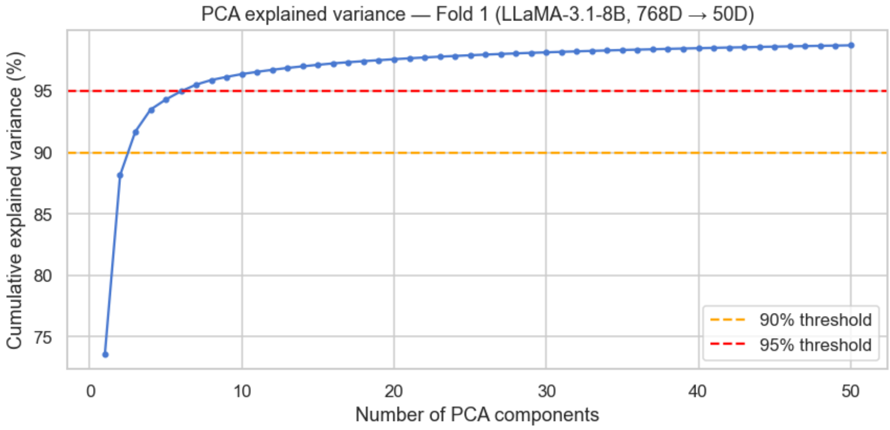

Figure 1: PCA explained-variance curve for Fold 1 (LLaMA 3.1 8B, 768D → 50D).

require *minor embedding*—representing each logical variable as a chain of physical qubits—the effective problem capacity is far below 5,627 logical variables. For a fully connected graph on n logical nodes, the $n(n-1)/2$ couplers required grow quadratically; in practice, problems with $n \gtrsim 150$ began to fail embedding in our experiments. Chain breaks arise when physical qubits within the same logical chain disagree after annealing; the chain strength λ (a ferromagnetic coupler penalty keeping chains aligned) must satisfy $\lambda > \max_{i,j} |Q_{ij}|$ to suppress breaks, but larger λ also reduces the effective energy gradient driving the optimisation [24].

This hardware reality is central to both tasks. Task 2 decomposes the global instance-selection problem into cluster-local QUBOs with an average of 20 variables for QA or 100 for SA each, making every subproblem trivially embeddable. Task 3A compresses 6,486 documents to $G \leq 80$ macro-clusters and selects 2 representatives per cluster (nearest and farthest from the centroid), before QUBO construction, limiting the logical graph to $2G \leq 160$ nodes. In both cases, decomposition is not an optimisation choice but a hardware necessity.

4. Task 2: Noise-Robust Instance Selection

4.1. Task and Data

Task 2 aims to select a compact training subset from noisy sentiment data while preserving classification quality. We use NYT Vader-based sentiment data split into 5-fold for Cross Validation (CV) and 20% injected label noise. Embeddings are generated from LLaMA 3.1 8B [25] and reduced from 768D to 50D via Principal Component Analysis (PCA) [26]. As shown in Figure 1, 50 components capture ≈98% of variance, with the curve flattening sharply after the first 8–10 components; this indicates that most semantic structure is encoded in a low-dimensional subspace, so reducing to 50D retains almost all discriminative signal while reducing QUBO matrix density by a factor of $(768/50)^2 \approx 236\times$. Noise is estimated with a kNN disagreement score ($k = 15$), following the intuition that mislabeled points are locally inconsistent.

4.2. QUBO Formulation

We introduce a binary variable $x_i \in \{0, 1\}$ for each instance, where $x_i = 1$ means instance i is selected. The QUBO objective is:

$$\min_{\mathbf{x}} \ \mathbf{x}^\top Q \mathbf{x} \tag{4}$$

where $Q \in \mathbb{R}^{n \times n}$ is an upper-triangular coefficient matrix defined as follows.

Diagonal terms (self-cost). For each instance i with label ℓ_i:

$$Q_{ii} = \underbrace{\frac{\text{count}(\ell_i)}{N}}_{\text{class-frequency bias}} + \underbrace{\lambda \cdot \eta_i}_{\text{noise penalty}} \tag{5}$$

The class-frequency term biases selection toward minority-class instances (which have a smaller frequency ratio), promoting class balance. The noise penalty $\eta_i \in [0, 1]$ is a per-instance score, weighted by hyperparameter λ.

We estimate per-instance label reliability using k-nearest-neighbour (k-NN) consistency [12]:

$$\eta_i = 1 - \frac{|\{j \in \mathcal{N}_k(i) : \tilde{y}_j = \tilde{y}_i\}|}{k} \tag{6}$$

where $\mathcal{N}_k(i)$ denotes the k nearest neighbours of instance i (excluding i itself). Instances in homogeneous neighbourhoods receive $\eta_i \approx 0$, while likely mislabelled instances on the "wrong side" of the class boundary receive $\eta_i \approx 1$. We use $k = 15$ throughout all experiments.

Off-diagonal terms (pair-cost). For each pair (i, j) with $i < j$, let $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2 / d_{\max}$ be the normalised Euclidean distance:

$$Q_{ij} = \begin{cases} -d_{ij} & \text{if } \ell_i = \ell_j \quad \text{(diversity reward)} \\ +d_{ij} & \text{if } \ell_i \neq \ell_j \quad \text{(boundary preservation)} \end{cases} \tag{7}$$

Rewarding distant same-class pairs encourages diversity within each class. Penalising distant cross-class pairs retains informative decision-boundary instances.

Cardinality constraint. We enforce the selection of exactly $k = \lfloor \rho N \rfloor$ instances (with target ratio ρ) via a quadratic penalty added to the BQM:

$$\gamma \left(\sum_i x_i - k \right)^2 \qquad \gamma = \Delta E_{\max} \tag{8}$$

where γ is set to the maximum energy gap of the unconstrained BQM, ensuring the constraint dominates the objective [7].

4.3. Decomposition and Solvers

To fit QPU limits, we partition points with k-medoids and solve cluster-local QUBOs. We use target cluster size 20 for QA and 100 for SA, with global selection ratio $\rho = 0.40$ and noise weight $\lambda = 0.50$. QA runs on Advantage_system4.1 with 50 reads and 20 μs anneal time; SA uses `dwave-neal` with 1000 reads. Figure 2 shows the PCA projection for Training Fold 1 alongside the QA subproblem size distribution; the majority of k-medoids clusters contain instances from both classes, enabling effective boundary-aware QUBO construction as described in Equation 5–(7) above.

4.4. Hyperparameter Tuning

To assess whether more QPU resources yield improved results, we ran a second QA configuration using `num_reads`$= 100$ and `annealing_time`$= 50\,\mu$s. As observed from the official QCLEF workspace statistics, the selected subset and resulting Macro-F1 were comparable to those obtained under the default configuration (`num_reads`$= 50$, `annealing_time`$= 20\,\mu$s) in respect to energy minimization and QPU processing time, confirming that for QUBO subproblems of size ≤ 20 variables on Pegasus, a moderate number of reads is sufficient to find high quality solutions.

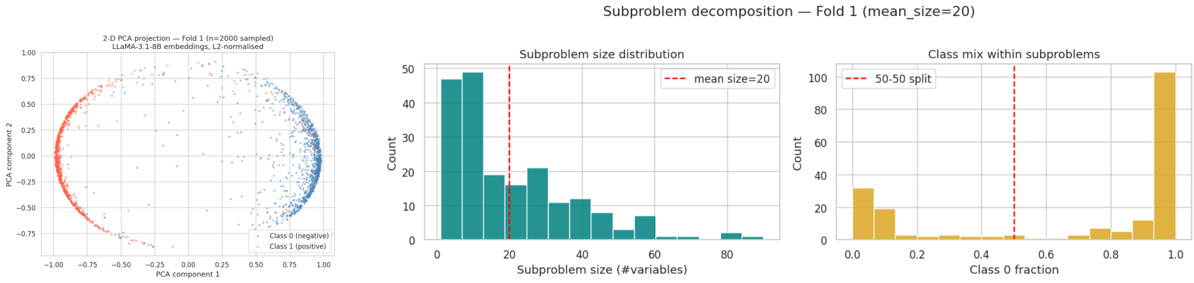

Figure 2: Left: PCA projection of LLaMA 3.1 8B embeddings (768D $\rightarrow$ 50D) coloured by sentiment class across one training fold. Right: Subproblem size distribution and class balance within k-medoids clusters (Fold 1, `target_cluster_size`$= 20$).

Table 1
Task 2: 5-fold cross-validation on NYT Vader (20% injected noise), using Logistic Regression as a proxy classifier (see protocol above). Both QA and SA retain ∼40% of training data while matching full-data Macro-F1. Reduction $\approx 60\%$ is consistent across all folds.

Method	Macro-F1 ↑	Retention	Reduction	QPU time
Full baseline	0.9482	100%	0%	—
SA-selected subset	0.9484	∼40%	∼60%	—
QA-selected subset	0.9484	∼40%	∼60%	<1% of 160 s budget

4.5. Task 2 Results

Internal validation protocol. Since fine-tuning the full LLaMA 3.1 8B model per fold was not feasible within our compute budget, we use Logistic Regression as a proxy classifier to evaluate the quality of QA- and SA-selected subsets prior to official submission. For each of the 5 folds, the fold's embeddings are split 80%/20% (stratified by class, fixed seed) into a training portion and a held-out validation portion that is never touched by the selection process. From the 80% training portion we retain only the instances selected by QA or SA (or all instances, for the baseline), train a Logistic Regression classifier (ℓ_2-regularised, `lbfgs` solver, max 1000 iterations) on the retained subset, and evaluate Macro-F1 on the untouched validation portion. Table 1 reports Macro-F1 averaged across all 5 folds. This proxy evaluation is used only to sanity-check subset quality internally; the official QCLEF evaluation (Table 2) instead fine-tunes LLaMA 3.1 on the submitted subsets, which is why the two tables are not directly comparable.

- Internal validation (Table 1) shows that QA and SA match full-data performance within 0.02 Macro-F1 points while discarding ∼60% of training data.
- Official test-set results (Table 2) demonstrate one of the most aggressive data reduction rates (∼60%) among participating teams, while maintaining competitive Macro F1 scores (63.9% QA, 65.9% SA).
- Our SA achieves the highest Macro F1 among methods using >50% data reduction, preserving predictive quality under substantial filtering.
- Our QA achieves a good F1 score with the lowest standard deviation (6.7), indicating strong robustness across folds.
- While a few competing methods achieve higher absolute F1 scores, they do so with significantly lower reduction rates, highlighting the trade-off between data efficiency and predictive accuracy.
- The validation–test gap suggests opportunities for further optimisation and improved generalisation.

As shown in Figure 3, both solvers closely track the 40% target selection ratio across folds, and selected instances consistently have lower k-NN noise scores than discarded instances, confirming that the QUBO noise penalty $\lambda \cdot \eta_i$ effectively filters mislabelled instances.

Table 2
Task 2: Official QCLEF 2026 test-set results [27] (Macro-F1, mean $\pm$ std across folds). Both QA and SA underperform the official full-data baselines on the held-out test set; the gap reflects distribution shift between the noisy training folds and the clean test evaluation.

Method	Official Macro-F1	$\pm$ std
Full baseline (official)	88.9	0.8
Alternative baseline (official)	84.3	2.5
QA-selected (our submission)	63.9	6.7
SA-selected (our submission)	65.9	10.0

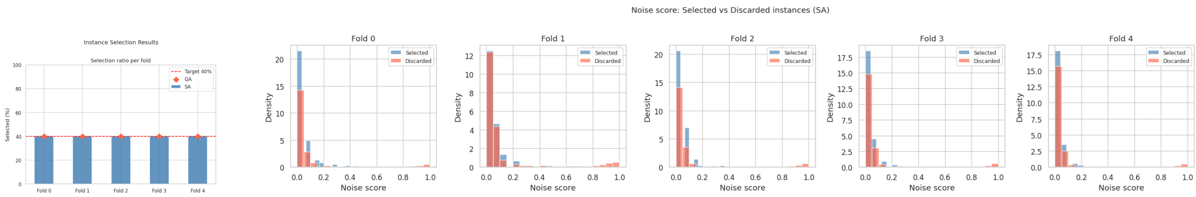

Figure 3: Left: actual selection ratio vs. the 40% target (dashed line). Right: Noise score density for selected (blue) vs. discarded (red) instances under SA selection.

5. Task 3A: Quantum K-Medoids for Retrieval Clustering

5.1. Task and Data

Task 3A clusters 6,486 ANTIQUE documents for retrieval routing and ranking, with quality measured by nDCG@10 over 50 queries and official labels [21, 20].

5.2. Hierarchical Pipeline

Direct Bauckhage QUBO at full corpus size is intractable. We use:

1. MiniBatchKMeans pre-clustering into G groups and selection of 2 representatives per group (nearest and farthest from the centroid),
2. QPU K-Medoids on $2G$ representatives,
3. full-corpus assignment of all documents to the nearest selected medoid.

We run 15 variants over $K \in \{10, 25, 50\}$, varying G, distance metric (Euclidean/cosine/correntropy), and PCA use.

5.3. QUBO Construction: Bauckhage K-Medoids

Following [16], binary variable $x_i = 1$ means representative i is selected as a medoid. The QUBO matrix over the $2G$ representatives is

$$Q_{ii} = \beta \sum_j \Delta_{ij} + \gamma(1 - 2K), \tag{9}$$

$$Q_{ij} = \gamma - \tfrac{\alpha}{2}\,\Delta_{ij}, \quad i \neq j, \tag{10}$$

where the correntropy loss [28]

$$\Delta_{ij} = 1 - e^{-d_{ij}/2} \ \in \ [0, 1] \tag{11}$$

bounds all pairwise interactions to the unit interval regardless of embedding scale or dimensionality, and the parameters are $\alpha = 1/K$, $\beta = 1/(2G)$, $\gamma = 2$.

The diagonal term Q_{ii} (Equation 9) penalises outlier documents: a high $\sum_j \Delta_{ij}$ (far from all others) adds a positive $\beta \sum_j \Delta_{ij}$ to Q_{ii}, making it *less negative* and therefore *less preferred* as a medoid when

Table 3
Best official nDCG@10 per K block for Task 3A and comparison with same-variant SA.

Setting	Best QPU nDCG@10	Variant	Δ(QPU-SA)
$K = 10$	0.6363	V1 (G=80, Euclidean)	+0.101
$K = 25$	0.5858	V9 (G=80, Cosine)	+0.035
$K = 50$	0.5466	V12 (G=70, Euclidean)	-0.009

minimising the QUBO energy. Central documents (low $\sum_j \Delta_{ij}$) receive a more negative Q_{ii} and are thus preferred—consistent with K-Medoids minimising total point-to-nearest-medoid distance. The cardinality offset $\gamma(1-2K)$ ensures the net Q_{ii} is always large and negative, arising from the expansion $\gamma(\sum_i x_i - K)^2 = \gamma \sum_i x_i - 2\gamma K \sum_i x_i + 2\gamma \sum_{i<j} x_i x_j + \text{const}$ (the same quadratic cardinality penalty used in Task 2). The off-diagonal term Q_{ij} (Equation 10) balances a global coverage reward (γ) against a co-selection penalty ($\frac{\alpha}{2}\Delta_{ij}$) that discourages choosing two near-identical documents as co-medoids.

Chain strength is set as $\lambda = 1.2 \times \max_{i<j} |Q_{ij}|$, where the maximum is taken over the *off-diagonal* entries only. The diagonal entries $Q_{ii} \approx \gamma(1-2K)$ carry the linear bias (not the quadratic coupling), so including them would inflate λ by a factor of K and suppress the energy landscape on QPU. The off-diagonal entries lie in $[2\gamma - \alpha,\ 2\gamma] = [4 - 1/K,\ 4]$ (the factor of 2 arises from the upper-triangular QUBO convention), so $\lambda \approx 1.2 \times 2\gamma = 1.2 \times 4 \approx 4.8$, essentially *constant* across all K. Critically, λ is also independent of G: because $\Delta_{ij} \in [0, 1]$, adding more representatives does not change $\max |Q_{ij}|$, so no retuning is needed as G varies—a direct practical benefit of the correntropy bounding that would not hold for raw Euclidean distance (where $\max |Q_{ij}|$ grows with G).

5.4. Hardware Behaviour: Chain Breaks

Across all 15 variants, chain break fractions are high (0.74–0.89), substantially above typical QUBO experiment reports. Three structural factors contribute: (1) at $G = 80$ the minor embedding of 160 logical variables demands long physical chains; (2) the correntropy-bounded energy landscape is relatively flat near the optimum ($\max |Q_{ij}| \approx 4$ in the upper-triangular QUBO passed to D-Wave), so the chain-alignment penalty ($\lambda \approx 4.8$) provides only $\approx 20\%$ headroom over the largest coupler values, which is insufficient to reliably suppress chain breaks; (3) chain break rates increase with K (from ∼0.77 at $K = 10$ to ∼0.88 at $K = 50$) as the number of active couplers grows. Despite these rates, post-processing (`reduce_medoids_to_k`) ensures exactly K medoids are returned from each restart; the best-of-restart strategy over three × 1,000 reads is sufficient to locate a cardinality-consistent solution. Embedding entirely fails at $G = 120$ ($N_{\text{reps}} = 240$, 28,680 couplers), setting the practical ceiling at $G = 80$ on Pegasus P16 [22].

5.5. Task 3A Results

QPU outperforms SA in 5/15 official variants, concentrated at lower-to-mid K. QPU also shows lower annealing-time scale in official logs, but quality advantage narrows as K grows and chain-break effects intensify [27].

Figure 4 illustrates the best Task 3A result (V1, $K = 10$, $G = 80$, Euclidean, nDCG@10 = 0.6363): the QPU panel shows 10 clusters spanning all major regions of the embedding space, with cluster sizes ranging from ∼200 to ∼1,200 documents, reflecting natural density variation in the corpus.

Figure 5 compares Silhouette, Calinski-Harabasz, and local Query Coherence across all five methods for V1. Ward achieves the highest Query Coherence (0.830), followed by K-Means (0.741); QPU ranks third (0.601), ahead of PAM (0.562) and SA (0.540). K-Means and Ward also dominate on internal geometry metrics (Silhouette, Calinski-Harabasz), yet QPU achieves the highest official nDCG@10 (0.6363) despite ranking fourth on Silhouette (0.0126) behind K-Means (0.028), Ward (0.016), and PAM (0.0133)—suggesting that internal geometry metrics may not be reliable proxies for official retrieval quality in this corpus.

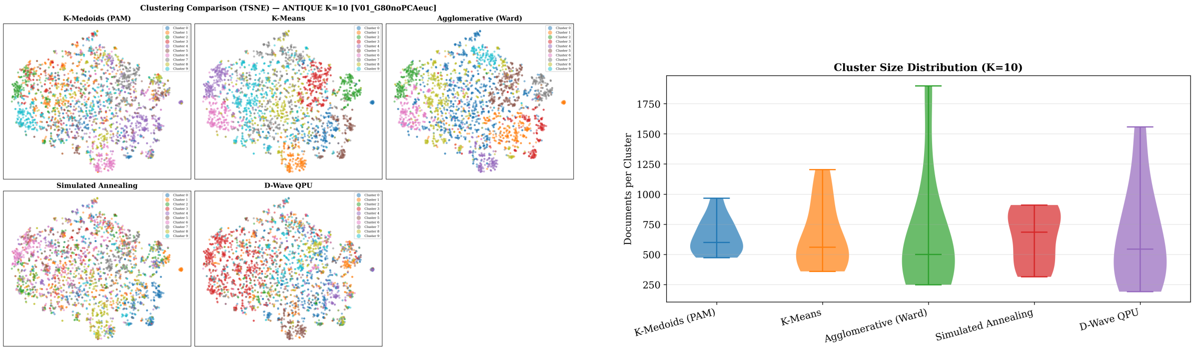

Figure 4: Task 3A best result: V1 ($K = 10$, $G = 80$, Euclidean, nDCG@10 = 0.6363). Left: t-SNE projection of a 3,000-document random subsample coloured by cluster assignment, shown side-by-side for all five methods (PAM, K-Means, Ward, SA, D-Wave QPU). Right: violin plot of cluster-size distributions across all five methods.

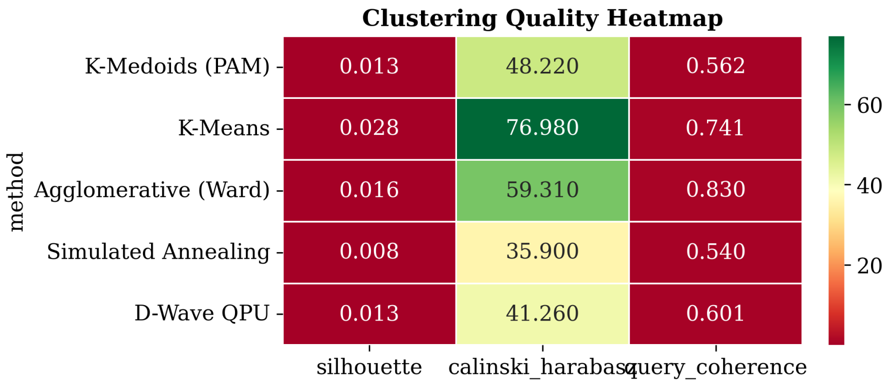

Figure 5: Method-comparison heatmap for V1 ($K = 10$, $G = 80$, Euclidean), showing Silhouette, Calinski-Harabasz, and local Query Coherence scores across QPU, SA, PAM, K-Means, and Ward clustering (colour scale spans the global data range; annotations show exact values).

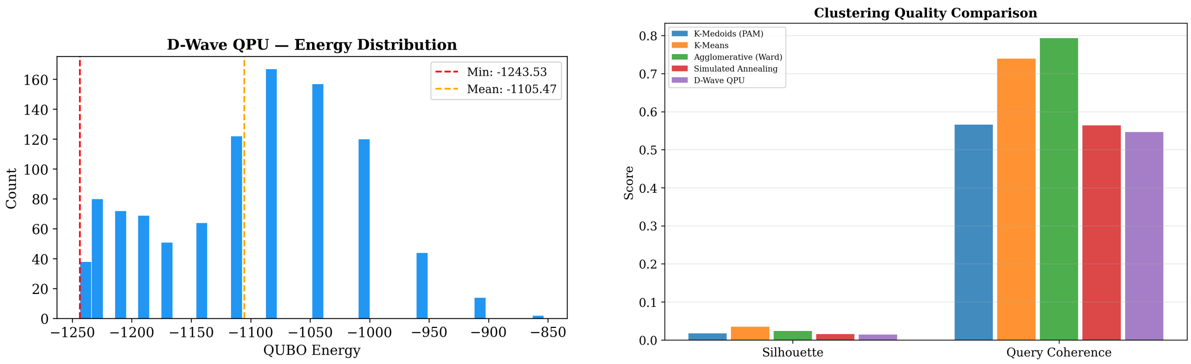

Figure 6: Task 3A hardware behaviour: V6 ($K = 25$, $G = 80$, Euclidean). Left: QPU QUBO energy distribution across 1,000 reads from the best-energy restart. Right: Silhouette and Query Coherence scores across all methods at $K = 25$.

Figure 6 shows V6 ($K = 25$, $G = 80$, Euclidean) hardware behaviour: the QPU energy distribution (best of 3 restarts $\times$ 1,000 reads) shows a wide spread reflecting the high chain break fraction ($\sim$0.84), yet post-processing (`reduce_medoids_to_k`) still ensures exactly K medoids from the best-energy read. At $K = 25$, SA (red) slightly exceeds QPU (purple) on both Silhouette and Query Coherence (SA QC$= 0.564$ vs. QPU QC$= 0.547$; SA Sil$= 0.015$ vs. QPU Sil$= 0.014$), illustrating the narrowing QPU advantage at larger K under high chain-break pressure.

6. Cross-Task Discussion

A shared pattern emerges across both tasks:

- **Decomposition is essential.** Hardware-feasible subproblems are the dominant design lever for both quality and runtime.
- **QPU gains are regime-dependent.** We see strong wins on medium-scale, well-structured subproblems (Task 2 local QUBOs, Task 3A at $K \leq 25$), but weaker gains under denser embeddings.
- **Metric alignment matters.** In Task 3A, cosine assignment improves official nDCG at $K = 25$ (V9 is the best $K = 25$ variant); at $K = 10$ Euclidean assignment achieves the best score (V1).
- **Hybrid robustness remains important.** SA is a strong baseline and can dominate in larger or noisier regimes, suggesting practical quantum pipelines should remain hybrid and adaptive. We note that this practical advantage of hybrid approaches largely reflects the current Noisy Intermediate-Scale Quantum (NISQ) era, in which annealers have a limited number of qubits and are heavily affected by chain breaks and embedding overhead; as hardware connectivity and coherence improve, this balance may shift toward fully quantum approaches becoming more competitive at larger problem scales.

7. Conclusion

This combined team note reports two QuantumCLEF 2026 submissions under a single methodological frame: QUBO-based discrete optimization with hardware-aware decomposition. For Task 2, we obtain near-baseline classification quality with around 60% data reduction. For Task 3A, we achieve the strongest official score at $K = 10$ while characterizing a clear scalability boundary at larger K under current NISQ constraints. Future work includes adaptive chain-strength calibration, decomposition guided by embedding quality indicators, and evaluation on newer annealing topologies with improved connectivity.

Acknowledgments

The authors thank the QuantumCLEF 2026 organizers for datasets, infrastructure, and evaluation support. QPU access was provided through the competition environment.

Declaration on Generative AI

During the preparation of this work, the author(s) used GitHub Copilot (Claude Sonnet 4.6) to assist with grammar and spelling checks, editing of LaTeX source, and general code improvements. After using these tools, the author(s) reviewed and edited all content as necessary and take full responsibility for the final publication.

References

[1] A. Pasin, M. F. Dacrema, W. Cunha, M. A. Gonçalves, P. Cremonesi, N. Ferro, Overview of quantumclef 2026: The third quantum computing challenge for information retrieval and recommender systems at CLEF, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction - 17th International Conference of the CLEF Association, CLEF 2026, Jena, Germany, September 21-24, 2026, Proceedings, Lecture Notes in Computer Science, Springer, 2026.

[2] A. Pasin, M. F. Dacrema, W. Cunha, M. A. Gonçalves, P. Cremonesi, N. Ferro, Quantumclef 2026: Overview of the third quantum computing challenge for information retrieval and recommender systems at CLEF, in: Working Notes of the Conference and Labs of the Evaluation Forum, CLEF 2026, Jena, Germany, 21-24 September 2026, CEUR Workshop Proceedings, CEUR-WS.org, 2026.

[3] S. Garcia, J. Derrac, J. R. Cano, F. Herrera, Prototype selection for nearest neighbor classification: Taxonomy and empirical study, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (2012) 417–435.
[4] L. Kaufman, P. J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, John Wiley & Sons, 1990.
[5] E. Schubert, P. J. Rousseeuw, Faster k-medoids clustering: Improving the PAM, CLARA, and CLARANS algorithms, Similarity Search and Applications (SISAP) (2019) 171–187.
[6] F. Glover, G. Kochenberger, Y. Du, Quantum bridge analytics I: A tutorial on formulating and using QUBO models, 4OR 17 (2019) 335–371.
[7] A. Lucas, Ising formulations of many NP problems, Frontiers in Physics 2 (2014) 5.
[8] T. Kadowaki, H. Nishimori, Quantum annealing in the transverse Ising model, Physical Review E 58 (1998) 5355–5363.
[9] E. Farhi, J. Goldstone, S. Gutmann, M. Sipser, Quantum computation by adiabatic evolution, arXiv preprint quant-ph/0001106 (2000).
[10] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, Optimization by simulated annealing, Science 220 (1983) 671–680.
[11] D-Wave Systems, dwave-neal: Simulated Annealing sampler for QUBO models, GitHub Repository (2021). https://github.com/dwavesystems/dwave-neal.
[12] D. L. Wilson, Asymptotic properties of nearest neighbor rules using edited data, IEEE Transactions on Systems, Man, and Cybernetics 2 (1972) 408–421.
[13] C. G. Northcutt, L. Jiang, I. L. Chuang, Confident learning: Estimating uncertainty in dataset labels, Journal of Artificial Intelligence Research 70 (2021) 1373–1411.
[14] B. Han, Q. Yao, X. Yu, G. Niu, M. Xu, W. Hu, I. Tsang, M. Sugiyama, Co-teaching: Robust training of deep neural networks with extremely noisy labels, Advances in Neural Information Processing Systems 31 (2018).
[15] A. Milne, M. Rounds, P. Goddard, Optimal Feature Selection in Credit Scoring and Classification Using a Quantum Annealer, White Paper, 1QBit, 2017.
[16] C. Bauckhage, N. Piatkowski, R. Sifa, D. Hecker, S. Wrobel, A QUBO formulation of the k-medoids problem, in: Lernen, Wissen, Daten, Analysen (LWDA), volume 2454 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 54–63. URL: https://ceur-ws.org/Vol-2454/paper_39.pdf.
[17] D-Wave Systems, D-Wave Hybrid Solver Service + Advantage: Technology Overview, Technical Report, D-Wave Systems Inc., 2020.
[18] F. Neukart, G. Compostella, C. Seidel, D. von Dollen, S. Yarkoni, B. Parney, Traffic flow optimization using a quantum annealer, in: Frontiers in ICT, volume 4, 2017, p. 29.
[19] D. Sculley, Web-scale k-means clustering, in: Proceedings of the 19th International World Wide Web Conference (WWW), 2010, pp. 1177–1178.
[20] K. Järvelin, J. Kekäläinen, Cumulated gain-based evaluation of IR techniques, ACM Transactions on Information Systems 20 (2002) 422–446.
[21] H. Hashemi, M. Aliannejadi, H. Zamani, W. B. Croft, ANTIQUE: A non-factoid question answering benchmark, in: Advances in Information Retrieval (ECIR), 2020, pp. 166–173.
[22] K. Boothby, P. Bunyk, J. Raymond, A. Roy, Next-generation topology of D-Wave quantum processors, arXiv preprint arXiv:2003.00133 (2020).
[23] D-Wave Systems, D-Wave advantage: New QPU topology and performance, D-Wave Technical Report (2020). https://www.dwavesys.com/media/s3qyd1he/14-1049a-a_d-wave_advantage_performance_update.pdf.
[24] J. Cai, W. G. Macready, A. Roy, A practical heuristic for finding graph minors, arXiv preprint arXiv:1406.2741 (2014).
[25] A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur, A. Schelten, A. Yang, A. Fan, et al., The Llama 3 herd of models, arXiv preprint arXiv:2407.21783 (2024).
[26] K. Pearson, On lines and planes of closest fit to systems of points in space, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2 (1901) 559–572.
[27] QCLEF 2026 Organizers, QuantumCLEF 2026 official results, https://qclef.dei.unipd.it/

clef2026-results, 2026. Accessed: May 2026.
[28] W. Liu, P. P. Pokharel, J. C. Príncipe, Correntropy: Properties and applications in non-Gaussian signal processing, IEEE Transactions on Signal Processing 55 (2007) 5286–5298.