QuantumCLEF 2025: Overview of the Second Quantum Computing Challenge for Information Retrieval and Recommender Systems at CLEF

Andrea Pasin[1,*], Maurizio Ferrari Dacrema[2], Washington Cunha[3], Marcos André Gonçalves[3], Paolo Cremonesi[2] and Nicola Ferro[2]

[1]University of Padua, Padua, Italy
[2]Politecnico di Milano, Milano, Italy
[3]Federal University of Minas Gerais, Belo Horizonte, Brazil

Abstract

The emerging field of Quantum Computing (QC) is attracting considerable research interest due to its potential. It is in fact believed that QC could revolutionize the way we approach complex problems by significantly reducing the time required to solve them. Although QC is still in its early stages of development, certain problems can already be addressed using quantum computers, offering a glimpse into its capabilities.

The goal of the QuantumCLEF lab is to raise awareness of QC and to design, develop, and evaluate new QC algorithms aimed at solving challenges typically encountered in the implementation of Information Retrieval (IR) and Recommender Systems (RS). Furthermore, the lab provides a valuable opportunity to engage with QC technologies, which are often difficult to access.

In this work, we present an overview of the second edition of QuantumCLEF, a lab focused on applying Quantum Annealing (QA), a specific QC paradigm, to three tasks: Feature Selection for IR and RS systems, Instance Selection for IR systems, and Clustering for IR systems. A total of 44 teams registered for the lab, with 5 teams successfully submitting their runs in accordance with the lab guidelines. Given the novelty of the topics, participants were provided with extensive examples and comprehensive materials to help them understand how QA works and how to program quantum annealers.

Keywords

Quantum Computing, Quantum Annealing, CLEF, Information Retrieval, Recommender Systems

1. Introduction

Even though IR and RS systems have been extensively studied and refined over the years, they continue to face significant challenges. The ever-increasing volume of data and the computational complexity required to process it pose difficult problems for these systems.

To tackle these issues, researchers are now exploring QC, an emerging computing paradigm with the potential to revolutionize the way problems are solved. QC is not just a new hardware alternative: it represents a fundamental shift in how problems are approached, leveraging principles of quantum mechanics. Unlike classical computing, which uses bits that are either 0 or 1, QC employs qubits, which can exist in multiple states simultaneously due to superposition. Furthermore, qubits can be entangled, meaning the state of one can influence another, even across long distances.

These properties allow quantum computers to theoretically explore exponentially larger problem spaces, offering advantages for certain types of problems, especially complex combinatorial problems or those where quantum principles can be effectively applied. This paradigm shift holds promise for possible improvements in terms of efficiency and effectiveness of IR and RS systems. Once QC technology

CLEF 2025: Conference and Labs of the Evaluation Forum, September 09–12, 2025, Madrid, Spain
*Corresponding author.
andrea.pasin.1@phd.unipd.it (A. Pasin); maurizio.ferrari@polimi.it (M. Ferrari Dacrema); washingtoncunha@dcc.ufmg.br (W. Cunha); mgoncalv@dcc.ufmg.br (M. A. Gonçalves); paolo.cremonesi@polimi.it (P. Cremonesi); nicola.ferro@unipd.it (N. Ferro)
https://www.dei.unipd.it/~pasinandre/ (A. Pasin); https://www.dei.unipd.it/~ferro/ (N. Ferro)
0009-0007-5193-0741 (A. Pasin); 0000-0001-7103-2788 (M. Ferrari Dacrema); 0000-0002-1988-8412 (W. Cunha); 0000-0002-2075-3363 (M. A. Gonçalves); 0000-0002-1253-8081 (P. Cremonesi); 0000-0001-9219-6239 (N. Ferro)

matures enough, it could provide innovative solutions that could be integrated into traditional pipelines to boost systems' performance. However, at the moment QC is in its early stages of development. While hardware is becoming more accessible and reliable, many challenges persist, primarily related to the size of the solvable problems and the qubit fragility. In fact, present quantum computers have a limited number of qubits, which must be isolated from environmental noise (e.g., electromagnetic interference or temperature fluctuations), since it can easily break computations. In contrast, classical systems are far more robust because they have been optimized over decades.

Given this exciting context, a natural question arises: can QC help solve some of the complex tasks faced by IR and RS systems? To explore this, we launched a new CLEF lab in 2024 called QuantumCLEF [1], dedicated to developing and evaluating QC algorithms for IR and RS. The lab has four main goals:

- Develop new QC algorithms for IR and RS, and evaluate their efficiency and effectiveness compared to traditional methods;
- Create datasets and resources to support reproducibility and future research;
- Provide participants with educational materials and access to real quantum computers, which are not yet widely available;
- Raise awareness about the potential of QC and foster a research community around this field.

This paper presents an overview of the second edition of QuantumCLEF, held in 2025. Similarly to the previous 2024 edition [2, 3, 4], also this one focused on QA, a specific QC paradigm tailored for optimization problems. Participants were granted access to cutting-edge quantum annealers developed by D-Wave, a leading company in the field.

QA is more approachable than the Universal Gate-Based paradigm and is supported by a range of tools and libraries provided by D-Wave, which simplify the development process. As a result, researchers could engage with quantum technology without needing deep expertise in quantum physics, focusing instead on algorithm design and testing.

The 2025 QuantumCLEF edition featured three main tasks [5]:

- **Task 1**: Feature Selection for IR and RS;
- **Task 2**: Instance Selection for IR [6];
- **Task 3**: Clustering for IR.

Participants were invited to develop their own algorithms using both QA and Simulated Annealing (SA). SA is a classical optimization technique that shares conceptual similarities with QA. Given the novelty of the subject, we provided extensive support materials such as videos, slides, and examples to help participants understand QA and how to program quantum annealers.

To facilitate access to real quantum devices, we also relied on a dedicated infrastructure, Kubernetes Infrastructure for Managed Evaluation and Resource Access (KIMERA) [7], that simplified workflows and promoted reproducibility. A total of 44 teams registered, with 5 actively participating and submitting for our proposed tasks.

Results were in line with the previous edition, showing that QA-based and hybrid approaches performed comparably to SA and traditional methods, often with improved efficiency. These findings confirm that QA is already a practical and effective option for tackling complex optimization challenges in IR, RS, and potentially other domains. As the technology improves, QC and QA could be integrated into the current state-of-the-art traditional systems' pipelines to boost performance in terms of efficiency and/or effectiveness.

The paper is organized as follows: Section 2 discusses related works; Section 3 presents the tasks of the QuantumCLEF 2025 lab while Section 4.1 introduces the lab's setup and the design and implementation of our ad-hoc infrastructure; Section 5 shows and discusses the results achieved by the participants; finally, Section 6 draws some conclusions and outlooks some future work.

2. Related Works

In this section, we provide an overview of QA and SA, followed by a summary of the tasks and outcomes from the 2024 edition of QuantumCLEF.

2.1. Background on Quantum and Simulated Annealing

2.1.1. Quantum Annealing

QA is a QC paradigm based on specialized hardware called *quantum annealers* designed to solve optimization problems framed according to specific mathematical formulations. The core principle is to encode the problem into the energy landscape of a physical system and use quantum phenomena such as superposition, entanglement, and tunneling to guide the system toward its lowest energy state, which represents the optimal solution. This evolution is driven by the tendency of each natural system to reach its minimum energy state.

To leverage quantum annealers, a problem must first be expressed in the Quadratic Unconstrained Binary Optimization (QUBO) format [8], a standard formulation for combinatorial optimization:

$$\min \quad y = x^T Q x \tag{1}$$

Here, x is a vector of binary variables, and Q is a matrix encoding the problem, representing the relationships between the considered variables.

Before execution on quantum hardware, a crucial step known as *minor embedding* is required to map the logical problem onto the specific topology of the Quantum Processing Unit (QPU). In fact, each QPU has a fixed hardware graph, with nodes representing qubits and edges representing couplers (i.e., connections between qubits). When a logical variable needs more connections than the ones physically available, a chain of physical qubits is used. As a result, the number of physical qubits required for a problem may exceed the number of logical variables. This embedding step is an *NP*-hard problem typically handled by heuristics [9]. When problems exceed the capacity of the QPU, D-Wave provides a Hybrid (H) approach that decomposes them into sub-problems solved via a hybrid classical-quantum method.

Constraints can be integrated into the objective function using penalty terms $P(x)$ [10], leading to the following formulation:

$$\min \quad C(x) = y + P(x) \tag{2}$$

These penalties act as *soft constraints*, discouraging infeasible solutions without enforcing strict exclusion. The effect of these constraints can be adjusted through hyperparameters.

Solving a problem with a quantum annealer generally involves the following pipeline [10]:

1. **Formulation**: Model the problem as a QUBO.
2. **Embedding**: Map the logical variables onto the physical architecture.
3. **Data Transfer**: Send the embedded problem to the quantum device.
4. **Annealing**: Run the annealing process, typically repeating it many times to sample a distribution of possible solutions. The best result is selected based on feasibility and optimality.

Once submitted, the actual annealing step takes only a few milliseconds, although preprocessing can take significantly longer.

2.1.2. Simulated Annealing

SA is a classical metaheuristic optimization method [11, 12, 13], capable of finding global optima even in landscapes with many local optimal solutions. Like QA, it is able to optimize cost functions that can be expressed as QUBO formulations, but it runs entirely on conventional hardware and does not require any embedding phase.

It is crucial to note that SA is not a simulation of QA done on traditional hardware: they are distinct algorithms which share only some of their aspects. However, SA can serve as a strong benchmark to compare against QA when evaluating performance and scalability on classical devices.

In the context of QuantumCLEF, access to quantum resources is limited to ensure fair usage. Therefore, SA can be used for preliminary tests to validate QUBO models without consuming quantum device time, thus understanding the validity of the proposed approaches.

2.2. QuantumCLEF 2024

The QuantumCLEF 2024 lab [2, 3, 4], presented at CLEF 2024, explored the application of QA in the fields of IR and RS. The lab was structured around two main tasks:

- **Feature Selection:** Focused on identifying the most relevant feature subsets for training IR and RS models using QA.
- **Clustering:** Based on document embeddings, this task aimed to group similar documents using QA to improve the efficiency of dense retrieval.

Participants accessed D-Wave's quantum annealers via the CINECA supercomputing center and utilized the KIMERA infrastructure [7] for streamlined access, experiment comparability, and reproducibility.

Of the 26 registered teams, 7 submitted official runs [14, 15, 16, 17, 18, 19, 20]. The results demonstrated the practical feasibility of using quantum annealers in IR and RS, encouraged cross-disciplinary research, and laid the groundwork for benchmarking future QC-based systems.

2.3. Related Challenges Outside CLEF

Outside CLEF, we are not aware of other challenges or shared tasks that have been done in the past involving the use of QC for IR and RS. There are instead other challenges offered by big-tech companies such as IBM[1] and Google[2]. These challenges involve the development of QC algorithms, which will be executed on quantum computers to solve some practical real-world challenges.

3. Tasks

QuantumCLEF 2025 addresses three distinct tasks involving computationally intensive problems: Feature Selection, Instance Selection, and Clustering. The main objectives across these tasks include:

- Designing suitable QUBO formulations for each problem;
- Mapping the formulated problems onto quantum annealing hardware;
- Comparing the performance of QA against traditional approaches in terms of both efficiency and effectiveness.

[1]https://challenges.quantum.ibm.com/2024
[2]https://www.xprize.org/prizes/qc-apps

To support participants, we provided Jupyter Notebooks demonstrating how to develop and run quantum annealing solutions, alongside tutorial slides presented at ECIR and SIGIR [21, 22], which introduce the foundational concepts of QC and QA. A video tutorial[3] also helps users through the usage of our KIMERA infrastructure and the provided materials.

Participants are asked to submit runs using both QA and SA for each task. The use of SA during development is strongly recommended due to the limited availability of quantum resources.

3.1. Task 1 - Quantum Feature Selection

This task addresses the challenge of solving the *NP-Hard* feature selection problem through QA, building on prior research [23, 24].

Feature Selection plays a critical role in both IR and RS, aiming to identify the most relevant subset of features for training learning models. By reducing feature dimensionality, models can benefit from improved efficiency and potentially better generalization.

In a QUBO context, this translates to mapping one binary variable per feature, indicating whether it is selected. The main challenge lies in designing the appropriate objective function, i.e., the matrix Q in Equation 1.

Task 1 includes two sub-tasks:

- **Task 1A**: Feature Selection for IR, using selected features to train a LambdaMART [25] model in a Learning-To-Rank framework.
- **Task 1B**: Feature Selection for RS, where the selected features are used to train a kNN-based recommender using cosine similarity and fixed hyperparameters.

For Task 1A, the MQ2007 [26] and Istella S-LETOR [27] datasets are used. MQ2007, with 46 features, allows direct embedding onto current QPU hardware, while Istella's 220 features require preprocessing or hybrid methods. Task 1B uses a custom music recommendation dataset with 1.9k users, 18k items, and 92k implicit interactions. It includes two item feature sets: a smaller 100-feature version and a larger 400-feature version. The larger set requires dimensionality reduction or hybrid methods. The official metric for both sub-tasks is nDCG@10.

Each sub-task will have a corresponding baseline approach:

- **Task 1A**: Recursive Feature Elimination with Linear Regression.
- **Task 1B**: kNN recommender using all features with fixed parameters (cosine similarity, shrinkage 5, 100 neighbors).

Participants may submit up to 5 runs per dataset using either QA/Hybrid or SA. Each QA/Hybrid run should have a corresponding SA run to ensure comparability.

3.2. Task 2 - Quantum Instance Selection

This task explores solving the Instance Selection problem through QA, aiming to reduce training data while preserving or improving model effectiveness.

Instance Selection is crucial in large-scale settings. Selecting representative data instances can reduce training time and resource usage. Previous work has demonstrated that QA can reduce dataset size significantly without degrading model performance [6]. In this task, participants select training instances to fine-tune a Llama3.1 7B model [28] for text classification and sentiment analysis.

Two datasets are considered, each split into 5 folds using cross-validation:

- **Vader NYT**: Sentiment-labeled news articles.
- **Yelp Reviews**: Sentiment-labeled customer reviews.

[3]https://www.youtube.com/watch?v=fKrnaJn40Kk/ (accessed June 17, 2025)

Different evaluation measures are used to evaluate both the efficiency and effectiveness of the model trained on the extracted subsets:

- Macro-F1 score on the test set from each fold;
- Training time for fine-tuning;
- Reduction rate of the dataset.

The baseline approach for this task is the Llama3.1 7B model trained on the full training set.

Participants may submit up to 5 runs per dataset using either QA/Hybrid or SA. Each quantum-based run should have a corresponding traditional one.

3.3. Task 3 - Quantum Clustering

This task involves clustering documents using QA, aiming to group similar documents and support efficient retrieval.

Clustering benefits IR and RS by organizing data, enhancing exploration, and improving retrieval speed. This task applies clustering to document embeddings derived from transformer models, with clustering performed for 10, 25, and 50 groups. Clustering is naturally suited to QUBO, though it poses scalability challenges. Existing methods often require one variable per document, making minor embedding difficult for large datasets. Approaches such as coarsening or hierarchical clustering can help overcome these limitations [29, 30, 31].

The task uses a split of the ANTIQUE dataset [32], containing 6,486 documents and 200 queries. Embeddings were generated using the `all-mpnet-base-v2` model. Of the queries, 50 are used for training, and 150 for testing. The effectiveness is measured through 2 evaluation metrics: the Davies-Bouldin Index to assess intrinsic clustering quality and nDCG@10 to measure retrieval effectiveness using the clustered structure.

A traditional k-Medoids clustering algorithm using cosine distance is used as a baseline approach. Participants can submit up to 5 runs for each number of clusters using either QA/Hybrid or SA.

4. Lab Setup

This section briefly describes the computational infrastructure used in the lab and outlines the guidelines participants were required to follow for submitting their runs.

4.1. Infrastructure

Direct access to quantum annealers is limited by D-Wave through the use of API keys and monthly time quotas. To address this and ensure fair and reproducible experimentation, we adopted KIMERA, a platform that enables participants to access quantum annealers without requiring individual API keys or separate agreements with D-Wave. KIMERA provides each team with an identical computational workspace, standardizing CPU and RAM resources across all participants. This ensures consistency in development environments and facilitates reproducible performance measurements. The platform allows participants to develop, test, and run code directly from their browsers, eliminating the need for local setup or dedicated hardware. All submissions are stored in a centralized database to track quota usage and support analysis and reporting. The infrastructure was deployed on a machine hosted at the Department of Information Engineering, University of Padua. Table 1 details the hardware specifications of the host machine and participant workspaces. Table 2 shows the monthly quotas allocated for quantum resource usage per task.

Table 1
Hardware specifications for the host machine and participant workspaces.

Component	CPU	RAM	Hard Drive
Infrastructure	32 cores	128 GB	1 TB
Workspace	1200 millicores	12 GB	20 GB

Table 2
Monthly quotas for quantum resource usage by task.

Task	February	March	April	May
Task 1: Feature Selection	30 seconds	30 seconds	30 seconds	30 seconds*
Task 2: Instance Selection	120 seconds	120 seconds	120 seconds	120 seconds*
Task 3: Clustering	120 seconds	120 seconds	120 seconds	120 seconds*

* Quantum annealers were unavailable from 05/05/2025 to 10/05/2025 (submission deadline).

Table 3
The teams that participated and submitted to QuantumCLEF 2025.

Team	Affiliation	Country
DS@GT qClef [33]	Georgia Institute of Technology	United States
FAST-NU [34]	National University of Computer and Emerging Sciences	Pakistan
GPLSI [35]	Language Processing and Information Systems (GPLSI), University of Alicante.	Spain
Malto [36]	Politecnico di Torino	Italy
SINAI-UJA [37]	Universidad de Jaén	Spain

Table 4
The breakdown of the runs submitted by the participating teams for each task and subtask.

Team	Task 1		Task 2	Task 3
	A	B		
DS@GT qClef	13	-	9	3
FAST-NU	5	-	-	-
GPLSI	-	-	12	12
Malto	-	3	2	-
SINAI-UJA	10	-	-	-
Total	28	3	23	15

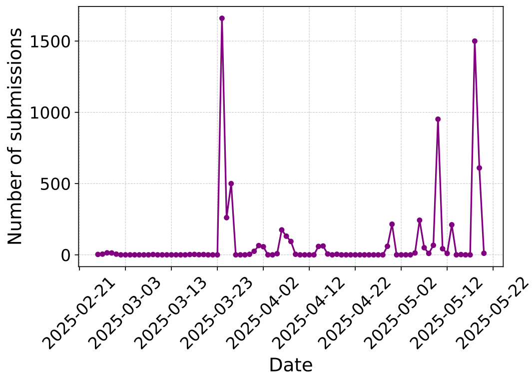
4.2. General Guidelines

Each team was provided with credentials granting access to their personal workspace within the infrastructure. All runs had to be executed exclusively within these designated environments, ensuring fairness and reproducibility across participants. Teams were required to stay within their allocated quantum usage quotas (see Table 2). A real-time dashboard was made available to each team to monitor their usage statistics across the different methods: QA, H, and SA. To support automated evaluation, all submissions were required to follow the standardized file formats.

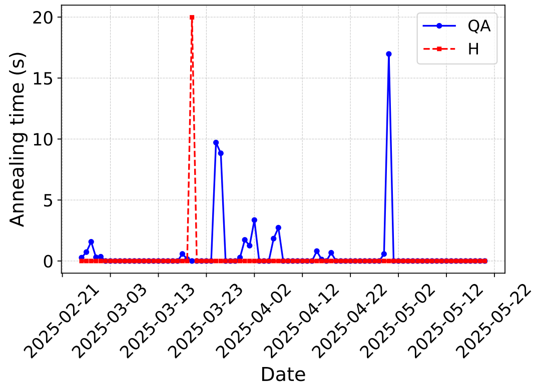
5. Results

In this section, we present the results achieved by the participants and discuss their approaches. Out of the 44 registered teams, 5 teams managed to upload some final runs. In total, the number of runs is 69, considering both SA, QA, and H (H was introduced in Section 2.1.1). Table 3 reports the 5 teams that correctly participated and submitted some final runs.

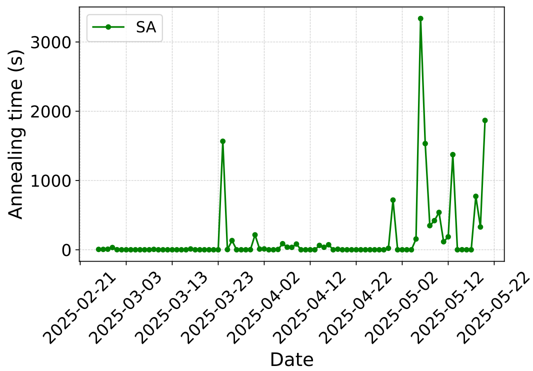
In total, participants submitted 7,183 problems throughout the lab, a significant increase compared to



(a) Number of submissions over time



(b) Distribution of the Annealing time (QA and H)



(c) Distribution of Annealing time (SA)

Figure 1: The distribution of the participating teams' submissions over time, considering also the Annealing time used per day.

the 976 submissions in QuantumCLEF 2024 [2, 3]. Of these, 6,333 were solved using SA, 848 using QA, and 2 using the H approach. The total execution time for SA exceeded 4 hours, whereas the combined execution time for QA and H was approximately 1 minute.

It is important to note that the reported QA execution time refers solely to the *Annealing* phase, as defined in Section 2.1.1. This includes QPU programming, sampling, and result readout, but excludes embedding and network latency, which are left for consideration in future QuantumCLEF editions.

Figure 1 shows the temporal distribution of participant submissions. A pronounced spike is observed during the final days of the lab, highlighting a period of intense usage that placed significant demand on the infrastructure. This trend shows that teams tend to intensify their activity and finalize their work during this final period.

5.1. Task 1A

Here we present the results achieved by the teams participating in task 1A.

5.1.1. MQ2007 dataset.

As it is possible to see in Table 5, teams considered different numbers of features in their submissions. In general, we can observe that most of the submissions achieve similar nDCG@10 values. In fact, Figure 2 shows that for most of the runs the Tukey HSD test performed after the Two-Way ANOVA hypothesis test shows no significant differences.

In terms of efficiency, specifically considering the Annealing time, QA runs consistently required significantly less time compared to SA. On average, QA completed the annealing process in approximately

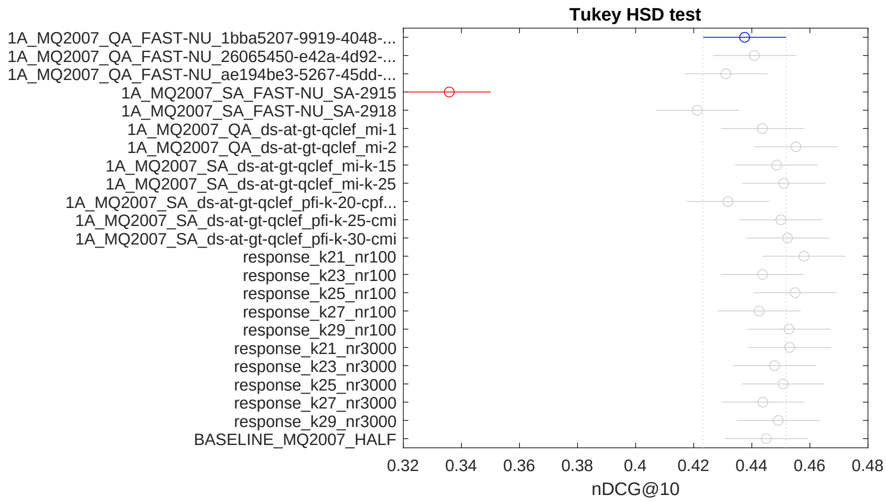
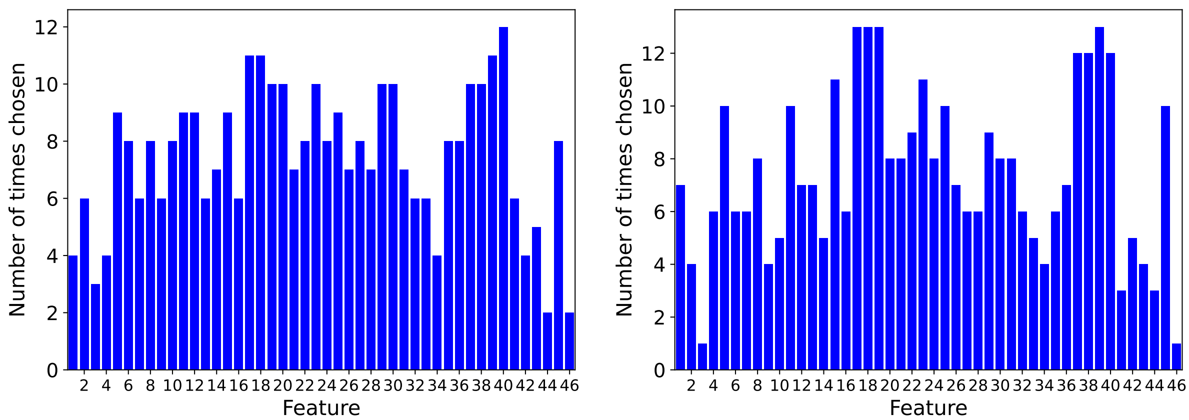


Figure 2: The Tukey HSD test considering the nDCG@10 values associated with different runs and queries for the MQ2007 dataset.



(a) Features kept using QA

(b) Features kept using SA

Figure 3: The number of times each feature of the MQ2007 dataset has been kept considering the different teams' approaches using QA and SA.

26.83 times less time than SA, making it a more time-efficient alternative. Regarding effectiveness, QA demonstrated a more stable and reliable performance across the board. On average, it performed better than SA by a factor of approximately 1.02, indicating a slight advantage. In contrast, SA exhibited greater variability, including two notable outliers that significantly underperformed compared to the rest of the submissions. This might suggest that while SA can occasionally yield competitive results, it may be less robust under certain conditions.

Figure 3 shows how many times each feature has been kept by the participants' approaches using both QA and SA. In general, we can see that both approaches have kept the same features most of the time, indicating that these were probably the most informative features.

To tackle this task, teams adopted a variety of strategies:

- **DS@GT qClef**, drawing inspiration from earlier research [38], explored a range of QUBO formulations that balanced different combinations of feature importance and redundancy measures to guide the selection process [33];
- **FAST-NU** implemented a Mutual Information-based feature selection approach, targeting those features that most strongly conveyed relevance-related information [34];

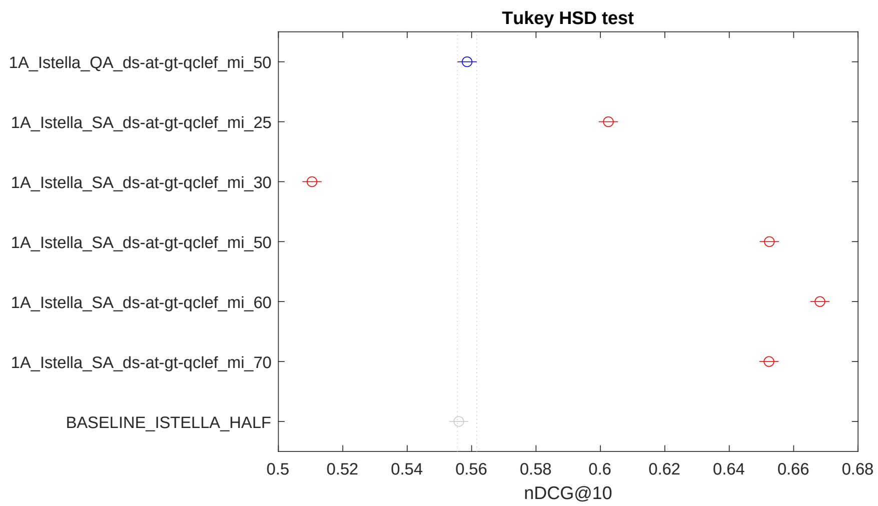


Figure 4: The Tukey HSD test considering the nDCG@10 values associated with different runs and queries for the Istella dataset.

- **SINAI-UJA** similarly employed a Mutual Information-based method, but further refined their results through post-processing techniques, including normalization and projection, to improve the quality and interpretability of the selected feature subsets [37].

5.1.2. Istella dataset.

As shown in Table 6, the submissions for this task considered varying numbers of features. A particularly noteworthy observation is that the baseline method, which employed Recursive Feature Elimination (RFE) to select the top 110 features, underperformed compared to most of the participating teams' submissions, many of which retained significantly fewer features. This can also be seen in Figure 4. Despite the larger feature set, this baseline approach not only yielded lower effectiveness but also required substantial computational resources: nearly two hours of processing time and approximately 24 GB of RAM, far exceeding the specifications allocated to participant workspaces.

These results underscore how the choice of feature subset can significantly influence performance. In particular, the baseline submission *RFE_HALF_FEATURES* exhibited poor effectiveness, likely due to suboptimal feature selection. In contrast, team **DS@GT qClef** employed QUBO-based formulations that integrated both importance and redundancy measures to guide feature selection [33], resulting in more competitive performance.

Furthermore, the hybrid (H) approach demonstrated a considerable advantage in terms of execution time, requiring substantially less Annealing time than the pure SA-based methods. This efficiency is due to a combination of QA with classical hardware computation, thus offering a more time-efficient solution while maintaining competitive performance.

5.2. Task 1B

In this section, we present the results obtained in Task 1B, where the focus was on evaluating the impact of feature selection on recommendation performance and computational efficiency. The results are organized according to the two provided feature sets: `ICM_100` and `ICM_400`.

Table 7 summarizes the performance of various submissions in Task 1B. Effectiveness was measured using nDCG@10, while efficiency was assessed via annealing time.

For the `ICM_100` dataset, the SA-based submission by team **Malto**, which retained 51 features, achieved an nDCG@10 of 0.0207. Although this is slightly below the baseline score of 0.0226, it represents only a minor performance drop. This modest degradation may be an acceptable trade-off, considering the 49% reduction in feature dimensionality, which could translate into significant efficiency gains at inference time.

The `ICM_400` dataset, on the other hand, revealed more diverse outcomes. The best-performing SA-based configuration using 200 features achieved an nDCG@10 of 0.0294, approaching the baseline performance of 0.0328. However, a configuration using only 53 features resulted in a substantial decline in performance (nDCG@10 of 0.0182), illustrating that excessive reduction in feature dimensionality can adversely affect recommendation quality. Moreover, the annealing times for this dataset were considerably higher (typically around 70 to 80 seconds) due to the larger feature space and increased optimization complexity.

Team **Malto** approached Task 1B by computing feature importance using a Random Forest classifier trained on the full feature set. They then formulated a QUBO objective function that incorporated these importance scores alongside pairwise Pearson correlation coefficients, aiming to penalize redundant features and encourage diversity in the selected subset [36].

Overall, the results indicate that SA-based feature selection can significantly reduce the number of features while preserving competitive recommendation performance, especially when a moderate number of features is retained. However, an aggressive feature reduction tends to degrade effectiveness. Additionally, the computational costs, particularly in terms of annealing time, increase with the size of the feature set.

5.3. Task 2: Quantum Instance Selection

Here we present the results obtained by the participating teams in Task 2, organized by dataset.

5.3.1. Yelp Dataset

Table 8 shows the performance results on the Yelp dataset. The teams explored a range of reduction rates, from approximately 25% up to 96% of the original dataset size. This diversity in reduction strategies highlights the varying priorities and experimental approaches of the participants.

A notable submission is *Yelp_SA_qclef_bcos_075*, which improved the effectiveness of the Llama3.1 7B model compared to the full-data baseline. This improvement may be attributed to the removal of noisy or redundant documents, which could otherwise hinder the fine-tuning process and negatively impact performance.

Equally significant is the submission *Yelp_QA_gplsi_2-SentimentKmeansCard*, where a QA-based method achieved an $\approx 87\%$ data reduction while maintaining a high level of performance (98.7 vs. 99.4 on the full dataset). Overall, QA demonstrated competitive effectiveness with respect to SA, while consistently requiring significantly less Annealing time. Below we briefly describe the main strategies adopted by the teams:

- **GPLSI** [35] explored multiple strategies, including *Sentiment Pairs*, which prioritized semantic diversity by selecting pairs of documents with either high or low similarity to reduce overlap; *Local Sets*, combining clustering and noise filtering based on Euclidean distance to select geometrically meaningful instances.

- **DS@GT qClef** [33] built upon a previous method [6] using cosine similarity for the off-diagonal terms of the Q-matrix. For diagonal terms, they introduced two innovative strategies: one based on the distance of instances to an Support Vector Machine (SVM) decision boundary, and another using logistic regression with leave-one-out scoring. These strategies were evaluated both independently and in combination, with batching applied to efficiently manage large datasets.

The results highlight the growing relevance of Instance Selection as a technique for balancing model effectiveness and computational sustainability. In particular, fine-tuning the Llama3.1 7B model on carefully selected subsets resulted in a training time reduction of up to $9\times$, with minimal performance degradation, typically less than one absolute point in macro-F1.

These findings demonstrate that significant efficiency gains can be achieved through strategic data reduction, reinforcing Instance Selection as a critical component in modern, resource-conscious model development workflows.

5.3.2. Vader Dataset

Table 9 presents the results obtained on the Vader dataset for Task 2. Similar to the Yelp dataset, the teams experimented with various reduction levels, ranging from around 25% to 96%. However, unlike the Yelp results, most reductions in this case led to a noticeable decrease in model effectiveness after fine-tuning.

An interesting observation is that the submission *Vader_SA_MALTO_2 - vader nyt_2L* produced a much smaller subset (about 25% of the original data), yet achieved a higher average Macro-F1 score than *Vader_SA_qclef_bcos_075*, which retained about 75% of the data. This discrepancy underscores the impact of dataset-specific characteristics on the outcome of Instance Selection and shows that a larger subset is not always more effective.

The methodologies used on the Vader dataset were generally consistent with those applied to the Yelp dataset. Again, QA-based approaches required significantly less Annealing time compared to their SA-based counterparts, while maintaining comparable performance trends in terms of effectiveness.

5.4. Task 3: Quantum Clustering Results

In this section, we present the outcomes achieved by the teams participating in Task 3. Table 10 summarizes the main results obtained in this task.

All participating teams in this task used only SA to address the clustering challenge. From the reported results, it is evident that both the **GPLSI** and **DS@GT qClef** teams succeeded in submitting solutions that outperformed the traditional k-medoids baseline in terms of nDCG@10 and the Davies–Bouldin Index. These findings suggest that their proposed methods were effective in identifying representative cluster centers that contributed to retrieving relevant documents more efficiently and accurately in response to user queries.

We briefly summarize the key methodologies adopted by the teams:

- **GPLSI** [35] proposed a technique that first reduces the embedding space to 150 pivot points. These pivots are selected using various heuristic methods such as Farthest Point Sampling (FPS), CLARA–CLARANS, k-Means, and SubMedoids (inspired by the qIIMAS approach from the first QuantumCLEF edition [20]). The aim of these techniques is to select pivots that ensure comprehensive coverage of the data space. Subsequently, they used SA to optimize the selection of cluster centroids and assign documents accordingly to maximize retrieval effectiveness.

- **DS@GT qClef** [33] employed a two-step approach. Initially, they applied classical clustering algorithms, such as k-Medoids, HDBSCAN [39], GMM, and a hybrid GMM-HDBSCAN method, with optional dimensionality reduction techniques like UMAP [40] or PaCMAP. This phase was used to identify a reduced and manageable subset of instances. In the second step, they formulated and solved a k-medoids clustering problem using a QUBO formulation on the reduced subset of data.

A particularly noteworthy submission is *50_SA_gplsi_3-FPS-Medoids*, which achieved a significantly higher nDCG@10 compared to the *BASELINE_10*. This is especially remarkable considering that this method used 50 clusters, five times more than the baseline's 10 clusters. Despite the risk of over-segmentation associated with a higher number of clusters, the method successfully maintained high retrieval quality, demonstrating that the clusters identified by the GPLSI approach were both fine-grained and representative. On the other hand, the DS@GT qClef submissions *10_SA_DS@GT_qClef_2* and *50_SA_DS@GT_qClef_3* (marked with an asterisk) employed UMAP for dimensionality reduction, reducing the original high-dimensional embeddings down to just 2 dimensions. While this aggressive reduction simplifies the optimization problem, it likely led to a considerable loss of information, which

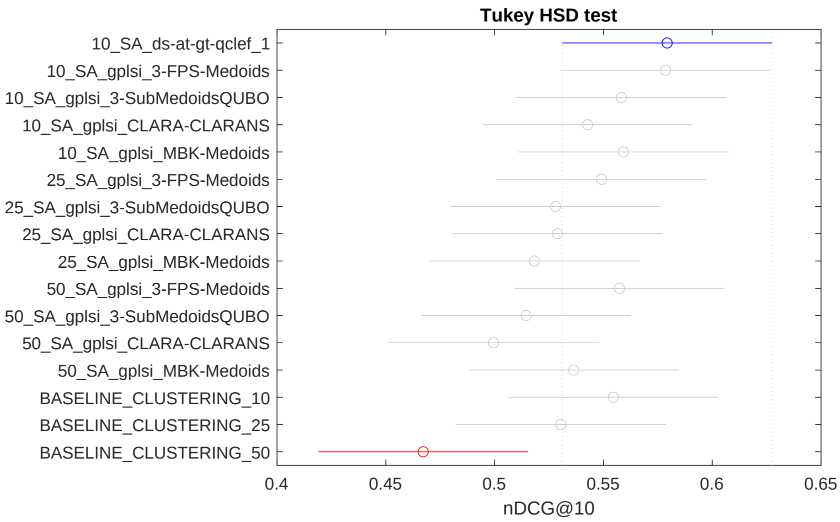


Figure 5: The Tukey HSD test considering the nDCG@10 values associated with different runs and queries for the Clustering dataset.

negatively affected the performance. Moreover, since dimensionality reduction was not expected, this resulted in poor results during the evaluation procedure.

Finally, Figure 5 reports the statistical analysis of the clustering results. The Tukey HSD test indicates no statistically significant differences among the various team submissions in terms of nDCG@10. This suggests that, while different in methodology, the effectiveness of the proposed approaches was statistically comparable.

6. Conclusions and Future Work

In this paper, we presented an overview of the second edition of the QuantumCLEF lab, which was held in 2025. QuantumCLEF is the first CLEF lab focused on the study, development, and evaluation of QC algorithms executed on real quantum hardware. This edition consisted of three tasks addressing the challenges of Feature Selection, Instance Selection, and Clustering, all computationally intensive problems commonly encountered in IR and RS systems.

Participants relied on the KIMERA infrastructure [7], which facilitated the workflow. The infrastructure granted access to both classical computing resources and state-of-the-art quantum annealers provided by D-Wave, allowing participants to experiment with real quantum computers.

A total of 44 teams registered for the lab, of which 5 successfully submitted their runs. The results demonstrated that both QA and H approaches achieved effectiveness levels comparable to those of SA, while offering significantly improved efficiency in terms of Annealing time. These findings support the potential of QC as a promising computational paradigm for tackling complex problems, particularly as the technology continues to mature. Notably, QA produced competitive results when compared to traditional baselines, confirming its capability to deliver effective solutions.

This second edition of QuantumCLEF served not only as an initiative to develop and evaluate QC algorithms on real quantum hardware, which remains today largely inaccessible to the broader research community, but also as an opportunity to raise awareness about the potential of quantum technologies. Participants were provided with educational material, including videos, slides, and practical examples, to help them understand the principles behind QC and QA. Furthermore, we emphasized transparency by allowing participants to directly interact with the D-Wave libraries, thus equipping them with the skills to independently program quantum annealers beyond the scope of this lab.

In the future, we plan to organize a third edition of QuantumCLEF, introducing new tasks and more advanced challenges. Additionally, we are exploring the possibility of extending the infrastructure to include gate-based quantum computers [41], complementing the quantum annealers already in use.

Declaration on Generative AI

We disclose that generative AI technologies were used solely to assist in grammar checking (i.e., Grammarly) during the preparation of this paper. No part of the scientific content or creative reasoning has been generated by generative AI tools.

Acknowledgments

We acknowledge the CINECA award under the ISCRA initiative for the availability of high-performance computing resources and support.

References

[1] A. Pasin, M. Ferrari Dacrema, P. Cremonesi, N. Ferro, qclef: A proposal to evaluate quantum annealing for information retrieval and recommender systems, in: A. Arampatzis, E. Kanoulas, T. Tsikrika, S. Vrochidis, A. Giachanou, D. Li, M. Aliannejadi, M. Vlachos, G. Faggioli, N. Ferro (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction - 14th International Conference of the CLEF Association, CLEF 2023, Thessaloniki, Greece, September 18-21, 2023, Proceedings, volume 14163 of *Lecture Notes in Computer Science*, Springer, 2023, pp. 97–108. URL: https://doi.org/10.1007/978-3-031-42448-9_9. doi:10.1007/978-3-031-42448-9_9.

[2] A. Pasin, M. Ferrari Dacrema, P. Cremonesi, N. Ferro, Quantumclef 2024: Overview of the quantum computing challenge for information retrieval and recommender systems at CLEF, in: G. Faggioli, N. Ferro, P. Galuscáková, A. G. S. de Herrera (Eds.), Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024), Grenoble, France, 9-12 September, 2024, volume 3740 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024, pp. 3032–3053. URL: https://ceur-ws.org/Vol-3740/paper-297.pdf.

[3] A. Pasin, M. Ferrari Dacrema, P. Cremonesi, N. Ferro, Overview of quantumclef 2024: The quantum computing challenge for information retrieval and recommender systems at CLEF, in: L. Goeuriot, P. Mulhem, G. Quénot, D. Schwab, G. M. D. Nunzio, L. Soulier, P. Galuscáková, A. G. S. de Herrera, G. Faggioli, N. Ferro (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction - 15th International Conference of the CLEF Association, CLEF 2024, Grenoble, France, September 9-12, 2024, Proceedings, Part II, volume 14959 of *Lecture Notes in Computer Science*, Springer, 2024, pp. 260–282. URL: https://doi.org/10.1007/978-3-031-71908-0_12. doi:10.1007/978-3-031-71908-0_12.

[4] A. Pasin, M. Ferrari Dacrema, P. Cremonesi, N. Ferro, Quantumclef - quantum computing at CLEF, in: N. Goharian, N. Tonellotto, Y. He, A. Lipani, G. McDonald, C. Macdonald, I. Ounis (Eds.), Advances in Information Retrieval - 46th European Conference on Information Retrieval, ECIR 2024, Glasgow, UK, March 24-28, 2024, Proceedings, Part V, volume 14612 of *Lecture Notes in Computer Science*, Springer, 2024, pp. 482–489. URL: https://doi.org/10.1007/978-3-031-56069-9_66. doi:10.1007/978-3-031-56069-9_66.

[5] A. Pasin, M. Ferrari Dacrema, P. Cremonesi, W. Cunha, M. A. Gonçalves, N. Ferro, Quantumclef 2025 - the second edition of the quantum computing lab at CLEF, in: C. Hauff, C. Macdonald, D. Jannach, G. Kazai, F. M. Nardini, F. Pinelli, F. Silvestri, N. Tonellotto (Eds.), Advances in Information Retrieval - 47th European Conference on Information Retrieval, ECIR 2025, Lucca, Italy, April 6-10, 2025, Proceedings, Part V, volume 15576 of *Lecture Notes in Computer Science*, Springer, 2025, pp. 450–458. URL: https://doi.org/10.1007/978-3-031-88720-8_66. doi:10.1007/978-3-031-88720-8_66.

[6] A. Pasin, W. Cunha, M. A. Gonçalves, N. Ferro, A quantum annealing instance selection approach for efficient and effective transformer fine-tuning, in: H. Oosterhuis, H. Bast, C. Xiong (Eds.), Proceedings of the 2024 ACM SIGIR International Conference on Theory of Information Retrieval,

ICTIR 2024, Washington, DC, USA, 13 July 2024, ACM, 2024, pp. 205–214. URL: https://doi.org/10.1145/3664190.3672515. doi:10.1145/3664190.3672515.

[7] A. Pasin, N. Ferro, Kimera: From evaluation-as-a-service to evaluation-in-the-cloud, in: Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2025, Padova, Italy, July 13-18, 2025, ACM, 2025. URL: https://doi.org/10.1145/3726302.3730298. doi:10.1145/3726302.3730298.

[8] F. Glover, G. Kochenberger, R. Hennig, Y. Du, Quantum bridge analytics I: a tutorial on formulating and using QUBO models, Annals of Operations Research 314 (2022) 141–183.

[9] J. Cai, W. G. Macready, A. Roy, A practical heuristic for finding graph minors, CoRR abs/1406.2741 (2014). URL: http://arxiv.org/abs/1406.2741. arXiv:1406.2741.

[10] S. Yarkoni, E. Raponi, T. Bäck, S. Schmitt, Quantum annealing for industry applications: introduction and review, Reports on Progress in Physics 85 (2022) 104001:1–104001:27.

[11] D. Bertsimas, J. Tsitsiklis, Simulated annealing, Statistical science 8 (1993) 10–15.

[12] P. J. Van Laarhoven, E. H. Aarts, P. J. van Laarhoven, E. H. Aarts, Simulated annealing, Springer, 1987.

[13] D. Bertsimas, O. Nohadani, Robust optimization with simulated annealing, J. Glob. Optim. 48 (2010) 323–334. URL: https://doi.org/10.1007/s10898-009-9496-x. doi:10.1007/S10898-009-9496-X.

[14] J. Niu, J. Li, K. Deng, Y. Ren, CRUISE on quantum computing for feature selection in recommender systems, in: G. Faggioli, N. Ferro, P. Galuscáková, A. G. S. de Herrera (Eds.), Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024), Grenoble, France, 9-12 September, 2024, volume 3740 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024, pp. 3096–3104. URL: https://ceur-ws.org/Vol-3740/paper-303.pdf.

[15] M. Fröbe, D. Alexander, G. Hendriksen, F. Schlatt, M. Hagen, M. Potthast, Team openwebsearch at CLEF 2024: Quantumclef, in: G. Faggioli, N. Ferro, P. Galuscáková, A. G. S. de Herrera (Eds.), Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024), Grenoble, France, 9-12 September, 2024, volume 3740 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024, pp. 3075–3081. URL: https://ceur-ws.org/Vol-3740/paper-300.pdf.

[16] W. Alvarez-Giron, J. Tellezz-Torres, J. Tovar-Cortes, H. Gómez-Adorno, Team qiimas on task 2 - clustering, in: G. Faggioli, N. Ferro, P. Galuscáková, A. G. S. de Herrera (Eds.), Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024), Grenoble, France, 9-12 September, 2024, volume 3740 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024, pp. 3064–3074. URL: https://ceur-ws.org/Vol-3740/paper-299.pdf.

[17] T. M. Almeida, S. Matos, Towards a hyperparameter-free QUBO formulation for feature selection in IR, in: G. Faggioli, N. Ferro, P. Galuscáková, A. G. S. de Herrera (Eds.), Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024), Grenoble, France, 9-12 September, 2024, volume 3740 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024, pp. 3054–3063. URL: https://ceur-ws.org/Vol-3740/paper-298.pdf.

[18] G. Shimi, J. M. C, D. Thenmozhi, Quantum feature selection, in: G. Faggioli, N. Ferro, P. Galuscáková, A. G. S. de Herrera (Eds.), Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024), Grenoble, France, 9-12 September, 2024, volume 3740 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024, pp. 3082–3086. URL: https://ceur-ws.org/Vol-3740/paper-301.pdf.

[19] A. Naebzadeh, S. Eetemadi, NICA at quantum computing CLEF tasks 2024, in: G. Faggioli, N. Ferro, P. Galuscáková, A. G. S. de Herrera (Eds.), Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024), Grenoble, France, 9-12 September, 2024, volume 3740 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024, pp. 3087–3095. URL: https://ceur-ws.org/Vol-3740/paper-302.pdf.

[20] E. Payares, E. Puertas, J. C. M. Santos, Team QTB on feature selection via quantum annealing and hybrid models, in: G. Faggioli, N. Ferro, P. Galuscáková, A. G. S. de Herrera (Eds.), Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024), Grenoble, France, 9-12 September, 2024, volume 3740 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024, pp. 3105–3114. URL: https://ceur-ws.org/Vol-3740/paper-304.pdf.

[21] M. Ferrari Dacrema, A. Pasin, P. Cremonesi, N. Ferro, Using and evaluating quantum computing

for information retrieval and recommender systems, in: G. H. Yang, H. Wang, S. Han, C. Hauff, G. Zuccon, Y. Zhang (Eds.), Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2024, Washington DC, USA, July 14-18, 2024, ACM, 2024, pp. 3017–3020. URL: https://doi.org/10.1145/3626772.3661378. doi:10.1145/3626772.3661378.

[22] M. Ferrari Dacrema, A. Pasin, P. Cremonesi, N. Ferro, Quantum computing for information retrieval and recommender systems, in: N. Goharian, N. Tonellotto, Y. He, A. Lipani, G. McDonald, C. Macdonald, I. Ounis (Eds.), Advances in Information Retrieval - 46th European Conference on Information Retrieval, ECIR 2024, Glasgow, UK, March 24-28, 2024, Proceedings, Part V, volume 14612 of *Lecture Notes in Computer Science*, Springer, 2024, pp. 358–362. URL: https://doi.org/10.1007/978-3-031-56069-9_47. doi:10.1007/978-3-031-56069-9_47.

[23] M. Ferrari Dacrema, F. Moroni, R. Nembrini, N. Ferro, G. Faggioli, P. Cremonesi, Towards feature selection for ranking and classification exploiting quantum annealers, in: E. Amigó, P. Castells, J. Gonzalo, B. Carterette, J. S. Culpepper, G. Kazai (Eds.), SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022, ACM, 2022, pp. 2814–2824. URL: https://doi.org/10.1145/3477495.3531755. doi:10.1145/3477495.3531755.

[24] R. Nembrini, M. Ferrari Dacrema, P. Cremonesi, Feature selection for recommender systems with quantum computing, Entropy 23 (2021) 970. URL: https://doi.org/10.3390/e23080970. doi:10.3390/E23080970.

[25] C. J. C. Burges, From RankNet to LambdaRank to LambdaMART: An Overview, Technical Report, Microsoft Research, MSR-TR-2010-82, 2010.

[26] T. Qin, T. Liu, Introducing LETOR 4.0 datasets, CoRR abs/1306.2597 (2013). URL: http://arxiv.org/abs/1306.2597. arXiv:1306.2597.

[27] C. Lucchese, F. M. Nardini, S. Orlando, R. Perego, F. Silvestri, S. Trani, Post-learning optimization of tree ensembles for efficient ranking, in: R. Perego, F. Sebastiani, J. A. Aslam, I. Ruthven, J. Zobel (Eds.), Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2016, Pisa, Italy, July 17-21, 2016, ACM, 2016, pp. 949–952. URL: https://doi.org/10.1145/2911451.2914763. doi:10.1145/2911451.2914763.

[28] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Grave, G. Lample, Llama: Open and efficient foundation language models, CoRR abs/2302.13971 (2023). URL: https://doi.org/10.48550/arXiv.2302.13971. doi:10.48550/ARXIV.2302.13971. arXiv:2302.13971.

[29] H. Ushijima-Mwesigwa, C. F. A. Negre, S. M. Mniszewski, Graph partitioning using quantum annealing on the d-wave system, CoRR abs/1705.03082 (2017). URL: http://arxiv.org/abs/1705.03082. arXiv:1705.03082.

[30] C. Bauckhage, N. Piatkowski, R. Sifa, D. Hecker, S. Wrobel, A QUBO formulation of the k-medoids problem, in: R. Jäschke, M. Weidlich (Eds.), Proceedings of the Conference on "Lernen, Wissen, Daten, Analysen", Berlin, Germany, September 30 - October 2, 2019, volume 2454 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 54–63. URL: https://ceur-ws.org/Vol-2454/paper_39.pdf.

[31] D. Arthur, P. Date, Balanced k-means clustering on an adiabatic quantum computer, Quantum Inf. Process. 20 (2021) 294. doi:10.1007/s11128-021-03240-8.

[32] H. Hashemi, M. Aliannejadi, H. Zamani, W. B. Croft, ANTIQUE: A non-factoid question answering benchmark, in: J. M. Jose, E. Yilmaz, J. Magalhães, P. Castells, N. Ferro, M. J. Silva, F. Martins (Eds.), Advances in Information Retrieval - 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14-17, 2020, Proceedings, Part II, volume 12036 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 166–173. URL: https://doi.org/10.1007/978-3-030-45442-5_21. doi:10.1007/978-3-030-45442-5_21.

[33] C. Pomeroy, A. Pramov, K. Thakrar, L. Yendapalli, Quantum annealing for machine learning: Applications in feature selection, instance selection, and clustering, in: G. Faggioli, N. Ferro, P. Rosso, D. Spina (Eds.), Working Notes of CLEF 2025 - Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, 2025.

[34] M. T. Shaikh, M. Hamza, S. B. Ali, M. Rafi, S. Zahid, Feature selection using quantum annealing: A mutual information based qubo approach, in: G. Faggioli, N. Ferro, P. Rosso, D. Spina (Eds.), Working Notes of CLEF 2025 - Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, 2025.
[35] J. P. Consuegra-Ayala, A. Morote-Martínez, F. Valero-Abellón, E. Lloret, P. Moreda, M. Palomar, Team gplsi at qclef 2025: Quantum-inspired instance selection and clustering, in: G. Faggioli, N. Ferro, P. Rosso, D. Spina (Eds.), Working Notes of CLEF 2025 - Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, 2025.
[36] F. Giobergia, C. Savelli, A. Koudounas, E. Baralis, Quantum feature selection from interpretable models using qubo formulation, in: G. Faggioli, N. Ferro, P. Rosso, D. Spina (Eds.), Working Notes of CLEF 2025 - Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, 2025.
[37] L. Molino-Piñar, J. Collado-Montañez, A. Montejo-Ráez, Sinai team at quantumclef 2025: Quantum feature selection based on energy with d-wave, in: G. Faggioli, N. Ferro, P. Rosso, D. Spina (Eds.), Working Notes of CLEF 2025 - Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, 2025.
[38] S. Mücke, R. Heese, S. Müller, M. Wolter, N. Piatkowski, Feature selection on quantum computers, Quantum Machine Intelligence 5 (2023) 11.
[39] L. McInnes, J. Healy, S. Astels, et al., hdbscan: Hierarchical density based clustering., J. Open Source Softw. 2 (2017) 205.
[40] E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I. W. Kwok, L. G. Ng, F. Ginhoux, E. W. Newell, Dimensionality reduction for visualizing single-cell data using umap, Nature biotechnology 37 (2019) 38–44.
[41] E. Rieffel, W. Polak, An introduction to quantum computing for non-physicists, ACM Computing Surveys (CSUR) 32 (2000) 300–335.

A. Task 1 - Team Results

Table 5
The results for Task 1A on the MQ2007 dataset. Rows marked in grey (●) represent the results achieved with QA/H, rows marked in yellow(●) refer to the baselines results, and the remaining refer to results SA results.

Group	Submission id	nDCG@10	Annealing time (ms)	Type	N° features
DS@GT qClef	1A_MQ2007_SA_DS@GT qClef_pfi-k-25-cmi	0.4500	2219	SA	25
DS@GT qClef	1A_MQ2007_SA_DS@GT qClef_pfi-k-20-cpfi	0.4318	2185	SA	20
DS@GT qClef	1A_MQ2007_SA_DS@GT qClef_mi-k-25	0.4510	2214	SA	25
DS@GT qClef	1A_MQ2007_SA_DS@GT qClef_mi-k-15	0.4485	2136	SA	15
DS@GT qClef	1A_MQ2007_SA_DS@GT qClef_pfi-k-30-cmi	0.4523	2157	SA	30
DS@GT qClef	1A_MQ2007_QA_DS@GT qClef_mi-1	0.4436	183	QA	15
DS@GT qClef	1A_MQ2007_QA_DS@GT qClef_mi-2	0.4552	160	QA	13
FAST-NU	MQ2007_SA_FAST-NU_SA-2918	0.4212	4073	SA	15
FAST-NU	1A_MQ2007_SA_FAST-NU_SA-2915	0.3358	4164	SA	15
FAST-NU	1A_MQ2007_QA_FAST-NU_ae194be3-5267-45dd-aa0e-36a58579d719	0.4311	339	QA	15
FAST-NU	1A_MQ2007_QA_FAST-NU_26065450-e42a-4d92-bfb9-ff367d132142	0.4409	287	QA	15
FAST-NU	1A_MQ2007_QA_FAST-NU_1bba5207-9919-4048-b4a0-80f89b03f603	0.4375	275	QA	15
SINAI-UJA	response_k21_nr3000	0.4530	3448	SA	21
SINAI-UJA	response_k23_nr3000	0.4478	6632	SA	23
SINAI-UJA	response_k25_nr3000	0.4510	2998	SA	25
SINAI-UJA	response_k27_nr3000	0.4438	6637	SA	27
SINAI-UJA	response_k29_nr3000	0.4491	6614	SA	29
SINAI-UJA	response_k21_nr100	0.4580	34	QA	21
SINAI-UJA	response_k23_nr100	0.4437	37	QA	23
SINAI-UJA	response_k25_nr100	0.4550	31	QA	25
SINAI-UJA	response_k27_nr100	0.4425	34	QA	27
SINAI-UJA	response_k29_nr100	0.4528	34	QA	29
BASELINE	ALL_FEATURES	0.4473	-	-	46
BASELINE	RFE HALF_FEATURES	0.4450	-	-	23

Table 6
The results for Task 1A on the Istella dataset. Rows marked in grey (●) represent the results achieved with QA/H, rows marked in yellow(●) refer to the baselines results, and the remaining refer to results achieved with SA.

Group	Submission id	nDCG@10	Annealing time (ms)	Type	N° features
DS@GT qClef	1A_Istella_SA_DS@GT qClef_mi_25	0.6025	126631	SA	25
DS@GT qClef	1A_Istella_SA_DS@GT qClef_mi_30	0.5104	133814	SA	30
DS@GT qClef	1A_Istella_SA_DS@GT qClef_mi_50	0.6524	159964	SA	50
DS@GT qClef	1A_Istella_SA_DS@GT qClef_mi_60	0.6682	173222	SA	60
DS@GT qClef	1A_Istella_SA_DS@GT qClef_mi_70	0.6523	184047	SA	70
DS@GT qClef	1A_Istella_QA_DS@GT qClef_mi_50	0.5586	9987	H	50
BASELINE	ALL_FEATURES	0.7146	-	-	220
BASELINE	RFE HALF_FEATURES	0.5560	-	-	110

Table 7
The results for Task 1B. Rows marked in grey (●) represent the results achieved with QA/H, rows marked in yellow(●) refer to the baselines results, and the remaining refer to results achieved with SA.

Dataset	Group	Submission id	nDCG@10	Annealing time (ms)	Type	N° features
ICM_100	Malto	1B_100_ICM_SA_MALTO_1B - 100_ICM submission	0.0207	6149	SA	51
	BASELINE	ALL_FEATURES	0.0226	-	-	100
ICM_400	Malto	1B_400_ICM_SA_MALTO_1B - 400_ICM submission - 200	0.0294	80781	SA	200
	Malto	1B_400_ICM_SA_MALTO_1B - 400_ICM submission	0.0182	70269	SA	53
	BASELINE	ALL_FEATURES	0.0328	-	-	400

B. Task 2 - Team Results

Table 8
The results for Task 2 on the Yelp dataset averaged over 5 folds. Rows marked in grey (●) represent the results achieved with QA/H, rows marked in yellow(●) refer to the baselines results, and the remaining refer to results achieved with SA.

Group	Submission id	Avg Macro F1	Avg Reduction	Avg Fine-Tuning time (s)	Avg Annealing time (ms)	Type
DS@GT qClef	Yelp_SA_qclef_bcos_075	99.5(0.2)	0.25	1548.5(2.8)	25997	SA
DS@GT qClef	Yelp_SA_qclef_it_del_075	99.3(0.3)	0.25	1549.2(1.5)	25784	SA
DS@GT qClef	Yelp_SA_qclef_svc_075	99.3(0.4)	0.25	1550.5(2.6)	25917	SA
DS@GT qClef	Yelp_QA_qclef_bcos	99.4(0.2)	0.274	1500(54.7)	1767	QA
GPLSI	Yelp_SA_gplsi_2-SentimentPairs(docs=just-final...	90.8(5.7)	0.963	170.8(3.8)	35810	SA
GPLSI	Yelp_SA_gplsi_2-SentimentPairs(docs=pair-related...	99.2(0.3)	0.627	822.2(395)	35810	SA
GPLSI	Yelp_SA_gplsi_2-LocalSets	99.4(0.2)	0.512	1045.5(5.3)	28789	SA
GPLSI	Yelp_SA_gplsi_2-SentimentKmeansCard	98.5(1.1)	0.875	338.8(21)	17823	SA
GPLSI	Yelp_SA_gplsi_2-emoconflictCard	98.6(0.5)	0.728	628.2(65.9)	34024	SA
GPLSI	Yelp_QA_gplsi_2-SentimentKmeansCard	98.7(0.2)	0.869	351(25.1)	553	QA
GPLSI	Yelp_QA_gplsi_2-emoconflictCard	98.8(0.6)	0.702	678.8(80.9)	549	QA
Malto	Yelp_SA_MALTO_2 - vader_nyt_2L_0	99.2(0.2)	0.751	582(2)	142949	SA
BASELINE	BASELINE_ALL	99.4(0.1)	-	2027.1(1.1)	-	-

Table 9

The results for Task 2 on the Vader dataset averaged over 5 folds. Rows marked in grey (●) represent the results achieved with QA/H, rows marked in yellow(●) refer to the baselines results, and the remaining refer to results achieved with SA.

Group	Submission id	Avg Macro F1	Avg Reduction	Avg Fine-Tuning time (s)	Avg Annealing time (ms)	Type
DS@GT qClef	Vader_SA_qclef_svc_075	65.4(7.1)	0.25	1529(2.4)	25530	SA
DS@GT qClef	Vader_SA_qclef_combined_075	65.9(4.7)	0.25	1529.4(3)	25300	SA
DS@GT qClef	Vader_SA_qclef_it_del_075	65.6(3)	0.25	1529.5(2.3)	25348	SA
DS@GT qClef	Vader_SA_qclef_bcos_075	62.5(10.4)	0.25	1528.6(2.2)	25735	SA
DS@GT qClef	Vader_QA_qclef_bcos	62.6(7.5)	0.283	1493.3(83)	1874	QA
GPLSI	Vader_SA_gplsi_2-LocalSets	63.3(4.9)	0.505	1048.3(6.7)	29110	SA
GPLSI	Vader_SA_gplsi_2-SentimentPairs-docs=just-final...	47.4(5.4)	0.962	172.8(5.7)	42408	SA
GPLSI	Vader_SA_gplsi_2-SentimentPairs-docs=pair-related...	62.2(4.1)	0.7	671.8(352.8)	42408	SA
GPLSI	Vader_QA_gplsi_2-SentimentPairs-docs=just-final...	50(64)*	0.835*	172.9(26.9)*	545*	QA
GPLSI	Vader_QA_gplsi_2-SentimentPairs(docs=pair-related...	62.1(1.8)*	0.658*	750.7(2653.2)*	545*	QA
Malto	Vader_SA_MALTO_2 - vader_nyt_2L	63.1(2.5)	0.751	574.5(1.7)	126087	SA
BASELINE	BASELINE_ALL	88.9(0.8)	-	1997.3(5.7)	-	-

* The submission did not include all 5 folds

C. Task 3 - Team Results

Table 10

The results for Task 3. Rows marked in grey (●) represent the results achieved with QA/H, rows marked in yellow(●) refer to the baselines results, and the remaining refer to results achieved with SA.

N° centroids	Team	Submission id	nDCG@10	DBI	Annealing Time (ms)	Type
10	GPLSI	10_SA_gplsi_3-FPS-Medoids	0.5783	7.5147	15375	SA
	GPLSI	10_SA_gplsi_3-SubMedoidsQUBO	0.5579	6.8779	15305	SA
	GPLSI	10_SA_gplsi_CLARA-CLARANS	0.5444	6.6710	15395	SA
	GPLSI	10_SA_gplsi_MBK-Medoids	0.5600	6.4258	15510	SA
	DS@GT qClef	10_SA_DS@GT qClef_1	0.5800	7.4776	83	SA
	DS@GT qClef	10_SA_DS@GT qClef_2 *	0.0172	4.4706	83	SA
	BASELINE	BASELINE_10	0.5509	7.9892	-	-
25	GPLSI	25_SA_gplsi_3-FPS-Medoids	0.5475	5.5577	20875	SA
	GPLSI	25_SA_gplsi_3-SubMedoidsQUBO	0.5298	5.6255	40687	SA
	GPLSI	25_SA_gplsi_CLARA-CLARANS	0.5310	5.6507	20532	SA
	GPLSI	25_SA_gplsi_MBK-Medoids	0.5193	5.3755	20758	SA
	BASELINE	BASELINE_25	0.5284	6.1201	-	-
50	GPLSI	50_SA_gplsi_3-FPS-Medoids	0.5592	4.4531	9869	SA
	GPLSI	50_SA_gplsi_3-SubMedoidsQUBO	0.5148	4.9325	23719	SA
	GPLSI	50_SA_gplsi_CLARA-CLARANS	0.5017	5.1703	9976	SA
	GPLSI	50_SA_gplsi_MBK-Medoids	0.5383	4.5025	24004	SA
	DS@GT qClef	50_SA_DS@GT qClef_3 *	0.0064	3.4217	228	SA
	BASELINE	BASELINE_50	0.4656	5.3679	-	-

* Dimensionality reduction was applied