QuantumCLEF 2026: Overview of the Third Quantum Computing Challenge for Information Retrieval and Recommender Systems at CLEF

Andrea Pasin[1,*], Maurizio Ferrari Dacrema[2], Washington Cunha[3], Marcos André Gonçalves[4], Paolo Cremonesi[2] and Nicola Ferro[2]

[1]*University of Padua, Padua, Italy*
[2]*Politecnico di Milano, Milano, Italy*
[3]*Universidade Estadual de Campinas, Campinas, Brazil*
[4]*Federal University of Minas Gerais, Belo Horizonte, Brazil*

Abstract

Although still at an early stage, Quantum Computing (QC) is increasingly regarded as a technology with the potential to reshape computational research. Its potential ability to tackle specific classes of highly complex problems more efficiently than classical approaches has motivated growing interest from both academia and industry. Current quantum hardware already allows researchers to experiment with practical applications, providing early insights regarding the capabilities of what quantum methods may eventually deliver once current hardware limitations are overcome.

Within this context, the QuantumCLEF lab was established to promote the adoption of quantum technologies and to stimulate research on quantum-based solutions for problems arising in Information Retrieval (IR) and Recommender Systems (RS). Beyond algorithmic development and evaluation, the initiative also gives participants hands-on experience with real cutting-edge quantum computers that are not commonly accessible.

This paper describes the third edition of QuantumCLEF, dedicated to the application of Quantum Annealing (QA) techniques to three practical tasks: Feature Selection for IR and RS, Instance Selection for IR, and Clustering for IR. The lab received registrations from several teams, among which 8 submitted valid experimental runs according to the lab rules. Since QC is still not well known within the IR and RS communities, participants were supplied with supporting materials and tutorials, covering the fundamentals of QA and the usage of quantum annealers.

Keywords

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

1. Introduction

Over the last decades, Information Retrieval (IR) and Recommender Systems (RS) technologies have achieved remarkable progress and have become fundamental in our everyday lives. Nevertheless, the rapid expansion of available information and the increasing computational demands of data processing continue to expose important limitations in current approaches. As datasets grow larger and optimization problems become more intricate, conventional computational methods often struggle to maintain a good trade-off between efficiency and effectiveness.

In recent years, the computational paradigm of Quantum Computing (QC) has emerged as a possible alternative to classical computation. Rather than representing a simple hardware improvement, QC introduces a radically different computational model based on the principles of quantum mechanics. Classical computers operate through binary bits that assume values of either 0 or 1, whereas quantum systems rely on qubits, which can exist in a superposition of 0 and 1 across a continuum of possible

CLEF 2026: Conference and Labs of the Evaluation Forum, September 21–24, 2026, Jena, Germany
*Corresponding author.
andrea.pasin.1@phd.unipd.it (A. Pasin); maurizio.ferrari@polimi.it (M. Ferrari Dacrema); wcunha@unicamp.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)

states. In addition, qubits exhibit uniquely quantum phenomena such as entanglement and tunnelling, which do not occur in classical systems.

Thanks to these properties, quantum computers are expected to provide substantial advantages for specific categories of problems, particularly optimization and combinatorial search tasks. Such capabilities make QC especially attractive for domains like IR and RS, where efficiency and effectiveness often depend on navigating extremely large solution spaces. It is important to note that quantum computers should not be thought of as a complete substitute for classical systems but rather, in the long term, as a possible complement or integration within traditional processing pipelines. Despite this potential, current quantum technologies remain limited by several factors, including hardware instability, restricted qubit counts, and sensitivity to external disturbances such as thermal fluctuations and electromagnetic noise. Unlike classical systems, which have benefited from decades of engineering refinement, quantum devices are still in an experimental and rapidly evolving phase.

Motivated by these developments, we introduced QuantumCLEF in 2023 [1], a CLEF lab designed to investigate the role of QC in addressing challenges related to IR and RS. The initiative was conceived with four primary objectives:

- Encourage the development and assessment of quantum-based approaches for IR and RS, comparing them against established classical solutions;
- Release datasets, benchmarks, and resources that facilitate reproducibility and stimulate further research;
- Offer educational support and practical access to quantum infrastructures, which are still difficult for many researchers to obtain independently;
- Promote interest in quantum technologies and support the growth of a research community focused on their application to information systems.

In this paper, we describe the third edition of QuantumCLEF, organized in 2026. Following the direction established during the previous two editions [2, 3, 4, 5, 6], the lab concentrated on Quantum Annealing (QA), a quantum-computing paradigm specifically suited to optimization problems. Participants were given access to cutting-edge QA hardware produced by D-Wave, one of the leading companies in the sector.

Compared with gate-based quantum computing, QA provides a more accessible entry point for researchers because of its simpler programming model. This allowed participants to focus primarily on experimentation and algorithmic design without requiring advanced expertise in quantum mechanics.

The 2026 edition of QuantumCLEF included the following three tasks [5]:

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

Research teams were encouraged to experiment with both QA and Simulated Annealing (SA), the latter being a classical optimization strategy conceptually related to QA but not a simulation of it. Since many participants were approaching these technologies for the first time, we prepared extensive supporting material, including tutorials, presentations, code examples, and explanatory videos covering the principles of QA and the programming of quantum annealers. To further simplify experimentation on real hardware, the lab adopted the Kubernetes Infrastructure for Managed Evaluation and Resource Access (KIMERA) infrastructure [7], which streamlined execution workflows and improved reproducibility across experiments. Overall, the initiative attracted several teams, with 8 teams eventually submitting valid runs for the proposed tasks.

The obtained results were consistent with those observed in the previous editions. In several cases, QA-driven and hybrid solutions achieved effectiveness comparable to classical techniques and SA, while also demonstrating promising efficiency characteristics. These outcomes reinforce the idea that QA already represents a viable strategy for addressing optimization problems in IR and RS. As quantum technologies continue to evolve, they may increasingly complement state-of-the-art classical pipelines and contribute to future improvements in both efficiency and effectiveness.

The remainder of this paper is organized as follows. Section 2 presents related work; Section 3 describes the QuantumCLEF 2026 tasks; Section 4.1 illustrates the infrastructure and experimental setup adopted for the lab; Section 5 presents and analyzes the participants' results; finally, Section 6 concludes the paper and outlines possible future research directions.

2. Related Works

In this section, we provide an overview of QA and SA, followed by a summary of the outcomes of the previous editions of QuantumCLEF.

2.1. 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.

2.2. 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.

2.3. Previous Editions of QuantumCLEF

The first edition of QuantumCLEF took place in 2024 [2, 3, 4], exploring the application of QA for IR and RS when tackling the tasks of Feature Selection and Clustering. In its second edition, the task of Instance Selection has been added, thus totalling 3 tasks. During both editions, participants were granted access to D-Wave's cutting-edge quantum annealers via the CINECA supercomputing center. Participants developed their approaches and submitted them using the KIMERA infrastructure [7], which played a crucial role in streamlining resource access, comparability, and reproducibility.

Overall, 7 teams submitted official runs for the first edition [14, 15, 16, 17, 18, 19, 20] and 5 teams for the second edition [21, 22, 23, 24, 25], most of which never have experienced using QC resources before the lab itself. The results obtained throughout these editions demonstrated that real quantum annealers can already be employed to address optimization problems in IR and RS. Furthermore, the lab contributed to establishing common benchmarks and methodologies that can support future research on the application of QC technologies to IR and RS systems. Finally, the lab has been instrumental in raising awareness of the potential of quantum resources and in providing researchers with the foundations to employ such technologies in the future.

2.4. 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 2026 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;

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

- Comparing the performance of QA against traditional approaches in terms of both efficiency and effectiveness.

To support participants, we provided Jupyter Notebooks demonstrating how to develop and run quantum annealing solutions, alongside tutorial slides presented at ECIR and SIGIR [26, 27], 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 [28, 29].

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 [30] 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 [31] and Istella S-LETOR [32] 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 in both tasks 1A and 1B can submit a maximum of 5 runs per dataset using QA or Hybrid methods and a maximum of 5 runs using SA. Each run that uses QA or Hybrid methods should correspond to a run that employs SA to make a fair comparison between them.

The results of the run must be a text file which lists the features that were selected, one per line. The discarded features are not reported in the run file. Furthermore, the last line must report the list of IDs associated with the problems solved using QA, SA, or Hybrid to obtain the final subset of features by the considered approach.

Each run file must be left in each team's workspace in a specific directory called */config/workspace/-submissions*, which is already available.

The submission file name should comply with the format
[Task]_[Dataset]_[Method]_[Groupname]_[SubmissionID].txt, where:

- **[Task]**: it should be either *1A* or *1B* based on the task the submission refers to;

[3] https://www.youtube.com/watch?v=fKrnaJn40Kk/ (accessed June 15, 2026)

- **[Dataset]**: it should be either *MQ2007*, *Istella*, *100_ICM* or *400_ICM* based on the dataset used;
- **[Method]**: it should be either *QA* or *SA* based on the method used;
- **[Groupname]**: the team name;
- **[SubmissionID]**: a custom submission ID that must be the same for the submissions using the same algorithm but performed with different methods (e.g., QA or SA).

3.2. Task 2 - Quantum Instance Selection and Noise Reduction

This task addresses the Instance Selection problem using QA. The goal of Instance Selection is to identify a subset of training samples that represents the original dataset, thereby reducing the amount of training data required while preserving model effectiveness [33, 34, 35]. By retaining only the most informative instances, it is possible to improve training efficiency and lower computational costs.

Compared to previous editions, the 2026 scenario introduces an additional challenge: the considered dataset contains artificially injected noise in the class labels. As a result, participants must not only identify representative instances but also detect and remove noisy or misleading samples that may negatively affect the learning process and compromise model performance [36]. Therefore, this task combines two complementary objectives: dataset reduction and noise filtering.

From an optimization perspective, this setting gives rise to a complex search problem in which the selected subset should simultaneously maximize representativeness and classification performance while minimizing redundancy and the inclusion of corrupted instances. Achieving an effective balance among these objectives is essential for obtaining compact, high-quality training subsets capable of supporting robust predictive models.

In this task, participants select training instances to fine-tune a Llama3.1 7B model [37] for text classification and sentiment analysis.

The Vader NYT Sentiment-labeled news articles dataset is considered, split into 5 folds using cross-validation. 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 can submit a maximum of 5 runs per dataset using QA or Hybrid methods and a maximum of 5 runs using SA. Each run that uses QA or Hybrid methods should correspond to a run that employs SA. In this way, it is possible to make a fair comparison between them.

The results of the run must be a text file which lists the documents that were selected, one per line. The discarded documents are not reported in the run file. Furthermore, the last line must report the list of IDs associated with the problems solved using QA, SA, or Hybrid to obtain the final subset of features by the considered approach.

Each run file must be left in each team's workspace in a specific directory called */config/workspace/-submissions*, which is already available.

The submission file name should comply with the format
Vader_[FoldNumber]_[Method]_[Groupname]_[SubmissionID].txt, where:

- **[FoldNumber]**: it should be a number in $[0, 4]$, representing the considered fold;
- **[Method]**: it should be either *QA* or *SA* based on the method used;
- **[Groupname]**: the team name;
- **[SubmissionID]**: a custom submission ID that must be the same for the submissions using the same algorithm but performed with different methods (e.g., QA or SA).

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 [38, 39, 40].

The task uses a split of the ANTIQUE dataset [41], 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 a maximum of 5 runs for each number of clusters (i.e., 10, 25, 50) using QA or Hybrid methods and a maximum of 5 runs using SA. Each run that uses QA or Hybrid methods should correspond to a run that employs SA. In this way, it is possible to make a fair comparison between them.

The run file must be a text file (JSON formatted) with a list of 10, 25, and 50 vectors that represent the final centroids achieved through their clustering algorithm. Each centroid should also be followed by the list of documents that belong to the given cluster. Furthermore, the last line must report the list of IDs associated with the problems solved using QA, SA, or Hybrid to obtain the final clusters by the considered approach.

Each run file must be left in each team's workspace in a specific directory called */config/workspace/-submissions*, which is already available.

The submission file name should comply with the format
[Centroids]_[Method]_[Groupname]_[SubmissionID].txt, where:

- **[Centroids]**: it should be either 10, 25, or 50 based on the number of centroids;
- **[Method]**: it should be either *QA* or *SA* based on the method used;
- **[Groupname]**: the team name;
- **[SubmissionID]**: a custom submission ID that must be the same for the submissions using the same algorithm but performed with different methods (e.g., QA 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 [7], 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

Table 1
The hardware resources corresponding to the machine where KIMERA was deployed and to the participants' workspaces.

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

Table 2
The monthly quotas to use quantum resources according to the tasks.

Task	March	April	May
Task 1: Feature Selection	30 seconds	50 seconds	50 seconds
Task 2: Instance Selection	120 seconds	180 seconds	180 seconds
Task 3: Clustering	60 seconds	90 seconds	90 seconds

Table 3
The teams that participated and submitted to QuantumCLEF 2026.

Team	Affiliation	Country
DS@GT QuantumCLEF [42]	Georgia Institute of Technology	United States
Entropy of Void [43]	-	India
FAST-NUCES [44]	National University of Computer and Emerging Sciences	Pakistan
GPLSI [45]	University of Alicante	Spain
QBoson	-	China
R2AI [46]	RMIT University	Australia
Skywalkers [47]	Indian Institute of Information Technology, Design and Manufacturing, Kurnool	India
SYZYGY [48]	Graz University of Technology	Austria

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.

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.

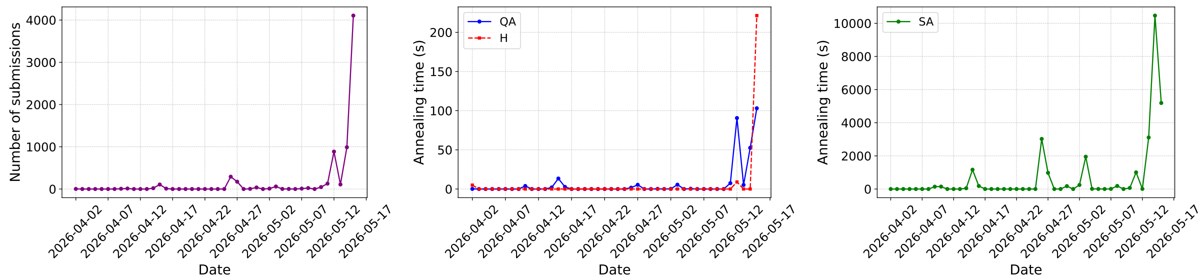
5. Results

In this section, we present the results achieved by the participants and discuss their approaches. 8 teams managed to upload some final runs. In total, the number of runs is 144, considering both SA, QA, and H (H was introduced in Section 2.1). Table 3 reports the 8 teams that correctly participated and submitted some final runs. Moreover, Table 4 presents the breakdown of submissions made by each team for each task.

In total, throughout the entire lab, participants have submitted 6787 problems (vs 7183 in QuantumCLEF 2025 and 976 in QuantumCLEF 2024). Specifically, 1820 of them were solved with SA, while 4889 were solved using QA and 78 with the H method. The total execution time of SA has been close to 8 hours, while the total QA execution time has been roughly 295 seconds, and H execution time has been

Table 4
The breakdown of the runs submitted by the participating teams for each task and subtask of QuantumCLEF 2026.

Team	**Task 1**		**Task 2**	**Task 3**
	A	**B**		
DS@GT QuantumCLEF	-	-	2	-
Entropy of Void	-	-	2	30
FAST-NUCES	2	4	3	15
GPLSI	-	-	10	30
QBoson	2	-	-	-
R2AI	-	20	2	-
Skywalkers	-	10	-	-
SYZYGY	10	2	-	-
Total	14	36	19	75



(a) Number of submissions over time (b) Distribution of 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.

roughly 235 seconds.

It is important to note that the reported QA execution time refers solely to the *Annealing* phase, as defined in Section 2.1. 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. 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 across QA, SA, and H.

It is possible to see that, in terms of efficiency (i.e., Annealing time), runs using QA required a substantially shorter amount of time with respect to SA. On average, QA required $\approx$ 15.6 times less compared to SA, thus representing a more efficient alternative. Considering effectiveness, they perform on par. In fact, on average SA performs $\approx$ 1.005 times better compared to QA, which is a negligible difference. This is also confirmed by the statistical significance analysis carried out using a two-way ANOVA [49] with Tukey HSD post-hoc test [50] reported in Figure 2, which confirms that no submission was statistically significantly better than any other in terms of nDCG@10.

Figure 3 shows how many times each feature has been kept by the participants' approaches using both QA and SA. In general, we can observe that there exist some common feature sets that were

Table 5
The results for Task 1A on the MQ2007 dataset. An adjacent couple of rows (marked with the same color) represents the results achieved with QA/H and SA using the same problem formulation. Results marked in yellow(●) refer to the baselines' results.

Team	Submission id	nDCG@10	Annealing time (ms)	Type	N° features
FAST-NUCES	1A_MQ2007_QA_FAST-NUCES_k12	0.4472	154	Q	18
FAST-NUCES	1A_MQ2007_SA_FAST-NUCES_k12	0.4412	4323	S	12
SYZYGY	1A_MQ2007_QA_SYZYGY_II+weighted_MI_1	0.4513	419	Q	23
SYZYGY	1A_MQ2007_SA_SYZYGY_II+weighted_MI_1	0.4498	5427	S	24
SYZYGY	1A_MQ2007_QA_SYZYGY_II+weighted_MI_2	0.4373	381	Q	22
SYZYGY	1A_MQ2007_SA_SYZYGY_II+weighted_MI_2	0.4482	5779	S	21
SYZYGY	1A_MQ2007_QA_SYZYGY_II	0.4450	348	Q	38
SYZYGY	1A_MQ2007_SA_SYZYGY_II	0.4473	3775	S	40
SYZYGY	1A_MQ2007_QA_SYZYGY_Weighted_MI_1	0.4477	267	Q	22
SYZYGY	1A_MQ2007_SA_SYZYGY_Weighted_MI_1	0.4479	4626	S	24
SYZYGY	1A_MQ2007_QA_SYZYGY_Weighted_MI_2	0.4505	403	Q	21
SYZYGY	1A_MQ2007_SA_SYZYGY_Weighted_MI_2	0.4519	4281	S	21
Qboson	1A_MQ2007_QA_Qboson_rankfs_qpu_only_v1	0.4400	139	Q	10
-	-	-	-	S	-
Qboson	1A_MQ2007_QA_Qboson_rankfs_v6	0.4361	4990	H	10
Qboson	-	-	-	S	-
BASELINE	ALL_FEATURES	0.4473	-	-	46
BASELINE	RFE HALF_FEATURES	0.4450	-	-	23

frequently selected among the participants' submissions, which could be a possible indication that some feature sets are more informative than others.

Different strategies were adopted by the participating teams. In particular:

- Team **FAST-NUCES** developed a QUBO-based formulation that integrates mutual information-driven relevance scoring with feature stability estimates obtained through cross-validation, while incorporating Pearson correlation penalties to reduce redundancy among selected features [44].
- Team **SYZYGY** introduced five information-theoretic QUBO formulations leveraging mutual information, conditional mutual information, interaction information, and two normalized weighted mutual information measures. Unlike conventional approaches, these formulations do not impose explicit cardinality constraints, allowing the intrinsic structure of the data to determine the size of the selected feature subset [48].

5.2. Task 1B: Quantum Feature Selection for RS

Here we report the results obtained for Task 1B. The outcomes are organized according to the two considered feature sets. Table 6 summarizes the performance of the different submissions for Task 1B on the ICM_100 dataset, highlighting the effect of feature selection on recommendation quality, measured through nDCG@10, as well as on computational requirements, quantified in terms of annealing time.

For the ICM_100 dataset, several teams were able to surpass the baseline configuration that retained all available features. In particular, the best nDCG@10 score was achieved by team **Skywalkers** with submission `1B_100_ICM_SA_skywalkers_012`, which retained 59 out of the original 100 features. Overall, the results indicate that QA tends to underperform compared with SA in terms of recommendation effectiveness in this setting. Conversely, the H approach appears to provide a robust and competitive alternative. Furthermore, both QA and H generally offer greater computational efficiency than SA, with QA being particularly advantageous, as it is typically at least one order of magnitude faster.

Table 7 reports the corresponding results for the ICM_400 dataset. A different trend emerges for the ICM_400 dataset. In this case, both H and QA achieve higher effectiveness than SA, while simultaneously

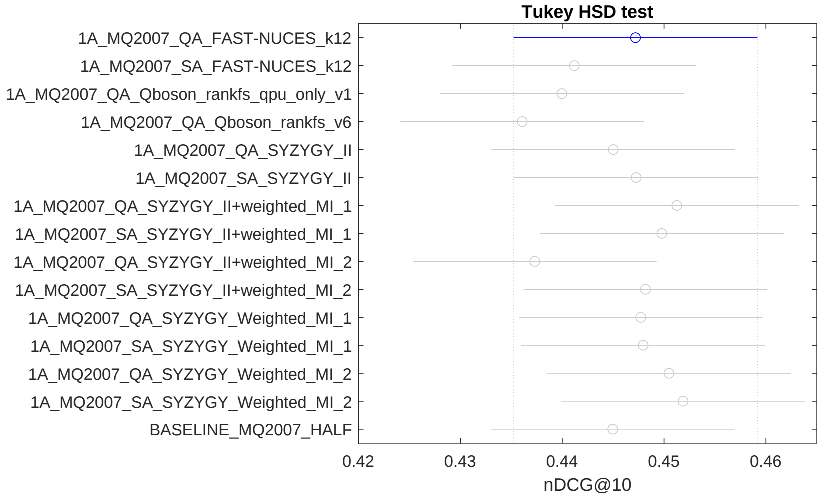
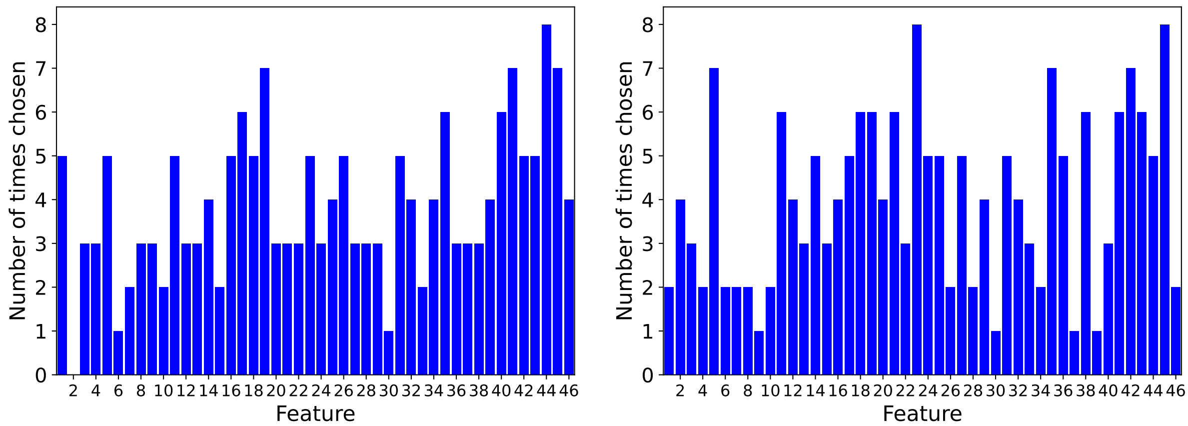


Figure 2: The Tukey HSD test considering the nDCG@10 values associated with different runs and queries for the MQ2007 dataset.



(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.

requiring lower computational effort. The best-performing configuration was proposed by team **R2AI**, which obtained an nDCG@10 score of 0.0396 while retaining 380 features, outperforming the baseline by a substantial margin. At the same time, submission `1B_400_ICM_QA_FAST-NUCES_Ensemble` from team **FAST-NUCES** achieved a score comparable to the baseline while retaining only 150 features, demonstrating a significant reduction in feature dimensionality without sacrificing effectiveness.

From a methodological perspective, the participating teams explored a variety of feature-selection strategies. Team **FAST-NUCES** adopted a QUBO formulation that combines collaborative relevance, separability, coverage, and redundancy signals extracted from both the Item Content Matrix and the User Rating Matrix [44]. In contrast, team **Skywalkers** developed a hybrid Bayesian-quantum pipeline that integrates Optuna-based hyperparameter optimization with a QUBO formulation leveraging Random Forest relevance scores and cosine-similarity redundancy penalties. Their approach automatically calibrates penalty weights, eliminating the need for manual tuning [47]. Team **SYZYGY**, instead, proposed a collaborative-driven QUBO formulation designed to align content-based similarity with user-interaction similarity through an agreement mask that is subsequently projected back into the feature space [48]. Finally, team **R2AI** introduced a performance-oriented QUBO framework in which both unary and pairwise coefficients are directly derived from the observed counterfactual degradation

Table 6
Task 1B results on the 100_ICM dataset. An adjacent couple of rows (marked with the same color) represents the results achieved with QA/H and SA using the same problem formulation. Results marked in yellow(●) refer to the baselines' results.

Team	Submission id	nDCG@10	Annealing time (ms)	Type	N° features
Skywalkers	1B_100_ICM_QA_skywalkers_003	0.0185	272	Q	59
Skywalkers	1B_100_ICM_SA_skywalkers_003	0.0254	9862	S	59
Skywalkers	1B_100_ICM_QA_skywalkers_010	0.0211	229	Q	59
Skywalkers	1B_100_ICM_SA_skywalkers_010	0.0277	9377	S	59
Skywalkers	1B_100_ICM_QA_skywalkers_011	0.0166	270	Q	59
Skywalkers	1B_100_ICM_SA_skywalkers_011	0.0269	9381	S	59
Skywalkers	1B_100_ICM_QA_skywalkers_012	0.0169	274	Q	59
Skywalkers	1B_100_ICM_SA_skywalkers_012	0.0287	9653	S	59
Skywalkers	1B_100_ICM_QA_skywalkers_016	0.0155	255	Q	48
Skywalkers	1B_100_ICM_SA_skywalkers_016	0.0264	20509	S	48
FAST-NUCES	1B_100_ICM_QA_FAST-NUCES_Ensemble	0.0218	456	Q	72
FAST-NUCES	1B_100_ICM_SA_FAST-NUCES_Ensemble	0.0229	15486	S	75
SYZYGY	1B_100_ICM_QA_SYZYGY_Standard	0.0152	551	Q	62
SYZYGY	1B_100_ICM_SA_SYZYGY_Standard	0.0243	14245	S	85
R2AI	1B_100_ICM_QA_r2ai_001	0.0238	2990	H	95
R2AI	1B_100_ICM_SA_r2ai_001	0.0241	24469	S	95
R2AI	1B_100_ICM_QA_r2ai_002	0.0224	2998	H	92
R2AI	1B_100_ICM_SA_r2ai_002	0.0211	25388	S	92
R2AI	1B_100_ICM_QA_r2ai_003	0.0245	2995	H	97
R2AI	1B_100_ICM_SA_r2ai_003	0.0256	22949	S	97
R2AI	1B_100_ICM_QA_r2ai_004	0.0233	2996	H	90
R2AI	1B_100_ICM_SA_r2ai_004	0.0218	26345	S	90
R2AI	1B_100_ICM_QA_r2ai_005	0.0243	2990	H	96
R2AI	1B_100_ICM_SA_r2ai_005	0.0244	21390	S	96
BASELINE	ALL_FEATURES	0.0226	-	-	100

Table 7
Task 1B results on the 400_ICM dataset. An adjacent couple of rows (marked with the same color) represents the results achieved with QA/H and SA using the same problem formulation. Results marked in yellow(●) refer to the baselines' results.

Team	Submission id	nDCG@10	Annealing time (ms)	Type	N° features
FAST-NUCES	1B_400_ICM_QA_FAST-NUCES_Ensemble	0.0307	37	Q	150
FAST-NUCES	1B_400_ICM_SA_FAST-NUCES_Ensemble	0.0289	129408	S	200
R2AI	1B_400_ICM_QA_r2ai_001	0.0374	2995	H	370
R2AI	1B_400_ICM_SA_r2ai_001	0.0344	703620	S	370
R2AI	1B_400_ICM_QA_r2ai_002	0.0382	2985	H	375
R2AI	1B_400_ICM_SA_r2ai_002	0.0340	678943	S	375
R2AI	1B_400_ICM_QA_r2ai_003	0.0376	2986	H	390
R2AI	1B_400_ICM_SA_r2ai_003	0.0316	229985	S	390
R2AI	1B_400_ICM_QA_r2ai_004	0.0396	2999	H	380
R2AI	1B_400_ICM_SA_r2ai_004	0.0341	651264	S	380
R2AI	1B_400_ICM_QA_r2ai_005	0.0363	2989	H	385
R2AI	1B_400_ICM_SA_r2ai_005	0.0338	651264	S	385
BASELINE	ALL_FEATURES	0.0328	-	-	400

in recommendation quality caused by removing individual features or pairs of features [46].

Overall, the results demonstrate that SA can successfully reduce the number of retained features while preserving competitive recommendation quality, particularly when the reduction remains moderate. Nevertheless, more aggressive feature-selection configurations may result in a substantial decline in performance. In addition, the method is associated with non-negligible computational costs, especially when applied to larger datasets. These findings emphasize the need to carefully balance recommendation effectiveness, computational efficiency, and the desired degree of feature reduction when deploying feature-selection techniques in practical recommender-system scenarios.

Table 8
Results for Task 2. An adjacent couple of rows (marked with the same color) represents the results achieved with QA/H and SA using the same problem formulation. Results marked in yellow(●) refer to the baselines' results.

Group	Submission id	Macro F1 Avg	Avg Reduction	Fine-Tuning + Prediction Time (s)	Annealing Time (ms)	Type
Entropy Of Void	QA_entropyOfVoid_QuboKMedoids	63.9(6.7)	59.83%	819.3(5.4)	21573	Q
Entropy Of Void	SA_entropyOfVoid_QuboKMedoids	65.9(10)	59.99%	816.8(0.5)	6392048	S
FAST-NUCES	QA_FAST-NUCES_DivideMix	58.5(11.9)	60.98%	797.4(11.1)	2643	Q
FAST-NUCES	SA_FAST-NUCES_DivideMix	56.3(9.1)	70.04%	642.4(28.6)	2918988	S
FAST-NUCES	QA_FAST-NUCES_Algorithm1	-	-	-	-	Q
FAST-NUCES	SA_FAST-NUCES_Algorithm1	57.9(8.8)	75.03%	541.7(0.3)	167180	S
GPLSI	QA_gplsi_LocalSets-step-1	55.5(8.7)	49.57%	1007(359.6)	997	Q
GPLSI	SA_gplsi_LocalSets-step-1	67.9(10.5)	12.74%	1675.5(330.5)	21914	S
GPLSI	QA_gplsi_LocalSets-step-2	55.5(8.7)	49.57%	1060.2(406.2)	1043	Q
GPLSI	SA_gplsi_LocalSets-step-2	67.9(10.5)	12.74%	1675.3(330.2)	25083	S
GPLSI	QA_gplsi_stratified-partitioning-RL-simulations-1000	49.8(9.7)	82.37%	410.7(2.9)	1404	Q
GPLSI	SA_gplsi_stratified-partitioning-RL-simulations-1000	50.2(2.8)	82.43%	409.6(2.0)	91374	S
GPLSI	QA_gplsi_stratified-partitioning-RL-simulations-2000	50.4(7.7)	82.30%	412.1(5.4)	2728	Q
GPLSI	SA_gplsi_stratified-partitioning-RL-simulations-2000	53.3(10.1)	82.42%	409.7(3.9)	18281	S
R2AI	QA_r2ai_001	75.7(3.7)	26.02%	1458.3(235.5)	191482	H
R2AI	SA_r2ai_001	67.8(8.9)	26.02%	1517.7(306.5)	976909	S
DS@GT QuantumCLEF	QA_ds-at-gt-qclef_svc_cluster_rep	64.8(8.9)	67.62%	678.7(75.2)	88104	Q
DS@GT QuantumCLEF	SA_ds-at-gt-qclef_svc_cluster_rep	60.4(7.4)	71.44%	609.8(3.4)	551114	S
BASELINE	Baseline_without_noise	88.9(0.8)	-	1997.3(5.7)	-	-
BASELINE	Baseline_with_noise	84.3(2.5)	-	1904.3(1.4)	-	-

5.3. Task 2: Quantum Instance Selection and Noise Reduction

Here we present the results achieved by the teams participating in Task 2. The participating teams investigated a diverse set of instance-selection strategies relying on different formulations:

- Team **GPLSI** introduced two complementary methodologies. The first combines the LocalSets geometric instance-selection algorithm with an SVM-based margin recovery stage and a subsequent QUBO refinement step. The second follows a compression-oriented pipeline that integrates stratified topological partitioning with a Deep Q-Network reinforcement learning agent, followed by a QUBO-based spatial diversification phase [45].

- Team **FAST-NUCES** proposed a noise-aware QUBO framework that incorporates GMM-based reliability estimates, predictive uncertainty measures, and local density information into a divide-and-mix composite scoring mechanism. The resulting optimization problem is solved over class-balanced batches to improve scalability and robustness [44].

- Team **DS@GT QuantumCLEF** developed a cluster-aware QUBO formulation characterized by a within-cluster exponential redundancy penalty and a representativeness-oriented diagonal term that combines SVM margin information with inverse distances from class centroids. Their approach additionally employs a Cleanlab confidence-learning filtering stage before constructing the final QUBO problem [42].

- Team **Entropy of Void** adopted a multi-objective QUBO framework integrating class-frequency priors, kNN-derived noise estimates, same-class diversity rewards, and decision-boundary preservation terms. To accommodate the limitations of current QPU topologies, the optimization problem is decomposed into smaller cluster-specific sub-problems [43].

- Team **R2AI** formulated the instance-selection task as a unified QUBO optimization problem that simultaneously captures four complementary objectives: noise robustness, representativeness, boundary awareness, and class balance. The resulting formulation is optimized through a two-stage procedure combining a global Simulated Annealing search with a local Hybrid Quantum Annealing refinement phase [46].

Table 8 reports the performance achieved by the participating teams on the Vader dataset for Task 2. This task attracted the largest number of participants among all the proposed challenges, receiving submissions from five teams. The evaluated approaches explored markedly different reduction levels,

Table 9
Task 3 results for 10 centroids. An adjacent couple of rows (marked with the same color) represents the results achieved with QA/H and SA using the same problem formulation. Results marked in yellow(●) refer to the baselines' results.

Team	Submission id	nDCG@10	DBI	Annealing time (ms)	Type	N° centroids
GPLSI	10_QPU_gplsi_FPS001	0.4862	7.1169	448	Q	10
GPLSI	10_SA_gplsi_FPS001	0.4654	7.2543	13703	S	10
GPLSI	10_QPU_gplsi_QAL003	0.4857	6.9728	490	Q	10
GPLSI	10_SA_gplsi_QAL003	0.5562	7.8100	19506	S	10
GPLSI	10_QPU_gplsi_QAL004	0.5003	7.0393	545	Q	10
GPLSI	10_SA_gplsi_QAL004	0.4956	6.8404	25766	S	10
GPLSI	10_QPU_gplsi_QMGC005	0.4832	6.8672	404	Q	10
GPLSI	10_SA_gplsi_QMGC005	0.5628	6.9122	8425	S	10
GPLSI	10_QPU_gplsi_QMGC006	0.5506	6.6033	407	Q	10
GPLSI	10_SA_gplsi_QMGC006	0.5564	6.7533	7220	S	10
FAST-NUCES	10_QA_FAST-NUCES_FPS-QUERYAWARE	0.5497	7.0623	16	Q	10
FAST-NUCES	10_SA_FAST-NUCES_FPS-QUERYAWARE	0.5709	6.6164	11819	S	10
FAST-NUCES	-	-	-	-	Q	10
FAST-NUCES	10_SA_FAST-NUCES_KMEDOIDS	0.5489	6.9336	2278	S	10
FAST-NUCES	-	-	-	0	Q	10
FAST-NUCES	10_SA_FAST-NUCES_KMEDOIDS_1	0.5901	7.2754	7882	S	10
FAST-NUCES	-	-	-	-	Q	10
FAST-NUCES	10_SA_FAST-NUCES_KMEDOIDS_QA	0.5901	7.2754	7979	S	10
Entropy of Void	10_QA_entropyOfVoid_G40noPCAeuc	0.5529	7.0068	702	Q	10
Entropy of Void	10_SA_entropyOfVoid_G40noPCAeuc	0.5914	6.8649	7205	S	10
Entropy of Void	10_QA_entropyOfVoid_G80noPCAcorr	0.5579	7.0836	859	Q	10
Entropy of Void	10_SA_entropyOfVoid_G80noPCAcorr	0.5357	7.4371	20577	S	10
Entropy of Void	10_QA_entropyOfVoid_G80noPCAcos	0.5948	6.3292	850	Q	10
Entropy of Void	10_SA_entropyOfVoid_G80noPCAcos	0.5357	7.4371	20532	S	10
Entropy of Void	10_QA_entropyOfVoid_G80noPCAeuc	0.6363	6.4572	848	Q	10
Entropy of Void	10_SA_entropyOfVoid_G80noPCAeuc	0.5357	7.4371	20629	S	10
Entropy of Void	10_QA_entropyOfVoid_G80P150euc	0.5164	6.9788	857	Q	10
Entropy of Void	10_SA_entropyOfVoid_G80P150euc	0.5328	6.8383	21634	S	10
BASELINE	BASELINE_10	0.5509	7.9892	-	-	10

ranging from approximately 13% to more than 82%, thereby reflecting distinct trade-offs between dataset compression and downstream classification performance. Overall, all instance-selection strategies resulted in a reduction of the Macro F1 score achieved by the fine-tuned Llama3.1 7B model when compared with the two baseline configurations trained on the complete dataset, both with and without artificially injected noise.

Among all submissions, the best overall result was obtained by the R2AI configuration (`QA_r2ai_001`), which achieved a Macro F1 score of 75.7 while reducing the dataset by 26.02%. Although the reduction rate is relatively moderate, the method was able to preserve the highest level of predictive performance among all competing approaches. In contrast, methods targeting substantially higher reduction rates, typically between 70% and 82%, generally exhibited a more pronounced decrease in effectiveness, indicating that aggressive data compression can negatively affect the quality of the resulting model. Nevertheless, several approaches, including those proposed by GPLSI and Entropy of Void, succeeded in maintaining competitive performance despite removing more than half of the available training instances.

A comparison between the QA/H and SA variants reveals largely consistent effectiveness trends across the different formulations. Although SA achieved better predictive performance in several configurations, the overall differences between the optimization strategies remained relatively limited. At the same time, QA consistently required substantially lower annealing times than SA, often reducing computational costs by one or more orders of magnitude. These findings highlight the computational efficiency of QA-based optimization while demonstrating its ability to maintain predictive performance comparable to that achieved by classical annealing approaches.

Table 10
Task 3 results for 25 centroids. An adjacent couple of rows (marked with the same color) represents the results achieved with QA/H and SA using the same problem formulation. Results marked in yellow(●) refer to the baselines' results.

Team	Submission id	nDCG@10	DBI	Annealing time (ms)	Type	N° centroids
GPLSI	25_QPU_gplsi_FPS001	0.4450	5.7445	433	Q	25
GPLSI	25_SA_gplsi_FPS001	0.4413	5.9002	17131	S	25
GPLSI	25_QPU_gplsi_QAL003	0.4435	5.8082	490	Q	25
GPLSI	25_SA_gplsi_QAL003	0.5071	5.6069	24311	S	25
GPLSI	25_QPU_gplsi_QAL004	0.5054	5.7333	560	Q	25
GPLSI	25_SA_gplsi_QAL004	0.4861	5.5941	31647	S	25
GPLSI	25_QPU_gplsi_QMGC005	0.5513	5.2352	327	Q	25
GPLSI	25_SA_gplsi_QMGC005	0.5799	5.3022	10172	S	25
GPLSI	25_QPU_gplsi_QMGC006	0.5448	5.1040	344	Q	25
GPLSI	25_SA_gplsi_QMGC006	0.5165	5.2551	8568	S	25
FAST-NUCES	25_QA_FAST-NUCES_FPS-QUERYAWARE	0.4793	5.6272	16	Q	25
FAST-NUCES	25_SA_FAST-NUCES_FPS-QUERYAWARE	0.5173	5.4729	14503	S	25
FAST-NUCES	-	-	-	-	Q	25
FAST-NUCES	25_SA_FAST-NUCES_KMEDOIDS	0.4592	6.9607	182119	S	25
FAST-NUCES	-	-	-	-	Q	25
FAST-NUCES	25_SA_FAST-NUCES_KMEDOIDS_1	0.5263	5.7743	7767	S	25
FAST-NUCES	-	-	-	-	Q	25
FAST-NUCES	25_SA_FAST-NUCES_KMEDOIDS_QA	0.4804	5.8211	7799	S	25
Entropy of Void	25_QA_entropyOfVoid_G60noPCAeuc	0.5229	5.8882	793	Q	25
Entropy of Void	25_SA_entropyOfVoid_G60noPCAeuc	0.5652	5.3448	15938	S	25
Entropy of Void	25_QA_entropyOfVoid_G80noPCAcorr	0.5286	5.4832	855	Q	25
Entropy of Void	25_SA_entropyOfVoid_G80noPCAcorr	0.5505	5.3252	25048	S	25
Entropy of Void	25_QA_entropyOfVoid_G80noPCAcos	0.5858	5.0280	843	Q	25
Entropy of Void	25_SA_entropyOfVoid_G80noPCAcos	0.5505	5.3252	25032	S	25
Entropy of Void	25_QA_entropyOfVoid_G80noPCAeuc	0.5248	5.6329	864	Q	25
Entropy of Void	25_SA_entropyOfVoid_G80noPCAeuc	0.5505	5.3252	24929	S	25
Entropy of Void	25_QA_entropyOfVoid_G80P150euc	0.5512	5.4883	864	Q	25
Entropy of Void	25_SA_entropyOfVoid_G80P150euc	0.5078	5.2803	25497	S	25
BASELINE	BASELINE_25	0.5284	6.1201	-	-	25

5.4. Task 3: Quantum Clustering Results

In this section, we present the results obtained by the teams participating in Task 3. Tables 9, 10, and 11 summarize the outcomes achieved when clustering the document representations into 10, 25, and 50 centroids, respectively.

Overall, the submitted solutions demonstrate that both quantum-based and classical optimization techniques can effectively support clustering-based retrieval tasks. The retrieval performance, however, is influenced by the specific clustering formulation adopted. Several submissions achieved competitive results in terms of both nDCG@10 and Davies-Bouldin Index (DBI), indicating that the resulting clusters were successful in capturing representative and meaningful document groups that support effective retrieval.

Among the participating teams, **Entropy of Void** delivered some of the strongest results across all centroid configurations. In particular, the submission `10_QA_entropyOfVoid_G80noPCAeuc` achieved the highest overall nDCG@10 score, reaching 0.6363 in the 10-centroid configuration. The same team also maintained competitive performance when the number of centroids was increased to 25 and 50, demonstrating a high level of robustness across different clustering granularities. Team **GPLSI**, in contrast, produced more balanced results across a variety of formulations. Although their submissions did not always achieve the highest retrieval effectiveness, they frequently obtained competitive DBI values while maintaining stable nDCG@10 performance. Notably, in the 50-centroid setting, configurations such as `50_QPU_gplsi_QMGC006` achieved particularly favorable trade-offs between retrieval effectiveness and clustering compactness.

A recurring pattern across all experimental settings concerns the computational requirements of the different optimization approaches. In general, SA-based solutions required substantially longer

Table 11
Task 3 results for 50 centroids. An adjacent couple of rows (marked with the same color) represents the results achieved with QA/H and SA using the same problem formulation. Results marked in yellow(●) refer to the baselines' results.

Team	Submission id	nDCG@10	DBI	Annealing time (ms)	Type	N° centroids
GPLSI	50_QPU_gplsi_FPS001	0.4483	4.6322	426	Q	50
GPLSI	50_SA_gplsi_FPS001	0.4608	4.8406	16853	S	50
GPLSI	50_QPU_gplsi_QAL003	0.4522	4.9158	476	Q	50
GPLSI	50_SA_gplsi_QAL003	0.4540	4.9734	29253	S	50
GPLSI	50_QPU_gplsi_QAL004	0.4377	4.9023	558	Q	50
GPLSI	50_SA_gplsi_QAL004	0.4073	4.9647	39437	S	50
GPLSI	50_QPU_gplsi_QMGC005	0.5547	4.2726	455	Q	50
GPLSI	50_SA_gplsi_QMGC005	0.5360	4.4472	11093	S	50
GPLSI	50_QPU_gplsi_QMGC006	0.5593	4.3579	450	Q	50
GPLSI	50_SA_gplsi_QMGC006	0.5509	4.4169	12286	S	50
FAST-NUCES	50_QA_FAST-NUCES_FPS-QUERYAWARE	0.5461	4.5136	16	Q	50
FAST-NUCES	50_SA_FAST-NUCES_FPS-QUERYAWARE	0.5411	4.3671	18008	S	50
FAST-NUCES	-	-	-	-	Q	50
FAST-NUCES	50_SA_FAST-NUCES_KMEDOIDS	0.5052	6.1287	182211	S	50
FAST-NUCES	-	-	-	-	Q	50
FAST-NUCES	50_SA_FAST-NUCES_KMEDOIDS_1	0.4661	5.8774	7978	S	50
FAST-NUCES	-	-	-	-	Q	50
FAST-NUCES	50_SA_FAST-NUCES_KMEDOIDS_QA	0.5158	6.0449	7812	S	50
Entropy of Void	50_QA_entropyOfVoid_G70noPCAeuc	0.5466	4.4982	813	Q	50
Entropy of Void	50_SA_entropyOfVoid_G70noPCAeuc	0.5555	4.5686	25956	S	50
Entropy of Void	50_QA_entropyOfVoid_G80noPCAcorr	0.5361	4.5300	851	Q	50
Entropy of Void	50_SA_entropyOfVoid_G80noPCAcorr	0.5376	4.5169	31640	S	50
Entropy of Void	50_QA_entropyOfVoid_G80noPCAcos	0.5205	4.6600	847	Q	50
Entropy of Void	50_SA_entropyOfVoid_G80noPCAcos	0.5376	4.5169	31216	S	50
Entropy of Void	50_QA_entropyOfVoid_G80noPCAeuc	0.5287	4.3410	854	Q	50
Entropy of Void	50_SA_entropyOfVoid_G80noPCAeuc	0.5376	4.5169	31638	S	50
Entropy of Void	50_QA_entropyOfVoid_G80P150euc	0.5406	4.7471	847	Q	50
Entropy of Void	50_SA_entropyOfVoid_G80P150euc	0.5453	4.5317	31576	S	50
BASELINE	BASELINE_50	0.4656	5.3679	-	-	50

execution times than their corresponding QA/QPU counterparts. Nevertheless, this additional computational effort did not consistently translate into superior retrieval performance. In several cases, the quantum-based submissions achieved retrieval effectiveness comparable to, or even exceeding, that of the corresponding SA variants while requiring significantly lower execution times.

The experiments conducted with varying numbers of centroids further illustrate the trade-off between retrieval granularity and cluster representativeness. Increasing the number of centroids naturally leads to the formation of smaller and more specialized clusters, thereby reducing the number of candidate documents associated with each query. Although this behavior can improve cluster compactness and separation, it may simultaneously reduce retrieval diversity and decrease the probability of retrieving highly relevant documents. Consequently, improvements in clustering quality, as measured by internal clustering metrics such as DBI, do not necessarily correspond to proportional gains in retrieval effectiveness measured by nDCG@10.

The participating teams explored a range of clustering methodologies, each relying on different mechanisms for representative selection and cluster construction:

- Team **GPLSI** proposed a unified quantum-inspired clustering framework centered on a shared Bauckhage-style k-medoids QUBO formulation. Within this framework, the team evaluated four alternative pivot-selection strategies: Farthest-Point Sampling, Query-Aligned Farthest-Point Sampling, Query-aware Multilevel Graph Clustering, and Quantum Clustering via Spectral Seeding [45].

- Team **FAST-NUCES** adopted a query-aware clustering strategy that combines document relevance and local-density information with a query-biased Farthest-Point Sampling procedure for pivot

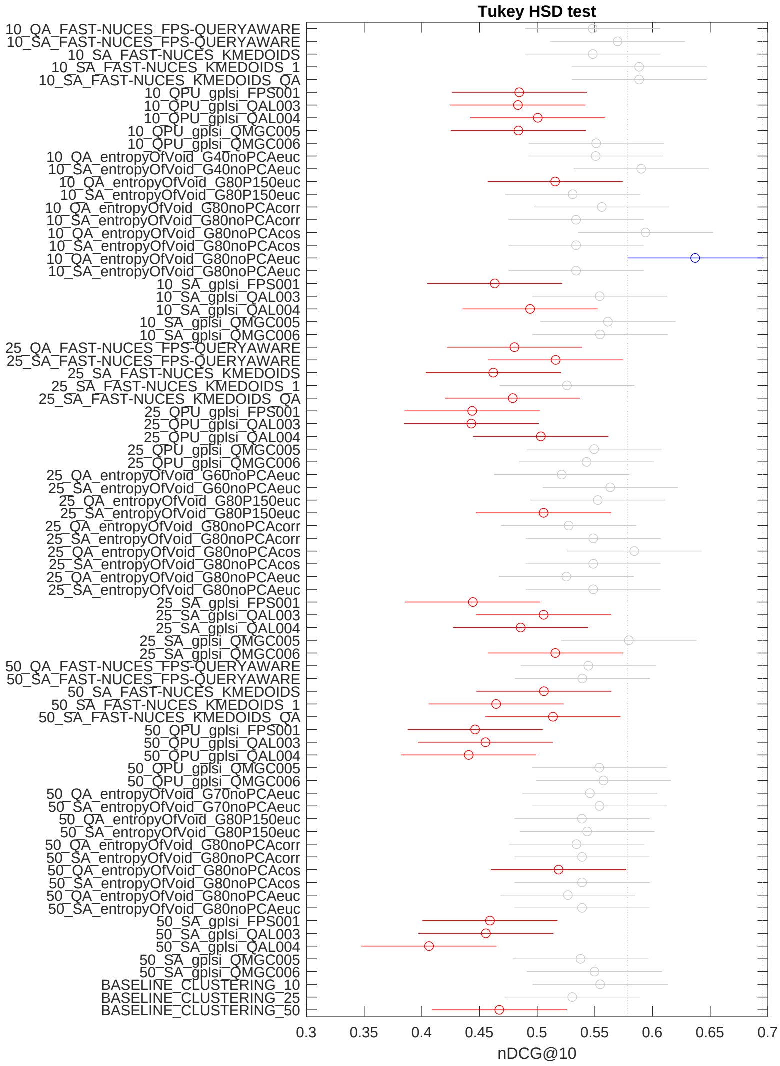


Figure 4: The Tukey HSD test considering the nDCG@10 values associated with different runs and queries for the Clustering dataset.

reduction. The selected pivots are subsequently processed through a QUBO-based k-medoids optimization stage followed by nearest-centroid assignment [44].

- Team **Entropy of Void** employed a three-stage hierarchical pipeline. The approach first compresses the full document collection using MiniBatchKMeans, then selects two representative documents from each macro-cluster, and finally applies a QUBO-based k-medoids formulation to the compressed representative set. To improve optimization stability across different configurations, the method relies on correntropy-bounded distances when constructing the optimization problem [43].

Figure 4 reports the statistical analysis of the clustering results using two-way ANOVA [49] and Tukey HSD post-hoc test [50]. The Tukey HSD test indicates statistically significant differences among the various team submissions in terms of nDCG@10, clearly indicating that some runs considering 10 clusters outperformed others that considered 25 or 50 clusters. Moreover, we can also see statistically significant differences among some of the runs considering the same number of clusters. This suggests that different approaches and QUBO formulations can substantially impact the overall effectiveness.

6. Conclusions and Future Work

In this paper, we presented an overview of the third edition of the QuantumCLEF lab, which was held in 2026. 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 Noise Reduction, and Clustering, all computationally intensive problems commonly encountered in IR and RS systems.

Participants relied on the KIMERA infrastructure [7], which facilitated the workflow and granted a high level of comparability and reproducibility. 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 8 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.

Declaration on Generative AI

We disclose that generative AI technologies were used solely to assist in grammar checking and writing 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, M. F. Dacrema, W. Cunha, M. A. Gonçalves, P. Cremonesi, N. Ferro, Quantumclef 2025: Overview of the second quantum computing challenge for information retrieval and recommender systems at CLEF, in: G. Faggioli, N. Ferro, P. Rosso, D. Spina (Eds.), Working Notes of the Conference and Labs of the Evaluation Forum, CLEF 2025, Madrid, Spain, 9-12 September 2025, CEUR Workshop Proceedings, CEUR-WS.org, 2025, pp. 4086–4104. URL: https://ceur-ws.org/Vol-4038/paper_338.pdf.

[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] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.
[28] 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.
[29] 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.
[30] C. J. C. Burges, From RankNet to LambdaRank to LambdaMART: An Overview, Technical Report, Microsoft Research, MSR-TR-2010-82, 2010.
[31] T. Qin, T. Liu, Introducing LETOR 4.0 datasets, CoRR abs/1306.2597 (2013). URL: http://arxiv.org/abs/1306.2597. arXiv:1306.2597.
[32] 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.
[33] 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.
[34] W. Cunha, C. França, G. Fonseca, L. Rocha, M. A. Gonçalves, An effective, efficient, and scalable confidence-based instance selection framework for transformer-based text classification, in: H. Chen, W. E. Duh, H. Huang, M. P. Kato, J. Mothe, B. Poblete (Eds.), Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23-27, 2023, ACM, 2023, pp. 665–674. URL: https://doi.org/10.1145/3539618.3591638. doi:10.1145/3539618.3591638.
[35] W. Cunha, F. Viegas, C. França, T. Rosa, L. Rocha, M. A. Gonçalves, A comparative survey of instance selection methods applied to non-neural and transformer-based text classification, ACM Comput. Surv. 55 (2023) 265:1–265:52. URL: https://doi.org/10.1145/3582000. doi:10.1145/3582000.
[36] W. Cunha, A. M. Fernández, A. Esuli, F. Sebastiani, L. Rocha, M. A. Gonçalves, A noise-oriented and redundancy-aware instance selection framework, ACM Trans. Inf. Syst. 43 (2025) 45:1–45:33. URL: https://doi.org/10.1145/3705000. doi:10.1145/3705000.
[37] 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.
[38] 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.
[39] 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.
[40] 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.
[41] 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.
[42] P. Awate, Cluster-aware qubo instance selection with representativeness scoring and noise pre-filtering for the quantumclef 2026 sentiment classification task, 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.
[43] G. Jain, H. Hitesh, S. Bedi, S. Sundriyal, Entropy of void at quantumclef 2026: Hardware-aware qubo decomposition via quantum annealing - instance selection and document clustering, 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.
[44] S. Zahid, M. A. Tahir, M. R. Sagri, Qubo based quantum annealing for feature selection, instance selection and clustering, 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.
[45] S. Galiano-Segura, F. Valero-Abellón, M. Sánchez-Bellido, J. Madrigal-Cutillas, J. P. Consuegra-Ayala, O. Francés, M. Palomar, P. Moreda, E. Lloret, Team gplsi at qclef 2026: Quantum-inspired instance selection and clustering, 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.
[46] J. Niu, Y. Wang, J. Li, K. Deng, A. Alavi, Y. Ren, Team r2ai at quantumclef 2026: Interaction-aware feature selection and complementary-signal instance selection via qubo optimization, 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.
[47] B. Prajapat, A. Kuila, Balancing relevance and redundancy: An optuna-tuned qubo formulation for feature selection in recommender systems, 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.
[48] D. Khemani, B. Kantz, Team syzygy at quantumclef 2026: Feature selection using quantum annealing, 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.
[49] L. St, S. Wold, et al., Analysis of variance (anova), Chemometrics and intelligent laboratory systems 6 (1989) 259–272.
[50] J. W. Tukey, Comparing individual means in the analysis of variance, Biometrics (1949) 99–114.