QuantumCLEF 2025 - The Second Edition of the Quantum Computing Lab at CLEF

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

[1] University of Padua, Padua, Italy
andrea.pasin.1@phd.unipd.it, nicola.ferro@unipd.it
[2] Politecnico di Milano, Milan, Italy
{maurizio.ferrari, paolo.cremonesi}@polimi.it
[3] Federal University of Minas Gerais, Belo Horizonte, Brazil
{washingtoncunha, mgoncalv}@dcc.ufmg.br

Abstract. Over the last few years, *Quantum Computing (QC)* has captured the attention of numerous researchers from different fields since QC resources have become more applicable in solving practical problems. In the current landscape, *Information Retrieval (IR)* and *Recommender Systems (RS)* need to perform computationally intensive operations on massive and heterogeneous datasets. Therefore, it could be possible to use QC technologies such as *Quantum Annealing (QA)* to boost systems' performance. The objective of this work is to present the second edition of the QuantumCLEF lab, which is composed of three tasks that aim at discovering and evaluating QA approaches compared to their traditional counterpart while also establishing collaborations among researchers from different fields to harness their knowledge and skills to solve the considered challenges and promote the usage of QA.
This lab will allow participants to use **real quantum computers** provided by CINECA, one of the most important computing centers worldwide.

1 Introduction

In the current challenging scenario where *Information Retrieval (IR)* and *Recommender Systems (RS)* systems face ever-increasing amounts of data and rely on computationally demanding approaches, *Quantum Computing (QC)* can be used to improve their performance. Although QC has already been applied in several domains, limited work has been done specifically for the IR and RS fields [6, 14, 19]. Indeed, the area of IR called Quantum IR [12, 23, 25] consists of exploiting the concepts of quantum mechanics to formulate IR models and problems but it does not deal with implementing IR and RS models and algorithms via QC technologies.
In this work we focus on *Quantum Annealing (QA)*, which exploits special-purpose devices able to rapidly find optimal solutions to optimization problems by leveraging quantum-mechanical effects. Our goal is to understand if QA can

improve the efficiency and effectiveness of IR and RS systems. So, we present the second edition of the evaluation lab called *QuantumCLEF (qCLEF)*[4] [16], which aims at:

- evaluating the performance of QA with respect to traditional approaches;
- identifying new ways of formulating IR and RS algorithms and methods, so that they can be solved with QA;
- growing a research community around this new field in order to promote a wider adoption of QC technologies for IR and RS.

Working with QA does not require particular knowledge about how quantum physics works underneath it. There are in fact available tools and libraries that can be easily used to program and solve problems through this paradigm.

The paper is organized as follows: Section 2 introduces related work; Section 3 presents the tasks in the qCLEF lab; Section 4 considers some critical evaluation aspects; Section 5 shows the design of the infrastructure for the lab; finally, Section 6 draws some conclusions and outlooks some future work.

2 Related Work

What is Quantum Annealing. QA is a QC paradigm that is based on special-purpose devices (quantum annealers) able to tackle optimization problems. A quantum annealer represents a problem as the energy of a physical system and then leverages quantum-mechanical phenomena to let the system find a state of minimal energy, corresponding to the solution of the original problem.

These problems need to be formulated as minimization ones using the *Quadratic Unconstrained Binary Optimization (QUBO)* formulation, defined as follows:

$$\min \quad y = x^T Q x$$

where x is a vector of binary decision variables and Q is a matrix of constant values representing the problem to solve. Through QUBO formulations, it is possible to represent many problems [8]. Then, the *minor embedding* step maps the QUBO problem into the quantum annealer hardware, accounting for its topology. This can be done automatically, relying on some heuristics. A QUBO problem is usually solved by quantum annealers in few *milliseconds*.

Applications of Quantum Annealing. QA can have practical applications in several fields due to its ability to tackle *NP-Hard* integer optimization problems.

QA has been previously applied to tackle IR and RS tasks such as Feature Selection [14], showing feasibility and promising improvements in efficiency and effectiveness. QA has also been applied to *Machine Learning (ML)* tasks. For example, Willsch et al. [26] proposes a formulation of kernel-based *Support Vector Machine (SVM)* on a D-Wave 2000Q quantum annealer, while Delilbasic et al.

[4] https://qclef.dei.unipd.it/

[4] proposes a quantum multiclass SVM formulation aiming to reduce the execution time for large training sets. Other works explore the application of QA to clustering; for example, Zaiou et al. [28] applies it to a balanced K-means method showing better performance according to the Davies-Bouldin Index.

Previous Editions. This is the second edition of the QuantumCLEF Lab. In the previous edition (2024), there were 26 subscribed teams, out of which 7 teams managed to provide official submissions for the tasks. The previous lab was composed of two tasks: Feature Selection and Clustering. The results in the previous edition suggest that quantum annealers are overall able to maintain a comparable level of effectiveness with respect to more traditional approaches (e.g., Simulated Annealing) while being able to solve the problems more efficiently considering just the time required for the annealing phase [17, 18]. In total, there have been 66 official submitted runs. The participating teams solved 976 problems using both traditional and quantum algorithms, for a total execution time of almost 12 hours for traditional solvers and 4 minutes for quantum solvers. We received very positive feedback from the participants, most of which never experienced using QC resources before.

3 Tasks

In the qCLEF lab there are three tasks, each with the following goals:

- find one or more possible QUBO formulations of the problem;
- evaluate the quantum annealer approach compared to a corresponding traditional approach to assess both its efficiency and its effectiveness.

In general, we expect QA to solve problems more quickly than traditional approaches, achieving results that are similar or better in terms of effectiveness.

3.1 Task 1 - Quantum Feature Selection

This task focuses on formulating the well-known *NP-Hard* Feature Selection problem and solving it with QA, similarly to other previous works [6, 14].

Feature Selection is a widespread problem for both IR and RS which requires the identification a subset of the available features (e.g., the most informative, less noisy, etc.) to train a learning model. This problem is very impacting since many IR and RS systems involve the optimization of learning models, and reducing the dimensionality and noise of the input data can improve their performance.

If the input data has n features, we can enumerate all the possible sets of input data having a fixed number k of features, thus obtaining $\binom{n}{k}$ possible subsets. Therefore, to find the best subset of k features the learning model should be trained on all the subsets of features, which is infeasible even for small datasets. So, in this task, we want to understand if QA can be used to solve this problem more efficiently and effectively.

We have identified some possible datasets such as MQ2007 [22] or Istella S-LETOR [11]. These datasets contain pre-computed features and the objective is to select a subset of these features to train a learning model, such as LambdaMART [1] or a content-based RS, and to achieve the best performance according to metrics such as nDCG@10.

3.2 Task 2 - Quantum Instance Selection

This task focuses on formulating the Instance Selection problem to solve it through QA [15].

Currently, transformer-based architectures, including 1st and 2nd generation transformers (e.g., RoBERTa [10]) as well as current large language models (e.g., Llama3 [24]), are used and considered state-of-the-art in several fields. Given the LLMs high-cost application, one of the big challenges is to fine-tune these models efficiently. Instance Selection focuses on selecting a representative subset of instances from a dataset to make the training of these models faster while maintaining a high level of effectiveness of the trained model [2, 3].

In this task, we aim at using QA to find a good subset of instances in a dataset in an efficient way, that allows the fine-tuning a **Llama3.1** model to perform a text classification task as effectively as it would on the entire original dataset.

We have identified some possible datasets such as Vader NYT or Yelp Reviews that will be provided in a five-fold cross-validation split. The extracted subsets will be then used to fine-tune the Llama3.1 model and the effectiveness will be measured with the Macro-F1 score [21].

3.3 Task 3 - Quantum Clustering

This task focuses on the formulation of the clustering problem and solving it with a quantum annealer. Clustering is a relevant problem for IR and RS which involves grouping items together according to their characteristics.

Clustering can be helpful for organizing large collections, helping users to explore a collection and providing similar results to a query. It can also be used to divide users according to their interests or build user models with the cluster centroids [27] boosting efficiency or effectiveness for users with limited data.

There are different clustering problem formulations, such as centroid-based Clustering or Hierarchical Clustering. In this task, each document or user can be represented as a vector in a similarity space and it is possible to cluster documents based on the similarity between each other.

For the IR task, we have identified ANTIQUE [9] as a possible dataset. From the dataset, we will produce embeddings using models such as BERT [5]. The cluster quality will be measured with user queries that undergo the same embedding process. These queries will match only the most representative embeddings of the clusters, avoiding computing similarities on the whole collection. For the recommendation task, the goal will be to partition the users in communities

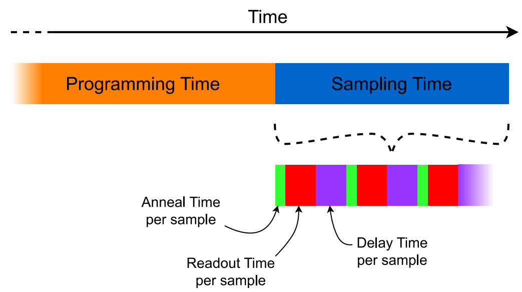

Fig. 1: The quantum annealer access time split in several steps.

based on their past interactions, in such a way that users within a community share similar interests [13]. The quality of the communities will be assessed based on the effectiveness of a non-personalized RS algorithm trained on each community.

The cluster quality will be measured according to the Davies-Bouldin Index and nDCG@10.

4 Evaluation of Quantum Annealing

Using a quantum annealer requires several stages:

Formulation: compute the QUBO matrix Q;
Embedding: generate the *minor embedding* of the QUBO for the hardware;
Data Transfer: transfer the problem and the embedding to the data center that hosts the quantum annealer;
Annealing: run the quantum annealer itself.

Considering effectiveness, there are at least two layers of stochasticity. First, the embedding phase in which heuristic methods transform the QUBO formulation in an equivalent one that will fit in the hardware. This process is not deterministic: it could produce different embeddings for the same problem, that are in principle equivalent but in practice may affect the result. Second, the annealing phase, which samples a low-energy solution. In some cases, many samples might be needed to get a reliable solution. Usually one selects the best solution found, but this may result in experiments with high variance. Therefore statistical evaluation measures are essential.

Considering efficiency, while the annealing phase in which the quantum annealer is actually used may last in the range of *milliseconds*, transferring the problem on the network introduces large delays, and generating the minor embedding may require even minutes for particularly large problems. Furthermore, the runtime can be split into several phases, see Fig. 1: first the device needs to be programmed for the problem, then the quantum-mechanical annealing process is run and lastly the result is read. The annealing process is extremely fast,

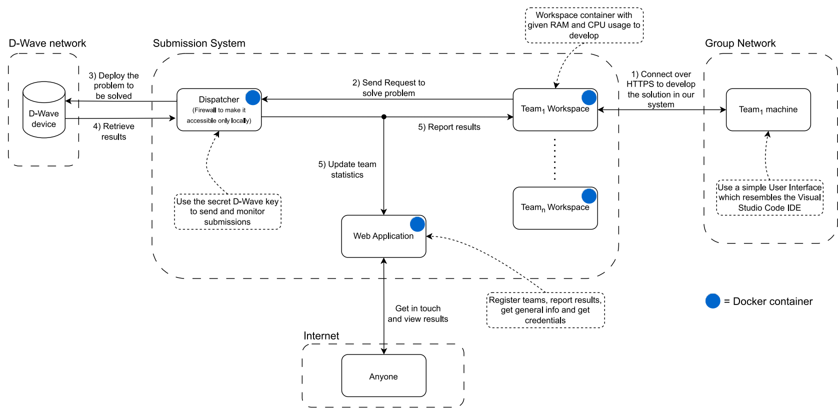

Fig. 2: High-level representation of the infrastructure.

requiring *few microseconds*, but it is repeated multiple times due to the stochastic nature of the device. It is indeed necessary to consider the time requirements of all the steps involved to measure efficiency.

5 QuantumCLEF Infrastructure

We present our custom infrastructure that is required since participants cannot have direct access to quantum annealers and we want measurements to be as fair and reproducible as possible. As depicted in Fig. 2, it is composed of several components with specific purposes:

- **Workspace**: each team has its own workspace which is accessible through the browser by providing the correct credentials. The workspace has a preconfigured git repository that is fundamental for reproducibility reasons.
- **Dispatcher**: it manages and keeps track of all the teams' submissions. It also holds the secret API Key that is used to submit problems to the quantum annealer. In this way, participants will never know the secret Key used.
- **Web Application**: it is the main source of information to the external users about the ongoing tasks. Moreover, it allows teams to view their quotas and some statistics through a dashboard. Also, organizers have their own dashboard through which it is possible to manage teams and tasks.

Through our infrastructure, participants can use real quantum computers. Furthermore, participants do not need any powerful machines since all their approaches will be executed directly in our servers or quantum computers. Our infrastructure plays for QA a role similar to others, such as TIRA [20] or TIREx [7], for more general evaluation purposes. We will use the QC resources provided by CINECA that will make available D-Wave's cutting-edge quantum annealers.

6 Conclusions and Future Work

In this paper we have discussed the qCLEF lab, a lab composed of three practical tasks aiming at evaluating the performance of QA applied to IR and RS. We have also discussed the potential benefits that QA can bring to the IR and RS fields and we have highlighted how the evaluation of both efficiency and effectiveness should be performed. Finally, we have presented an infrastructure designed and implemented to satisfy both participants and organizers' needs.

qCLEF can represent a starting point for many researchers worldwide to know more about these new cutting-edge technologies that will likely have a big impact on the future of several research fields. Through this lab it will also be possible to assess whether QA can be employed to improve the current state-of-the-art approaches, hopefully delivering new performing solutions.

References

1. Burges, C.J.C.: From RankNet to LambdaRank to LambdaMART: An Overview. Tech. rep., Microsoft Research, MSR-TR-2010-82 (June 2010)
2. Cunha, W., França, C., Fonseca, G., Rocha, L., Gonçalves, M.A.: An effective, efficient, and scalable confidence-based instance selection framework for transformer-based text classification. In: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 665–674 (2023)
3. Cunha, W., Viegas, F., França, C., Rosa, T., Rocha, L., Gonçalves, M.A.: A comparative survey of instance selection methods applied to non-neural and transformer-based text classification. ACM Computing Surveys **55**(13s), 1–52 (2023)
4. Delilbasic, A., Saux, B.L., Riedel, M., Michielsen, K., Cavallaro, G.: A single-step multiclass svm based on quantum annealing for remote sensing data classification. arXiv preprint arXiv:2303.11705 (2023)
5. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
6. Ferrari Dacrema, M., Moroni, F., Nembrini, R., Ferro, N., Faggioli, G., Cremonesi, P.: Towards Feature Selection for Ranking and Classification Exploiting Quantum Annealers. In: Proc. 45th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2022), pp. 2814–2824, ACM Press, New York, USA (2022)
7. Fröbe, M., Reimer, J.H., MacAvaney, S., Deckers, N., Reich, S., Bevendorff, J., Stein, B., Hagen, M., Potthast, M.: The Information Retrieval Experiment Platform. In: Chen, H.H., Duh, W.J., Huang, H.H., Kato, M.P., Mothe, J., Poblete, B. (eds.) Proc. 46th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2023), pp. 2826–2836, ACM Press, New York, USA (2023)

8. Glover, F.W., Kochenberger, G.A., Du, Y.: Quantum bridge analytics I: a tutorial on formulating and using QUBO models. 4OR **17**(4), 335–371 (2019), https://doi.org/10.1007/s10288-019-00424-y
9. Hashemi, H., Aliannejadi, M., Zamani, H., Croft, W.B.: Antique: A non-factoid question answering benchmark. In: Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part II 42, pp. 166–173, Springer (2020)
10. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: Roberta: A robustly optimized bert pretraining approach. arXiv preprint 1907.11692 (2019)
11. Lucchese, C., Nardini, F.M., Orlando, S., Perego, R., Silvestri, F., Trani, S.: Post-learning optimization of tree ensembles for efficient ranking. In: Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, pp. 949–952 (2016)
12. Melucci, M.: Introduction to Information Retrieval and Quantum Mechanics, The Information Retrieval Series, vol. 35. Springer (2015)
13. Nembrini, R., Carugno, C., Ferrari Dacrema, M., Cremonesi, P.: Towards recommender systems with community detection and quantum computing. In: RecSys '22: Sixteenth ACM Conference on Recommender Systems, Seattle, WA, USA, September 18 - 23, 2022, pp. 579–585, ACM (2022), https://doi.org/10.1145/3523227.3551478
14. Nembrini, R., Ferrari Dacrema, M., Cremonesi, P.: Feature selection for recommender systems with quantum computing. Entropy **23**(8), 970 (2021)
15. Pasin, A., Cunha, W., Gonçalves, M.A., Ferro, N.: A quantum annealing instance selection approach for efficient and effective transformer fine-tuning. In: Proceedings of the 2024 ACM SIGIR International Conference on Theory of Information Retrieval, pp. 205–214 (2024)
16. Pasin, A., Ferrari Dacrema, M., Cremonesi, P., Ferro, N.: qclef: A proposal to evaluate quantum annealing for information retrieval and recommender systems. In: International Conference of the Cross-Language Evaluation Forum for European Languages, pp. 97–108, Springer (2023)
17. Pasin, A., Ferrari Dacrema, M., Cremonesi, P., Ferro, N.: Overview of quantumclef 2024: The quantum computing challenge for information retrieval and recommender systems at clef. In: International Conference of the Cross-Language Evaluation Forum for European Languages, pp. 260–282, Springer (2024)
18. Pasin, A., Ferrari Dacrema, M., Cremonesi, P., Ferro, N., et al.: Quantumclef 2024: Overview of the quantum computing challenge for information retrieval and recommender systems at clef. In: CEUR WORKSHOP PROCEEDINGS, vol. 3740, pp. 3032–3053, CEUR-WS (2024)
19. Pilato, G., Vella, F.: A survey on quantum computing for recommendation systems. Inf. **14**(1), 20 (2023), https://doi.org/10.3390/info14010020
20. Potthast, M., Gollub, T., Wiegmann, M., Stein, b.: TIRA Integrated Research Architecture. In: Ferro, N., Peters, C. (eds.) Information Retrieval Evaluation in a Changing World – Lessons Learned from 20 Years of CLEF,

The Information Retrieval Series, vol. 41, pp. 123–160, Springer International Publishing, Germany (2019)
21. Powers, D.M.W.: Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. CoRR **abs/2010.16061** (2020), URL https://arxiv.org/abs/2010.16061
22. Qin, T., Liu, T.Y.: Introducing LETOR 4.0 Datasets. arXiv.org, Information Retrieval (cs.IR) **arXiv:1306.2597** (June 2013)
23. van Rijsbergen, C.J.: The Geometry of Information Retrieval. Cambridge University Press, Cambridge, UK (2004)
24. Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., Lample, G.: Llama: Open and efficient foundation language models (2023)
25. Uprety, S., Gkoumas, D., Song, D.: A survey of quantum theory inspired approaches to information retrieval. ACM Computing Surveys (CSUR) **53**(5), 1–39 (2020)
26. Willsch, D., Willsch, M., De Raedt, H., Michielsen, K.: Support vector machines on the d-wave quantum annealer. Computer physics communications **248**, 107006 (2020)
27. Wu, Y., Cao, Q., Shen, H., Tao, S., Cheng, X.: INMO: A model-agnostic and scalable module for inductive collaborative filtering. In: SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, pp. 91–101, ACM (2022), https://doi.org/10.1145/3477495.3532000
28. Zaiou, A., Bennani, Y., Matei, B., Hibti, M.: Balanced k-means using quantum annealing. In: 2021 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1–7 (2021), https://doi.org/10.1109/SSCI50451.2021.9659997