

Reproducibility Report for ACM SIGMOD 2023 Paper: “InfiniFilter: Expanding Filters to Infinity and Beyond”

SONIA HORCHIDAN, KTH Royal Institute of Technology, Sweden
MAJID SAEEDAN, University of California, Riverside, United States

The paper introduces and evaluates InfiniFilter, a novel technique for filter expansion. The experiments included in the paper are reproducible. Our reproduced figures showcase similar trends and relationships between the alternatives compared and support the major findings of the evaluation. We noted variations in the raw latency measurements of some experiments, which can be attributed to hardware differences.

1 INTRODUCTION

This report presents the findings of the reproducibility review conducted for the paper entitled “InfiniFilter: Expanding Filters to Infinity and Beyond” [1] written by Niv Dayan, Ioana Bercea, Pedro Reviriego, and Rasmus Pagh. The authors are affiliated with the University of Toronto, BARC, IT University of Copenhagen, and Universidad Politécnica de Madrid. The work proposes InfiniFilter, a filter-expanding technique that addresses the major drawbacks of current methods. The key figures of the original paper were reproduced successfully.

2 SUBMISSION

The authors provided a link to a publicly available repository (github.com/nivdayan/FilterLibrary). The submission contained the following:

- the source code.
- a detailed README file describing the compilation of the Java project, instructions to run the experiments, and instructions to install the dependencies.
- one Shell script to run all the experiments and to produce the plots (`run_exps.sh`).

The instructions to set up the code are simple and easy to follow. The script `run_exps.sh` executes all the experiments shown in Figures 13 through 16 in the paper. It produces both the plots and six directories, named `exp1` to `exp6`, which contain the raw data measurements. All the dependencies required are clearly stated and trivial to install.

3 HARDWARE AND SOFTWARE ENVIRONMENT

Table 1 presents the hardware specification used in the original paper alongside the one used in our reproducibility assessment.

Table 1. Hardware & Software environment

	Paper	Repro Review 1	Repro Review 2
CPU	2 × Intel Xeon E5-2690v4	2 × Intel Xeon	Apple M2 Max
Cores	2 × 14 (× 2 threads)	2 × 16 (× 2 threads)	12
GHz	2.6	2.30	2.42 - 3.7
RAM	512GB	120GB	64GB
L1 Cache	32KB	512KB	3.3MB
L2 Cache	256KB	4MB	36MB
L3 Cache	35MB	45MB	48MB
Java	11.0.16	11.0.20	21.0.1

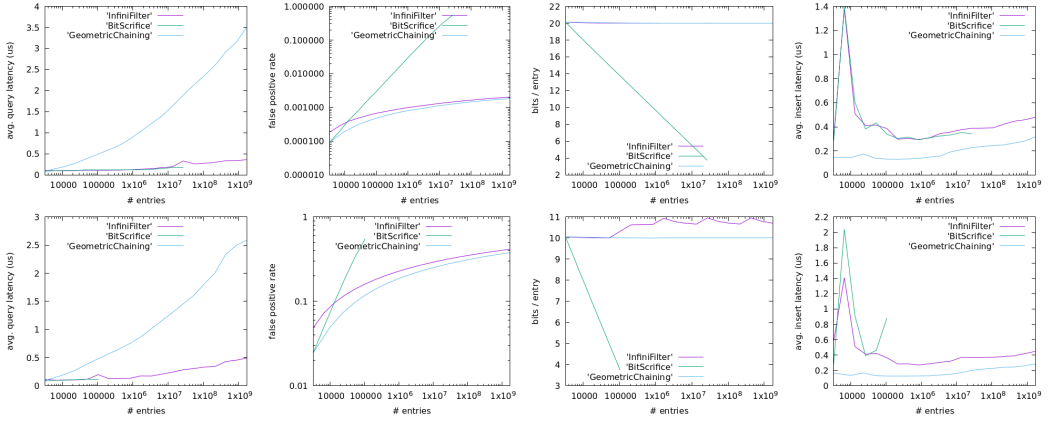


Fig. 1. Reproduction 1 of Figure 13.

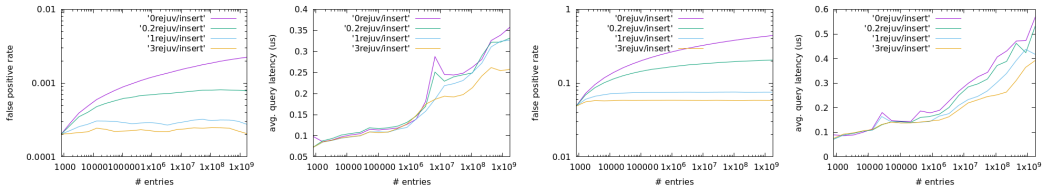


Fig. 2. Reproduction 1 of Figure 14.

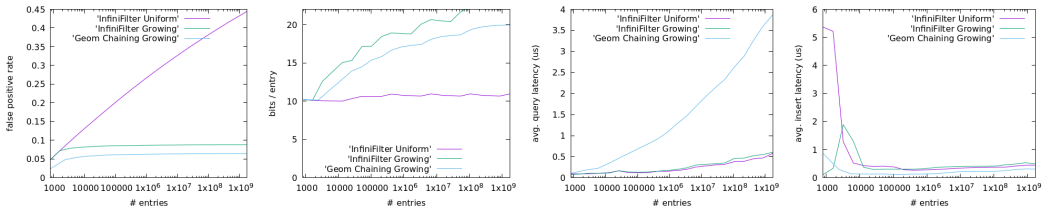


Fig. 3. Reproduction 1 of Figure 15.

4 REPRODUCIBILITY EVALUATION

4.1 Process

We followed the instructions included in the repository. We installed the required dependencies and compiled the library accordingly. The script provided by the authors was then used to execute all the experiments and produce the plots. The reproducibility process ran smoothly and was completed in under 6 hours.

4.2 Results

All the experiments presented in the paper were reproduced, as can be noted in Figures 1, 2, 3, and 4. Our results exhibit the same general trends and confirm the findings of the paper, as the relationships between baselines and alternative methods are consistent with the ones reported by the authors. We only note higher absolute values observed for the experiments measuring the

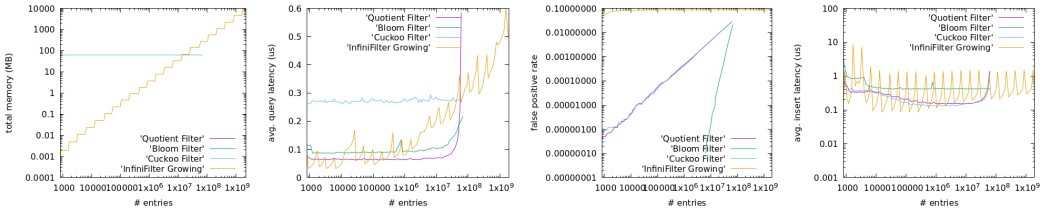


Fig. 4. Reproduction 1 of Figure 16.

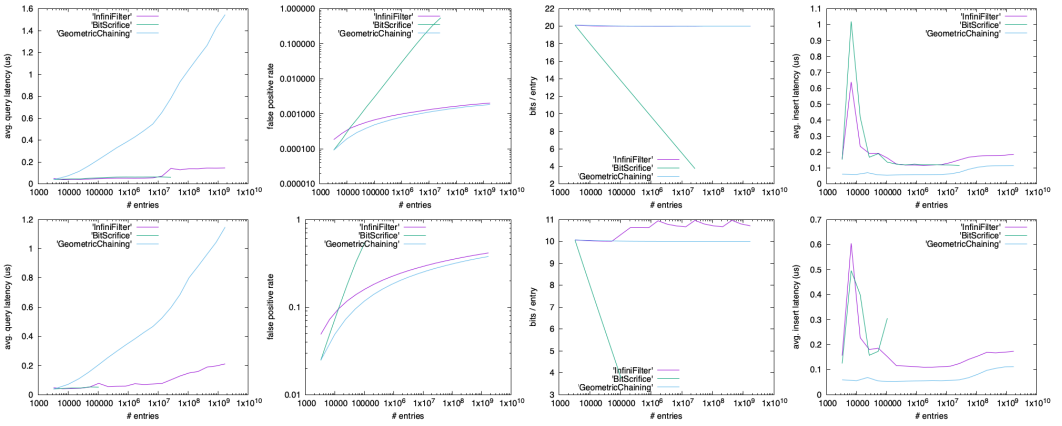


Fig. 5. Reproduction 2 of Figure 13.

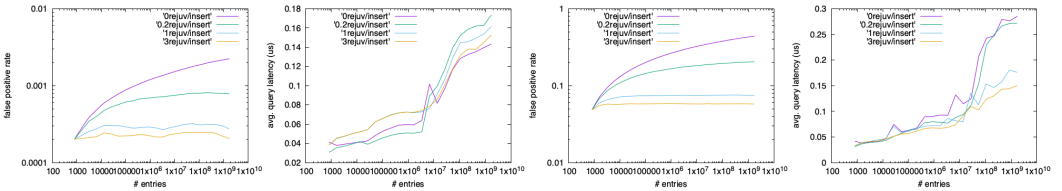


Fig. 6. Reproduction 2 of Figure 14.

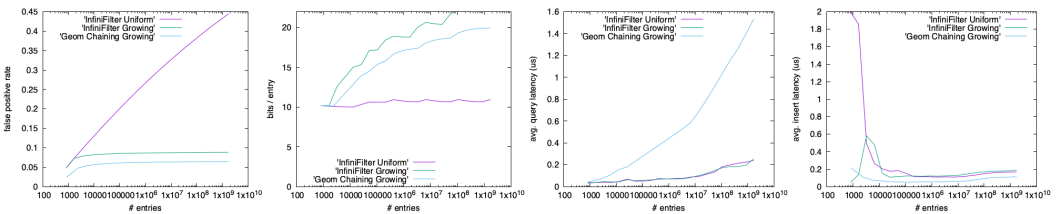


Fig. 7. Reproduction 2 of Figure 15.

average insert latency for a number of entries lower than 10^5 (i.e., Figure 13 D and H, and Figure 15 D in the original paper). The differences are attributed to the use of hardware with different specifications.

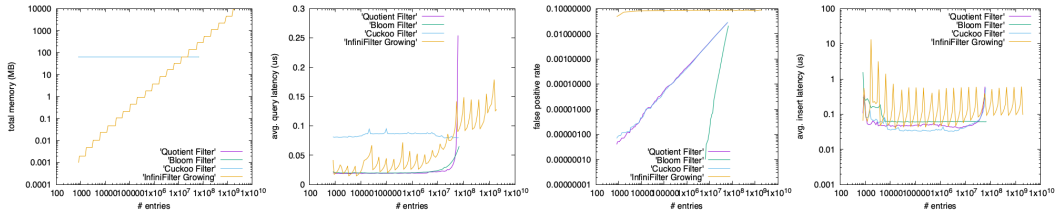


Fig. 8. Reproduction 2 of Figure 16.

The second reproducibility review also produced similar results using different hardware, as shown in Figures 5, 6, 7, and 8. Similarly to the original paper, we used a value of 31 for the "power of the maximum number of slots" (PMS). We test with additional lower values and show the results in the figures included in the Appendix. Changing this parameter still shows similar behavior, with some minor differences.

5 SUMMARY

The reproducibility process produced results consistent with the trends and conclusions presented in the original paper. The authors provided clear instructions, and the reproducibility assessment has been carried out with no issues.

REFERENCES

- [1] Niv Dayan, Ioana O. Bercea, Pedro Reviriego, and Rasmus Pagh. 2023. InfiniFilter: Expanding Filters to Infinity and Beyond. *Proc. ACM Manag. Data* 1, 2 (2023), 140:1–140:27. <https://doi.org/10.1145/3589285>

6 APPENDIX

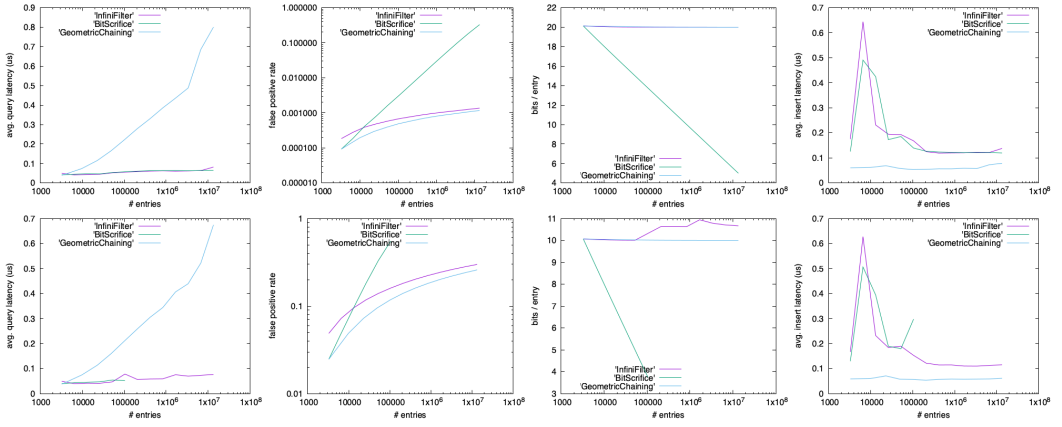


Fig. 9. Reproduction of Figure 13 (PMS=24).

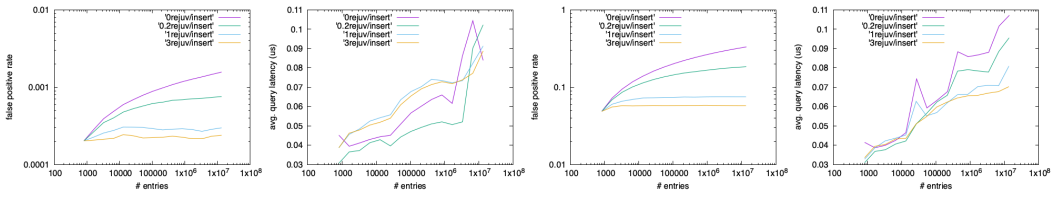


Fig. 10. Reproduction of Figure 14 (PMS=24).

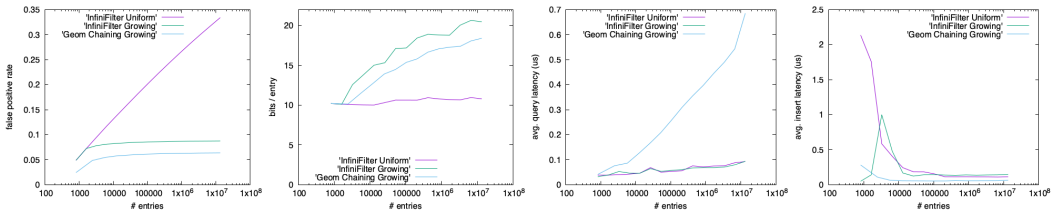


Fig. 11. Reproduction of Figure 15 (PMS=24).

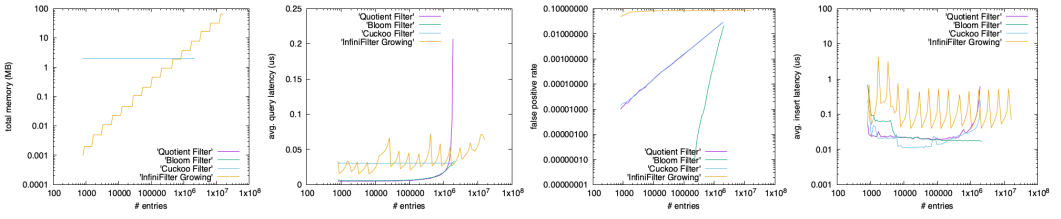


Fig. 12. Reproduction of Figure 16 (PMS=24).

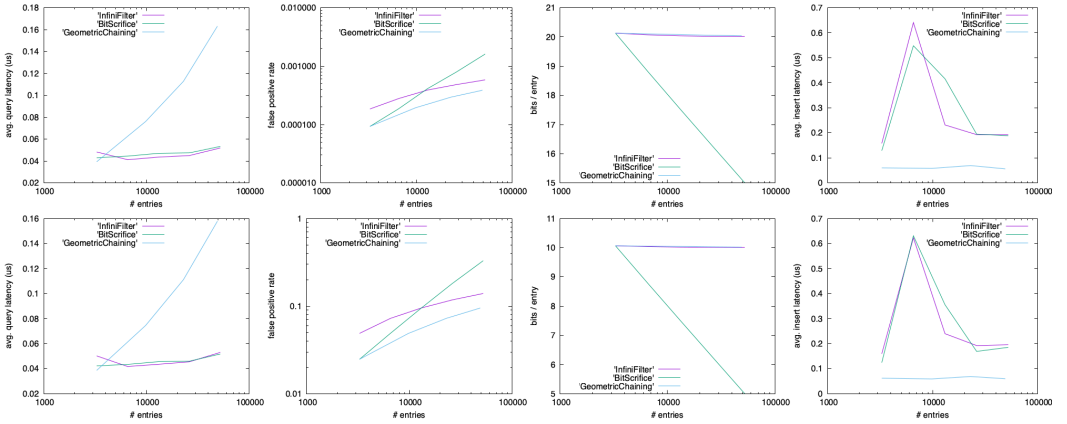


Fig. 13. Reproduction of Figure 13 (PMS=16).

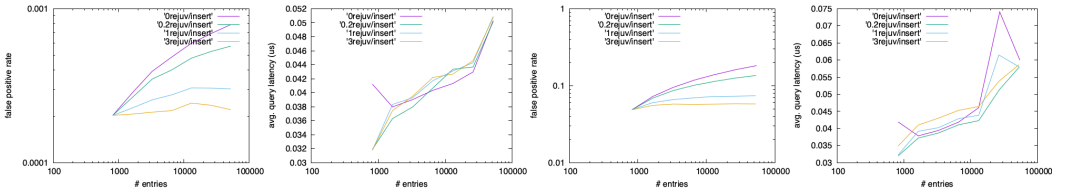


Fig. 14. Reproduction of Figure 14 (PMS=16).

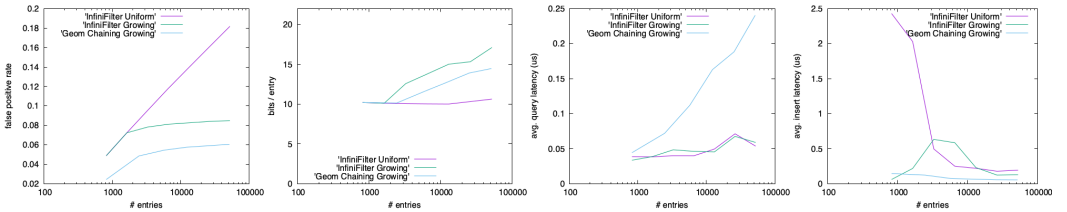


Fig. 15. Reproduction of Figure 15 (PMS=16).

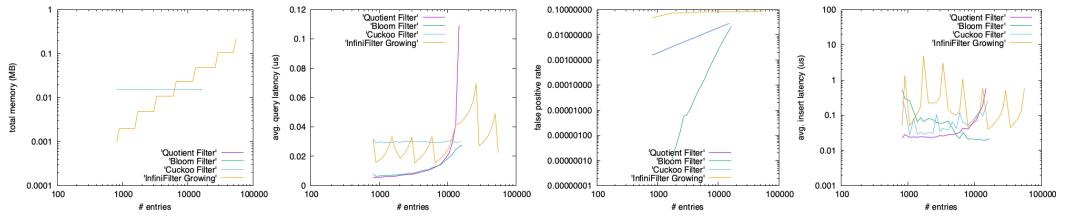


Fig. 16. Reproduction of Figure 16 (PMS=16).