

THÔNG TIN VỀ LUẬN ÁN TIẾN SĨ

1. Họ và tên nghiên cứu sinh: Lê Hoàng Bình Nguyên
2. Giới tính: Nam
3. Ngày sinh: 27/4/1993
4. Nơi sinh: Cam Lộ, Quảng Trị
5. Quyết định công nhận nghiên cứu sinh: Quyết định số 3304/QĐ-ĐHKHTN ngày 11/11/2022 của Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội.
6. Các thay đổi trong quá trình đào tạo (nếu có): Quyết định gia hạn số 3684/QĐ-ĐHKHTN ngày 06/10/2025 của Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội.
7. Tên đề tài luận án: **Phát triển thuật toán khai thác hiệu quả mẫu trọng số hữu ích phổ biến trên luồng dữ liệu**
8. Chuyên ngành: Cơ sở toán học cho tin học
9. Mã số: 9460110
10. Cán bộ hướng dẫn khoa học:

PGS.TS. Nguyễn Thị Hồng Minh

PGS.TS. Nguyễn Duy Hàm

11. Tóm tắt các kết quả mới của luận án:

11.1 Mục đích và đối tượng nghiên cứu

- Luận án tập trung nghiên cứu bài toán khai thác mẫu trọng số hữu ích phổ biến (FWUP) trên cơ sở dữ liệu (CSDL) định lượng trong bối cảnh dữ liệu ngày càng lớn, biến động theo thời gian và được sinh ra liên tục dưới dạng luồng dữ liệu. Mục tiêu chính là khắc phục những hạn chế của các phương pháp khai thác truyền thống khi xử lý dữ liệu động, đặc biệt là trường hợp dữ liệu và trọng số của các mục thay đổi theo thời gian.

- Đối tượng nghiên cứu của luận án là các phương pháp và thuật toán khai thác FWUP trên nhiều loại CSDL, đặc biệt là CSDL định lượng động và luồng dữ liệu. Đồng thời, luận án xem xét các kỹ thuật liên quan như khai thác mẫu đóng, mẫu top- k trên nhiều loại CSDL, và phân tích các phương pháp khai thác trên luồng dữ liệu với trọng tâm là mô hình cửa sổ trượt.

11.2 Các phương pháp nghiên cứu đã sử dụng

- Phương pháp nghiên cứu lý thuyết: Luận án tiến hành tổng hợp, phân tích và đánh giá có hệ thống các tài liệu khoa học trong và ngoài nước liên quan đến khai thác mẫu, đặc

biệt là khai thác FWUP. Trên cơ sở đó, đề xuất các bài toán, cấu trúc dữ liệu và thuật toán mới nhằm nâng cao hiệu quả khai thác FWUP trong môi trường dữ liệu động.

- Phương pháp nghiên cứu thực nghiệm: Dựa trên các kết quả lý thuyết, luận án tiến hành xây dựng và cài đặt các thuật toán đề xuất. Các thuật toán này được thực nghiệm trên các bộ dữ liệu chuẩn nhằm đánh giá hiệu năng theo các tiêu chí như thời gian xử lý, mức sử dụng bộ nhớ và khả năng mở rộng.

11.3 Các kết quả chính của luận án

- Luận án đã phát triển các phương pháp khai thác FWUP từ CSDL định lượng động. Khác với các nghiên cứu trước đây giả định trọng số là cố định, luận án đã xây dựng cơ chế xử lý sự biến động của trọng số theo thời gian. Trên cơ sở đó, các thuật toán dFWUT sử dụng cấu trúc Tidset và dFWUNL sử dụng cấu trúc WUNList được đề xuất nhằm giải quyết hiệu quả bài toán này. Tiếp đó, hướng nghiên cứu được mở rộng sang bài toán khai thác FWUP đóng, là dạng rút gọn nhưng không làm mất thông tin của FWUP, với thuật toán FWUCWP và các chiến lược cắt tỉa nhằm thu hẹp không gian tìm kiếm.

- Luận án đã nghiên cứu bài toán khai thác FWUP trên luồng dữ liệu định lượng động sử dụng mô hình cửa sổ trượt và đề xuất hai thuật toán FWUDS-CT và FWUDS-DWT để giải quyết bài toán này. Thuật toán FWUDS-CT sử dụng cấu trúc Circular Tidset cải tiến nhằm quản lý hiệu quả định danh giao dịch, tránh hiện tượng tràn số. Thuật toán FWUDS-DWT dựa trên cấu trúc DSWUN-tree và sự hiệu quả của cấu trúc WUNList, cho phép cập nhật dữ liệu nhanh chóng và thực hiện khai thác chỉ với một lần quét dữ liệu. Từ đó, luận án phát triển thuật toán TKFWUDS để khai thác top- k FWUP trên luồng dữ liệu định lượng động, sử dụng cơ chế nâng ngưỡng động kết hợp các chiến lược cắt tỉa, giúp loại bỏ sự phụ thuộc vào ngưỡng tối thiểu.

- Kết quả thực nghiệm trên nhiều bộ dữ liệu khác nhau cho thấy các phương pháp đề xuất đạt hiệu quả cao hơn về thời gian xử lý, sử dụng bộ nhớ và khả năng mở rộng so với các phương pháp hiện có.

12. Các hướng nghiên cứu tiếp theo:

Các hướng nghiên cứu tiềm năng bao gồm khai thác sức mạnh của các kiến trúc tính toán song song và phân tán để xử lý dữ liệu quy mô lớn; mở rộng mô hình sang các dạng dữ liệu không chắc chắn; cũng như đa dạng hóa các mô hình cửa sổ và tích hợp các kỹ thuật khai thác tăng trưởng. Những hướng đi này hứa hẹn sẽ tiếp tục hoàn thiện và nâng cao giá trị ứng dụng của bài toán này trong tương lai.

13. Các công trình công bố liên quan đến luận án:

[I] **Le N.**, Bui H., Nguyen M., Nguyen H., and Le T. (2022), “Mining frequent weighted utility patterns in dynamic quantitative databases”, *2022 RIVF International Conference on Computing and Communication Technologies (RIVF)*, pp. 536–541.

[II] Nguyen H., **Le N.**, Bui H., and Le T. (2023), “Mining frequent weighted utility patterns with dynamic weighted items from quantitative databases”, *Applied Intelligence* 53(16), pp. 19629–19646.

[III] **Lê Hoàng Bình Nguyên**, Nguyễn Duy Hàm, Nguyễn Thị Hồng Minh (2023), “Thuật toán song song khai thác nhanh các mẫu trọng số hữu ích phổ biến từ cơ sở dữ liệu định lượng động”, *Chuyên san Các công trình nghiên cứu, phát triển và ứng dụng Công nghệ Thông tin và Truyền thông* 2023(2).

[IV] **Le N.**, Nguyen H., Bui H., and Nguyen M. (2025), “The Problem Of Mining Top-K Frequent Pattern Over Dynamic Quantitative Data Stream”, *The 18th National Conference on Fundamental and Applied Information Technology Research (FAIR'2025)*, pp. 615-621.

[V] **Le N.**, Nguyen H., Nguyen M., Bui H., Yun U., and Vo B. (2026), “Efficiently mining frequent weighted utility closed patterns with pruning strategies from dynamic quantitative databases”, *The Journal of Supercomputing* 82(3), pp. 129.

[VI] **Le N.**, Nguyen H., Nguyen M., Bui H., Yun U., and Vo B. (2026), “Fast and scalable sliding-window-based algorithms for mining frequent weighted utility patterns over dynamic quantitative data streams”, *Information Sciences* 748, pp. 123488.

TM. Tập thể hướng dẫn
(Ký và ghi rõ họ tên)

Hà Nội, ngày tháng năm 2026
Nghiên cứu sinh
(Ký và ghi rõ họ tên)

PGS.TS. Nguyễn Thị Hồng Minh

Lê Hoàng Bình Nguyên

INFORMATION ON DOCTORAL THESIS

1. Full name: Le Hoang Binh Nguyen
2. Sex: Male
3. Date of birth: 27th April 1993
4. Place of birth: Cam Lo, Quang Tri
5. Admission decision number 3304/QĐ-ĐHKHTN dated 11th November 2022 by VNU University of Science, Vietnam National University, Hanoi.
6. Changes in academic process: Extension decision number 3684/QĐ-ĐHKHTN dated 6th October 2025 by VNU University of Science, Vietnam National University, Hanoi.
7. Official thesis title: Developing efficient algorithms for mining frequent weighted utility patterns over data streams
8. Major: Mathematical Foundation for Computers
9. Code: 9460110
10. Supervisors:

Assoc.Prof. Dr. Nguyen Thi Hong Minh

Assoc.Prof. Dr. Nguyen Duy Ham

11. Summary of the new findings of the thesis

11.1 Research objectives and subjects

- The thesis investigates the problem of frequent weighted utility pattern (FWUP) mining on quantitative databases within the context of increasingly large-scale, time-varying data that are continuously generated as data streams. The primary objective is to address the limitations of traditional mining methods when handling dynamic data, particularly in scenarios where item weights change over time.

- The thesis focuses on methods and algorithms for FWUP mining in various types of databases, specifically dynamic quantitative databases and data streams. Furthermore, the thesis examines related techniques such as closed pattern mining and top-k pattern mining across multiple database models, with special emphasis on stream mining techniques using the sliding window model.

11.2 Research methods

- Theoretical research methods: The thesis performs a systematic review, synthesis, analysis, and critical evaluation of both domestic and international scientific literature on

pattern mining, with a particular focus on FWUP mining. On this basis, new problem formulations, data structures, and algorithms are proposed to enhance the efficiency of FWUP mining in dynamic data environments.

- Empirical research methods: Based on the theoretical results, the thesis implements the proposed algorithms and evaluates their performance on benchmark datasets. The evaluation is conducted according to key performance criteria, including execution time, memory consumption, and scalability.

11.3 Major results

- The thesis has developed methods for mining FWUPs from dynamic quantitative databases. Unlike previous studies that assume fixed item weights, this research establishes a mechanism for handling changes in item weights over time. Based on this approach, two algorithms, dFWUT using the Tidset structure and dFWUNL using the WUNList structure, are proposed to efficiently address this problem. The research is further extended to mining frequent weighted utility closed patterns, which provide a compact, lossless representation of FWUPs, through the development of the FWUCWP algorithm, incorporating pruning strategies to reduce the search space.

- The thesis has investigated the problem of mining FWUPs from dynamic quantitative data streams using the sliding-window model and proposed two algorithms, FWUDS-CT and FWUDS-DWT, to address this problem. The FWUDS-CT algorithm employs an enhanced Circular Tidset structure to efficiently manage transaction identifiers and prevent integer overflow. The FWUDS-DWT algorithm leverages the DSWUN-tree and WUNList structures to support efficient data updates and pattern mining with a single scan of incoming data. Building upon this foundation, the thesis further develops the TKFWUDS algorithm for mining top- k FWUPs from dynamic quantitative data streams, incorporating a dynamic threshold-raising mechanism and pruning strategies to eliminate the need for a user-specified minimum threshold.

- Experimental results across various datasets demonstrate that the proposed methods achieve superior performance in terms of processing time, memory usage, and scalability compared with existing approaches.

12. Further research directions

The thesis suggests several future research directions aimed at further improving the performance and adaptability of the proposed methods. Potential directions include leveraging parallel and distributed computing architectures to process large-scale data, extending the model to handle uncertain data, exploring alternative window-based

mechanisms, and integrating incremental mining techniques. These directions hold significant promise for advancing the field and enhancing its practical applicability.

13. Thesis-related publications

[I] **Le N.**, Bui H., Nguyen M., Nguyen H., and Le T. (2022), “Mining frequent weighted utility patterns in dynamic quantitative databases”, *2022 RIVF International Conference on Computing and Communication Technologies (RIVF)*, pp. 536–541.

[II] Nguyen H., **Le N.**, Bui H., and Le T. (2023), “Mining frequent weighted utility patterns with dynamic weighted items from quantitative databases”, *Applied Intelligence* 53(16), pp. 19629–19646.

[III] **Le Hoang Binh Nguyen**, Nguyen Duy Ham, Nguyen Thi Hong Minh (2023), “A parallel algorithm for fast mining frequent weighted utility patterns from dynamic quantitative databases”, *Research, Development and Application on Information and Communication Technology* 2023(2).

[IV] **Le N.**, Nguyen H., Bui H., and Nguyen M. (2025), “The Problem Of Mining Top-K Frequent Pattern Over Dynamic Quantitative Data Stream”, *The 18th National Conference on Fundamental and Applied Information Technology Research (FAIR'2025)*, pp. 615-621.

[V] **Le N.**, Nguyen H., Nguyen M., Bui H., Yun U., and Vo B. (2026), “Efficiently mining frequent weighted utility closed patterns with pruning strategies from dynamic quantitative databases”, *The Journal of Supercomputing* 82(3), pp. 129.

[VI] **Le N.**, Nguyen H., Nguyen M., Bui H., Yun U., and Vo B. (2026), “Fast and scalable sliding-window-based algorithms for mining frequent weighted utility patterns over dynamic quantitative data streams”, *Information Sciences* 748, pp. 123488.

Date:

On behalf of academic supervisors

PhD. Student

Assoc.Prof. Dr. Nguyen Thi Hong Minh

Le Hoang Binh Nguyen