Vietnam SciMath Database Project

The Vietnam SciMath Database Project is a product of the mathematical research history initiative led by Professor Ngô Bảo Châu, a Vietnamese-French mathematician who won the Fields Medal in 2010.

Keywords: SciMath ; Vietnam mathematical research ; Ngô Bảo Châu

---

Overview

The Vietnam SciMath Database Project is a product of the mathematical research history initiative led by Professor Ngo Bao Chau (Ngô Bảo Châu), a Vietnamese-French mathematician who won the Fields Medal in 2010. This particular project received early assistance from the research team at ISR, Phenikaa University, and AISDL (Vuong & Associates), under the leadership of Dr. Vuong Quan Hoang (Vương Quân Hoàng).

Other members of the project are: Mr. La Viet Phuong (Lã Việt Phương), Prof. Le Tuan Hoa (Lê Tuấn Hoa), Assoc. Prof. Le Minh Ha (Lê Minh Hà), Dr. Trinh Thi Thuy Giang (Trịnh Thị Thúy Giang), Dr. Pham Hung Hiep (Phạm Hùng Hiệp), Ms. Nguyen Thanh Thanh Huyen (Nguyễn Thanh Thanh Huyền), Ms. Nguyen Thanh Dung (Nguyễn Thanh Dung), Ms. Nguyen Thi Linh (Nguyễn Thị Linh), Mr. Nguyen Minh Hoang (Nguyễn Minh Hoàng), and Mr. Ho Manh Toan (Hồ Mạnh Toàn)

The project A Database of Vietnam Mathematics was initiated in 2019 by the Vietnam Institute for Advanced Study in Mathematics (VIASM). In August 2019, the SciMath database was constructed. The website for data input and storage (accessed here: http://SciMath.aisdl.com) was later launched and tested on 16th December 2019. The first data lines were officially inputted into the database on 23rd December 2019.

Structure of the database

The SciMath database consists of three main systems: 1) data curation, 2) data analysis, and 3) data presentation.

- **Data curation system:** includes tools automatically collecting data of publications from scientific literature crawlers and tools for data inputting, editing, filtering, and verification.
- **Data analysis system:** includes tools for searching, extracting, and producing results for statistical reports.
- **Data presentation system:** includes a website for public use and the Application Programming Interface (API) that provides statistical information to authorized agencies or organizations.

Besides these systems, there also exists a **master database** that stores all information related to authors, affiliations, publications, sources, etc., which meet the criteria for being included in the SciMath database. Scientific databases, such as Google Scholar, zbMATH, and MathSciNet, are the main sources of data collection. For being inputted into the SciMath database, a publication must meet the following two criteria:

- At least one author of the publication is Vietnamese.
- The publication must be indexed by at least one of these databases: MathSciNet, zbMATH, Scopus, or Web of Science.

Data types

The SciMath database contains four primary groups of data: 1) article, 2) author, 3) affiliation, and 4) source/publisher.
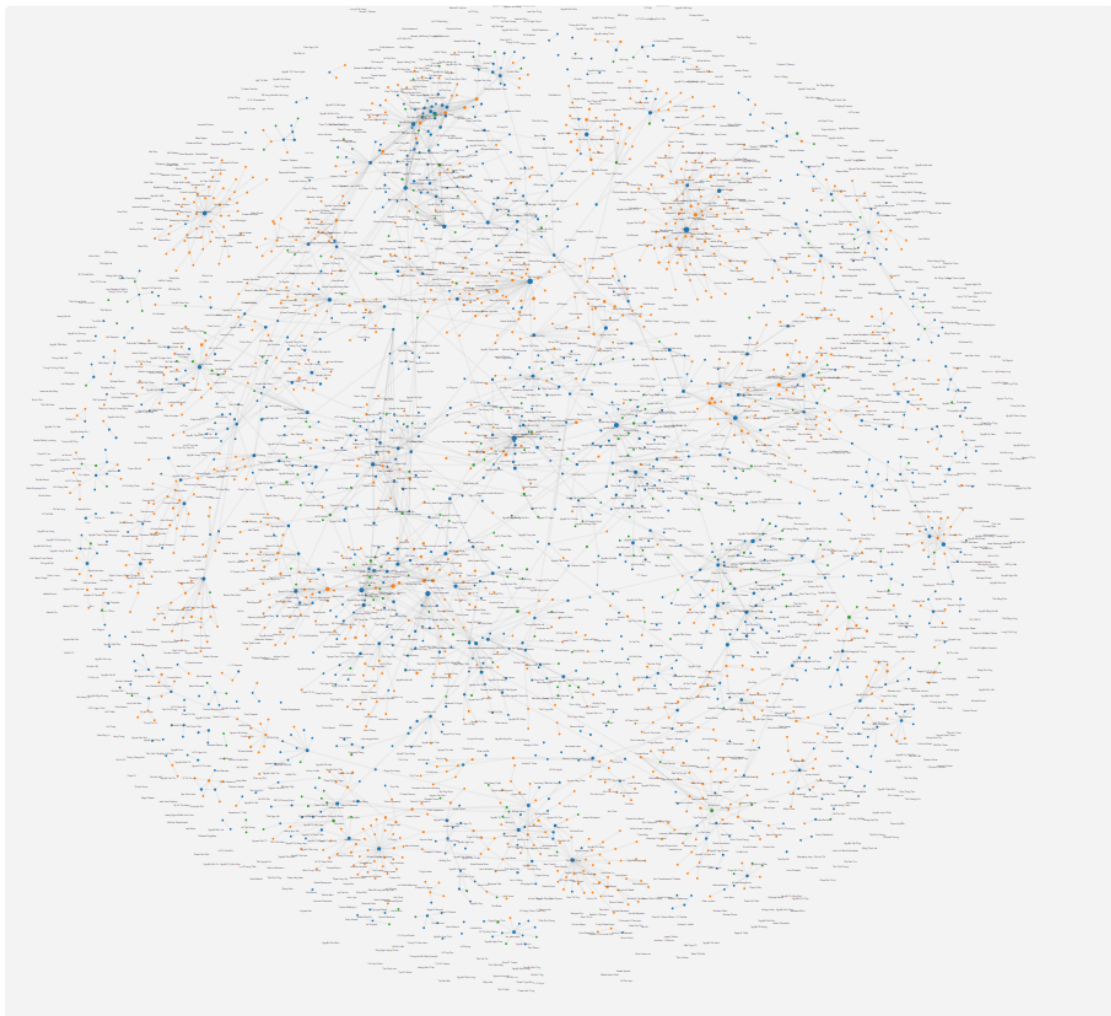
- **Article-related group:** includes information about the publication at article-level, such as title, types of publications (journal article, book, proceedings), year of publication, source, DOI/link, abstract, the keyword(s), subject(s), author(s),

and their affiliation(s). Filling information about the title, year of publication, and source of an article is mandatory during the data inputting phase.

- **Author-related group:** includes information about the author(s) of the publication, such as author's name, gender, nationality,
  affiliation, and other information such as website, phone number, and email. Each author is assigned to an ID based on their nationality, gender, and order of entry. For example, an author who is a Vietnamese woman and is the 999th author recorded into the database is given ID as vf.999 (v: Vietnamese; f: female; 999: the 999th author in the database). Similarly, the 800th author in the database, a foreign male, is given the ID as fm.800 (f: foreign; m: male; 800: the 800th author in the database). For distinguishing authors who have similar names or whose names are written in abbreviation, data collectors use information, such as website, email, and phone number.
- **Affiliation-related group:** includes information about the organization where the author(s) of the publication work(s).
- **Source/publisher-related group:** includes information about the journal/ book/ proceeding and publisher of the publication.

Project in progress

The SciMath Database was used to produce the first technical report for the meeting of the Vietnam Institute for Advanced Study in Mathematics (VIASM) on November 13, 2020 [1].

Its early results have been encouraging as the database enables researchers and the public to better appreciate the past and present contributions of generations of Vietnamese mathematicians to the development of the field. In addition, their networks and significance of works can also be observed in a context.

The project has been implemented based on the principle of cost-efficiency and cost-effectiveness[2]. It is still ongoing.

Pictures

Copyrights: Authors.

The first meeting to report the progress of the SciMath project at VIASM Headquarters (September 16, 2020)

Second meeting to report the progress of SciMath project at VIASM Headquarters (October 26, 2020)

References

1. Chau NB, Hoang VQ, Phuong LV, Hoa LT, Ha LM, Giang TTT, et al. (2020). The 80-year development of Vietnam mathematical research: Preliminary insights from the SciMath database on mathematicians, their works and their networks. arXiv preprint, arXiv:2011.09328.

2. Vuong QH. (2018). The (ir)rational consideration of the cost of science in transition economies. Nature Human Behaviour, 2(1), 5.