CS594:
Vector Databases

Introduction and Overview

Abolfazl Asudeh, asudeh@uic.edu

Aug. 2026

Instructor

Instructor:

- Abolfazl Asudeh
- https://www.cs.uic.edu/∼asudeh/
- Email address: asudeh@uic.edu
- Drop-In Office Hours (in-person): **Mondays, 1:00 PM to 3:00 PM**
- Office number: 5452 CDRLC

TA: No TA!

Course Information

About the course

CS 594: Vector Databases is a graduate special-topics course on vector databases.

This is a research-oriented course!

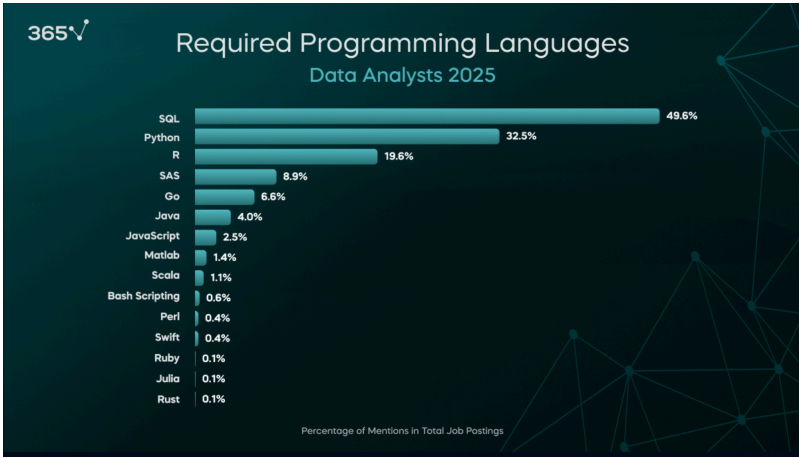
The course connects database systems, information retrieval, machine learning, and algorithms.

Why Should You Care?

49 Years of Queries + 2 years
Don Chamberlin, SIGMOD Keynote 2023

Abstract: "The relational data model, proposed by **Ted Codd** in the 1970s, has been **the dominant paradigm** for storing and accessing business data for *several decades*."

50+ years of SQL, still at the top! (365datascience.com)



Why to take the course?

The rapid rise of large language models (LLMs) and Retrieval-Augmented Generation (RAG) for tasks such as information retrieval, question answering, and knowledge-intensive reasoning, together with the increasing prevalence of large-scale unstructured data, has driven the emergence of vector database management systems. These systems are designed to extend traditional database architectures by efficiently storing, indexing, and querying vector representations such as embeddings, which are central to modern machine learning and AI-driven applications.

Why to take the course?

You will learn about information retrieval fundamentals, embedding models, vector database architectures, indexing techniques for high-dimensional data, and modern approaches for efficient query processing in vector databases.

Prerequisit?

Background in algorithms and data structures, basic database systems, and familiarity with machine learning or information retrieval is needed.

1. CS 401 (Algorithms I) or equivalent background in algorithm analysis and data structures.
2. CS 480: Database systems.
3. Prior exposure to Machine Learning, Data Mining, or Information Retrieval.

Topics – Part I: Fundamentals

1. Introduction / Fundamentals of Information Retrieval / Ranking
2. Introduction to Vector Embeddings
3. Vector Database Architectures and Challenges
4. Query Models
5. Similarity Measures
6. Performance Evaluation Measures

Topics – Part II: Algorithms and Data Structures

1. Vector indexing: Basics
2. Low-dimension Indices: Voronoi Diagrams
3. Indexing in High Dimensions
4. Local Sensitive Hashing
5. Quantized / Table-Based Indexing
6. Tree and Graph-Based Indexing
7. Filtered Search

Topics – Part III: Beyond VectorDBs

1. (maybe) Introduction to Knowledge Graphs
2. (maybe) A new VectorDB Model
3. Introduction to RAGs and Challenges
4. Vector Databases for RAG

References?

Mostly papers! (This is a research-oriented course)

- **A few reference papers**
  - Pan, James Jie, et al. Survey of Vector Database Management Systems. The VLDB Journal 33, no. 5 (2024): 1591-1615.
  - Han, Yikun, et al. ”A comprehensive survey on vector database: Storage and retrieval technique, challenge.” arXiv preprint arXiv:2310.11703 (2023)
  - Gao, Yunfan, et al. ”Retrieval-augmented generation for large language models: A survey.” arXiv preprint arXiv:2312.10997 2, no. 1 (2023)
- **Two related books**
  - Introduction to Information Retrieval: `https://nlp.stanford.edu/IR-book/information-retrieval-book.html`
  - Speech and Language Processing: `https://web.stanford.edu/~jurafsky/slp3/`

Implementation References

1. Online materials for Vector Database
2. pgvector documentations and tutorials:
   - GitHub (official documentation): https://github.com/pgvector/pgvector
   - “PostgreSQL as a Vector Database: A Pgvector Tutorial”

Related Courses

- (Princton) Algorithmic aspects of Vector DB and ANN: `https://github.com/edoliberty/vector-search-class-notes`
- (U of Torronto) Vector Databases and Unstructured Data Query Processing: `https://www.cs.toronto.edu/~koudas/csc2508_2025.html`
- (Purdue) Systems aspects of Vector Databases: `https://cs.purdue.edu/homes/csjgwang/CS592VectorDB/`

Preformance Evaluation

Grading Policy
Points Breakdown
Research Project
Coding Project

Grading Policy

Let's Discuss!
Initial thoughts:

1. **Scribe Notes** 10%
2. **Research Project Phase I:** 20%
3. **Research Project Phase II:** 30%
4. **Active Class Participation:** 10%
5. **Programming Project** 30%

- A: > 85
- B: > 70
- C: > 55
- D: > 40
- F: < 40

Academic Honesty & Late Submission Policy

Cheating will NOT be tolerated!

- **All work has to be original.**
- **Cheating = 0 points for assignment**. Possibly F in the course and **further administrative sanctions**.
- Every dishonesty **will be reported** to office of academic dishonesty

Late policy for assignments and the project:

- **-20% per day**
- No exceptions!

Use of LLMs

MUST EXPLICITLY STATE in detail how LLM was used. Anything else is cheating!

Allowed for

- Deep dive into the concepts and understand the details.
- Finding/Understanding the papers for the literature review
- Syntax level (not semantic)
  - Grammar and language errors
  - Coding syntax (not for algorithm design)

Not Allowed

- Vibe coding is NOT ALLOWED
- in any of the major activities, including (but not limited to) preparing any report or presentation (only checking for language errors is fine).

Extended Scribe Note

- You will be in charge of 1 lecture: a sign-up sheet will be sent later.
- Goal: to prepare clear technical scribe notes for the assigned course material and critically engage with the associated research papers.
- **The notes should go beyond the lecture** and explain the problem, core technical ideas, assumptions, guarantees, or design tradeoffs, and limitations rather than only summarizing the lecture.

Research Project Phase I

- **Individual:** (almost the entire grade is assigned to this part)
  - Present at least one paper that lays the foundation of their research project.
  - Provide a critical review of the work.
- **Team:** Propose your initial ideas and research direction to get feedback from students.
- **Team:**
  * Proposal:

Research Project Deliverable 1: Proposal

- Prepare and submit the refined Proposal (a 2-3 page document outlining your planned research project) based on the feedback.
- A detailed description of your proposed methodology (including datasets, algorithms, or systems you plan to use and how you will evaluate your results), and a discussion of your expected outcomes. The goal of the proposal is to demonstrate a well-scoped, feasible, and intellectually meritous plan for your term-long project, which will serve as your roadmap for the rest of the semester.

Research Project Phase II

- Expected team size: 2 (3 requires my approval)
- Evaluation based on: Technical Merit, Presentation, Proposal, and Report.

Research Project Deliverable 2: Final Report

The final project report is a comprehensive account of your completed work, structured as a 6-10 page research paper following the standard ACM Conference Proceedings format. This report must go beyond the initial proposal to include a full introduction, detailed background, a complete explanation of your methodology, a thorough presentation and analysis of your experimental results, and a discussion of their significance. You must connect your findings back to the core concepts of the course and the related work you cited. The report should tell the complete story of your project: what you set out to do, what you actually did, what you discovered, and what it all means, including an honest reflection on any limitations and potential avenues for future work. Adherence to the ACM formatting guidelines and page limit is mandatory.

Active Class Participation

- This is NOT free points that all students will receive!
- Participation is NOT enough. You should be ACTIVE!
- Evaluation:
  - Part 1: by myself. I will use my honest judgment based on how active you were in the class.
  - Part 2: peer review (most of the grade). After each Phase I presentation, you should write your summary of understanding, feedback, and research questions (hand-written), and give it to me at the end of the class. The presenting student will grade your document.

Coding Project

- A VectorDB + RAG Recommendation for Pathology (This is EXCITING! and a real problem)
- Pathology images are rich with little descriptive information. Question to explore: can existing image descriptions in a RAG setting provide reasonable descriptive information?
- Recommended Dataset: PathVQA
  - 4,998 pathology images and 32,799 question/answer pairs, derived from pathology textbooks and the Pathology Education Informational Resource (PEIR).
  - Link to dataset: `https://huggingface.co/datasets/flaviagiammarino/path-vqa`
- Embedding: I think this should be good: `https://github.com/pathologyfoundation/plip`

Thank you!