

Data-Driven Discovery of Anchor Points for PDC Content

Matthew McQuaigue¹, **Erik Saule**¹, Kalpathi Subramanian¹, Jamie Payton²
esaule@uncc.edu

¹UNC Charlotte, ²Temple University

EduHPC 2023

Towards HPC in early CS

Improving PDC education in undergrad education

- ▶ We have provided curriculum guidelines for PDC
- ▶ We have inspired ACM/IEEE CS guidelines.
- ▶ We train instructors
- ▶ We have produced Peachy assignments

We have a matching problem

- ▶ We have people who may want to teach PDC in early CS
- ▶ We have people with PDC expertise.
- ▶ Yet we are running into early CS instructors who don't know what to teach
- ▶ And PDC experts who don't know how to help

The core problem

Not every “CS1” lecture/assignment will fit in every “CS1” class

- ▶ Instructors have a lack of understanding of how to find materials that work for them.
- ▶ PDC experts have have a lack of understanding of how to develop materials that will integrate well.
- ▶ We need to understand how CS is being taught to both find materials that work and develop PDC content **with a target** of where it could be adopted.

Curriculum Guidelines Computer Science Curricula 2013

Curriculum Guidelines for
Undergraduate Degree Programs
in Computer Science

December 20, 2013

The Joint Task Force on Computing Curricula
Association for Computing Machinery (ACM)
IEEE Computer Society

memory" across all higher-level topics -- in addition to the hours allocated to related topics such as "SMD," "tasks and threads," and "synchronization." In contrast, the hours allocated to Algorithms topics represent our estimation of the effort required to achieve the desired level of competence solely within the context of Algorithms instruction. This decision reflects our recognition that many Algorithms topics develop concepts and tools that will pervade the coverage of many disparate non-Algorithms topics -- the specific list of topics varying from institution to institution. The cumulative number of hours to master a topic is, therefore, impossible to estimate in isolation.

8.2 Architecture Topics

Table 1: Architecture

Topics	II CS 11 12 13 14 15	II CS 11 12 13 14 15	Where Covered	Learning Outcome
Classes				
Economy	0	0.5	Systems	Topic: taxonomy, data vs. control, parallelism, shared-distribution memory
Data vs. control/parallelism				
Supercalculator (LFP)	0	0.25 to 1	Systems	Describe opportunities for multiple instructions issue and execution (based on level)
SMD/Venue (e.g., SDR)	0	0.1 to 0.5	Systems	Describe uses of SMD/Venue (same operations on multiple data items) e.g., accelerating graphics for games
Pipeline				
• Single vs. multistage	0	1 to 2	Systems	Describe basic pipeline process (single-stage instructions can execute at the same time); describe stages of instruction execution
• Data and control hazards			Compilers (A), Basic 2-14C	Understand how one pipe stage can depend on a result from another; for advanced branch resolution can start the wrong instructions in a

17

Map classes to curriculum

Material Form

+ Add

Title: UNOC (CS 2214) Scale-Complexity

Material type: Notes

Material visibility: Public

Version url: https://webpages.uncc.edu/isaiah/Classes/2013_01_CS2214/lecture/2013

Description:

Author: [Link to file](#)

Course: [CS 2214](#)

+ Add

+ ACM CSC 2013

+ Add

NEW-UNOC (CS 2214)

Root-ACM/IEEE Curriculum Guidelines for Undergraduate Degree Programs in Computer Science

Knowledge Area: Algorithms and Complexity

Knowledge Unit: Basic Analysis

Learning Outcome: Explain what is meant by "best," "worst," and "average" case behavior of an algorithm.

Learning Outcome: In the context of specific algorithms, identify the characteristics of data and/or other conditions or assumptions that lead to different performance.

Learning Outcome: Determine informally the time and space complexity of simple algorithms.

Learning Outcome: State the formal definition of Big O.

Learning Outcome: List and compare standard complexity classes.

Learning Outcome: Perform empirical studies to validate hypotheses about runtime behavior from mathematical analysis. Run algorithms on input data.

Learning Outcome: Describe opportunities for multiple instructions issue and execution (based on level).

Learning Outcome: Give examples that illustrate time-space trade-offs of algorithms.

Learning Outcome: Use Big O notation to give asymptotic upper bounds on time and space complexity of algorithms.

Learning Outcome: Use Big O notation to give expected case bounds on time complexity of algorithms.

Learning Outcome: Explain the use of log, sqrt, log^2, and log^3 notation to describe the amount of work done by an algorithm.

Learning Outcome: Use recurrence relations to determine the time complexity of recursively defined algorithms.

Learning Outcome: State necessary recurrence relations, e.g., using substitution method of a Master Theorem.

Topic: Differences among best, expected, and worst case behaviors of an algorithm.

Topic: Asymptotic analysis of upper and expected complexity bounds.

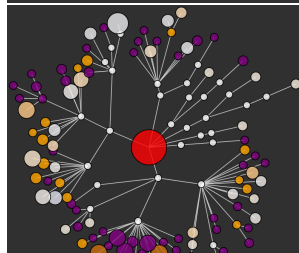
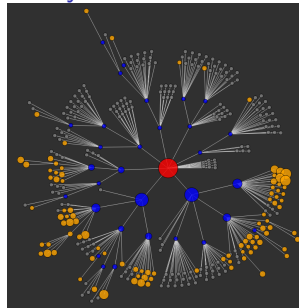
Topic: Big O notation: formal definition.

Topic: Complexity classes, such as constant, logarithmic, linear, quadratic, and exponential.

Topic: Empirical measurements of performance.

Topic: Time and space trade-offs in algorithms.

Analyze classes



Data collection

Workshops

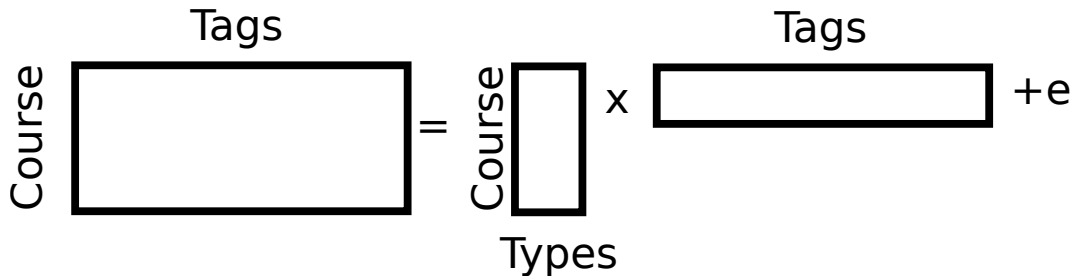
- ▶ Conducted Workshops online during COVID
- ▶ Conducted 2 workshops summer 2023 to train instructors in analyzing their courses using CS Materials
- ▶ 31 courses
- ▶ About 1700 materials (lectures, assignments, quizzes, etc.) in CS Materials system

Dataset

- ▶ Retained 20 courses in total for analysis
- ▶ Courses were tagged with course type based on course names

Discovering Types of Courses

Non Negative Matrix Factorization



By setting the number of types, you can investigate different clustering of courses.

Results and Conclusions

What we saw

- ▶ Confirmed that NNMF enables to recover the type of courses
- ▶ Identified 3 types of CS1 courses: Object Oriented, Imperative, Algorithmic Thinking
- ▶ Identified 3 types of Data Structure courses: Focus on interfaces and OOP, focus on application, cover combinatorial algorithms.

What it means for PDC in early CS courses

- ▶ Object Oriented CS1 may not be able to integrate loop-based parallelism, but may support promise style concurrency.
- ▶ Only Imperative CS1 talks about number representation, so representation based discussion of parallel reduction probably only makes sense there.
- ▶ OOP style Data structure probably can support thread safe DS discussions.
- ▶ Dependency extraction and PTG is probably easier in DS with combinatorial alg.

Thank You!

Learn more

- ▶ <https://cs-materials.herokuapp.com>
- ▶ Goncharow et al. CS-Materials: A system for classifying and analyzing pedagogical materials to improve adoption of parallel and distributed computing topics in early cs courses. JPDC 2021.
- ▶ Goncharow et al. Mapping materials to curriculum standards for design, alignment, audit, and search. SIGCSE 21.
- ▶ Goncharow et al. Classifying pedagogical material to improve adoption of parallel and distributed computing topics. EduPar 19.
- ▶ We will be running CS Materials workshops Summer 24.
- ▶ Contact us: esaule@uncc.edu

We are recruiting!

UNC Charlotte is recruiting: PhD Students, Faculty, Chaired Faculty.

Acknowledgement

This work was supported by grants from the NSF: OAC-1924057, CCF-1652442.