Vikas Kushwaha

vikask23@cse.iitk.ac.in

https://vikaskushwaha9oct.github.io

Educational Qualifications

2008 – 2013	B Tech – M Tech Dual Degree Computer Science & Engineering	Indian Institute of Technology, Kanpur	BTech : 9.0/10.0 MTech : 8.4/10.0
2024 – Present	Phd Computer Science & Engineering	Indian Institute of Technology, Kanpur	9.3/10.0

Areas of Interest

- **Phd Research Area**: Modeling Understandability, **Advisors**: Prof. Subhajit Roy, Prof. Sruti S Ragavan
- Programming Language Design, Software Design & Architecture, Artificial Intelligence
- Psychology, Culture, History & Future of Science & Civilization

Work Experience

- Vice President, Data Engineering, Goldman Sachs, Bangalore (2020-23)
- Vice President, Research & Development Engineering, Goldman Sachs, Bangalore (2017-20)
- Associate, High Impact Project Pods, Goldman Sachs, Bangalore (2015-17)
- Analyst, Data Architecture, Goldman Sachs, Bangalore (2013-15)

Research Papers

- Vikas Kushwaha, Sruti Srinivasa Ragavan, and Subhajit Roy. A measure based generalizable approach to understandability. arXiv preprint arXiv:2503.21615, 2025 (https://arxiv.org/abs/2503.21615)

Key Projects

- (**Legend**) Created a testing framework to run db-agnostic test suites against various databases. It was part of data modelling and querying engine called Legend. (2022-2023) (https://github.com/finos/legend-engine)
- (**Alchemy**) Created a framework to speedup data access by algorithmically remodelling how it is stored, and queried, based on automated analysis of their shapes. (2020-2021) (part of https://github.com/finos/legend-engine)
- (**AutoStyle**) Created a library to define multi dimensional patterns on rich text containing visual styling and positional information. The result of pattern match could be obtained as matrices and vectors, for integration with various machine learning techniques. Used it to extract semi structured data from documents like table of contents, definitions, and section headers. (2018-19)
- (**DocKnight**) Implemented novel algorithms to extract section hierarchy, paragraphs, page headers and footers, tables, key value pairs, etc from document models representing their visual structure. It was used to automate extraction of information from documents. (2017-18) (https://patents.justia.com/patent/11657101)
- (**DocKnight**) Designed and implemented a document modelling framework to parse/render documents and represent their visual structure. It supported various input formats such as pdf, scanned doc, html, etc and could render the the document model back as html/pdf. It also featured a rich API to navigate and manipulate the document structure. (2016-17) (https://github.com/goldmansachs/docknight_lib)
- (**Excelerate**) Designed and implemented a version control system for relational data, supporting various data sources, and concurrent users. It scaled horizontally and supported locking, merging, conflict resolution and milestoning. (2015-16)
- (**Data Exchange**) Designed and implemented a framework for authoring and executing data contracts in a data store agnostic manner. It was designed to scale both horizontally and vertically, and to support various data stores such as hdfs, files, relational databases, etc. for storage and execution. It could validate both batch as well as streaming data. (2014-15)
- (**MTech Thesis**) Designed a programming language to support multiple paradigms such as message passing, reactive programming, distributed computing, parallelism and concurrency, without programmer needing to deal with low level plumbing and concerns. The language core was minimalistic and consisted of four operations on top of a graph based structure. A prototype interpreter was implemented for the language. (Prof. Harish Karnick, Prof. R. K. Ghosh) (2012-13) (https://vikaskushwaha9oct.github.io/assets/pdf/MasterThesis.pdf)

Patent

- Document Information Extraction System Using Sequenced Comparators, Prabhdeep Singh Walia, Vikas Kushwaha, Granted May 23, 2023 (https://patents.justia.com/patent/11657101)