

Vaibhav KASTURIA

PERSONAL DATA

PLACE OF BIRTH: Delhi, India
DATE OF BIRTH: 18 November 1992
NATIONALITY: Indian (Permanent Resident of Germany)
EMAIL: vbh18kas@gmail.com
LINKEDIN: [vkasturia](https://www.linkedin.com/in/vkasturia)
TWITTER: [vbhvkas18](https://twitter.com/vbhvkas18)
GITHUB: [vkasturia](https://github.com/vkasturia)
SKYPE: vaibhav.kasturia1
WEBPAGE: www.vaibhavkasturia.com



EDUCATION

- 2015 - 18 Master of Science in INTERNET TECHNOLOGIES AND INFORMATION SYSTEMS (ITIS)
Leibniz University Hannover, Hannover
Major: Data and Information
Research Project: "Building & Querying Semantic Layers for Web Archives"
Thesis: "Ranking Archived Documents for Structured Queries on Semantic Layers"
GPA: 1.1 (German Scale)
- 2011 - 15 Bachelor of Engineering (Hons.) in COMPUTER SCIENCE
Birla Institute of Technology and Science - Pilani, Dubai
Thesis: "Software Development Practices at ESRI"
GPA: 1.1 (German Scale), 9.77 (Indian Scale)

WORK EXPERIENCE

- AUG 2018 - PRESENT Research Associate at UNIVERSITY OF HALLE-WITTENBERG, Halle (Saale)
Big Data Analytics, [Webis Group](#)
- Entity-Based Query Interpretation**
- **Goal:** Interpret ambiguous search engine queries to show more relevant results to the user, answer the query or help fill search engine's knowledge boxes.
 - Designed and developed an automatic approach that uses query segmentation and entity linking to identify the most reasonable interpretations of a query based on the contained entities.
 - Conducted an experimental comparison on a new corpus of 2,800 queries. It proves that my approach has better interpretation accuracy at a better run time than the previously best methods.
- Total Recall in Systematic Reviews**
- **Goal:** Find all relevant documents ("total recall") given a collection of potentially several thousands of documents somewhat related to a user-specified topic. A single systematic review may take up to 2 years without any machine-assistance.
 - Built a system that reduces the review period by ordering these documents in descending relevance.
 - Implemented several machine learning methods from an existing total recall approach ([HiCAL](#)) and tested these on botanical research datasets. The results show that machine learning reduces the human effort by almost 80 percent.
 - Development of a new algorithm that continuously adapts the feature set to the growing user feedback, and combines the current feature set with machine learning (learning-to-rank) to a ranking score.
- Argumentative Axiomatic Re-Ranking for Medical Search Queries**
- People use search engines to seek health advice online.
 - Using search engines to complete such decision making tasks, users are not able to discern authoritative from unreliable information.
 - As part of a team, we developed an axiomatic approach to re-rank search results obtained by traditional search models, in order to promote more argumentative results for medical queries.

Activities

- Prepared and took exercises for undergraduate and graduate courses (Object Oriented Programming in Java, C Programming, Search Algorithms, Foundations of Computer Science and Concepts of Modelling).
- Supervised a team of 10 students in their software project internship.
Topic: Develop a system to automatically migrate a company's old Excel-records in Excel to a database.
- Maintenance of the [Big Data Analytics](#) webpage.

MAY 2018 - JUL 2018 **Research Associate at FRAUNHOFER IAIS, Bonn**

- Prototyped a Question Answering system for an accounting firm.
- As part of the team, contributed to the development of algorithms in the area of Deep Learning as well as Speech Processing for intelligent smart car systems.
- Small contributions to the project GEISER which dealt with the analysis of spatial data.

OCT 2016 - JAN 2018 **Student Assistant at L3S RESEARCH CENTER, Hannover**

[ALEXANDRIA Project](#)

- Research on methods for the semantic and entity-based exploration of Web Archives.
- Aim of the project was to significantly advance semantic and time-based indexing for Web Archives, to efficiently index, retrieve and explore information about entities and events from the past.
- Built semantic profiles ("layers") that describe semantic information about the contents of Web Archives using Entity Linking Tools.
- Evaluated the semantic layers for complex information needs against keyword-based search systems like Google, Bing and HistDiv.
- Designed and evaluated statistical and advanced models (PageRank-like) to rank results returned by running queries on these layers.

AUG 2014 - JAN 2015 **Software Developer (Intern) at ESRI, Sharjah**

- Handled Multidimensional Geo-data (GRIB, NetCDF, HDF, etc.)
- Analyzed Raster, Mosaic and Image Service Data Layers.
- Fixed bugs and changes requested for ArcGIS 10.
- Validated UI functioning of Raster and Geo-Processing Tools of ArcGIS Pro.
- Removed potential defects (by Coverity Analysis) in Raster Solutions of ArcGIS 10.

TECHNICAL PROFICIENCY

Programming Languages:	JAVA (Advanced), Python (Intermediate), C/C++ (Intermediate)
IDE Software:	IntelliJ, Visual Studio, Eclipse, NetBeans, Jupyter Notebook
Document Preparation:	LaTeX, MS Office, Apple Office Suite
Version-Control Software:	Gitlab/Github, GitKraken, SVN, CVS
Database Systems:	Virtuoso, RocksDB, MySQL, PostgreSQL
Information Retrieval:	Apache Lucene
Natural Language Processing:	Entity Recognition, Entity Linking, Entity Disambiguation, Word Embeddings, Query Segmentation
Java Libraries:	Standard Libraries (Apache Commons, Lang, etc.), JSON/XML Parsing
Semantic Web Technologies:	RDF/RDFa, OWL, Turtle, SPARQL, Apache Jena, SPARQLWrapper
Web Technologies:	HTML5, CSS, Materialize, Bootstrap
Geo-Information Systems:	ArcGIS
Others:	Apache Maven, Weka, Apache Tika, Multithreaded Programming
Basic Knowledge:	Intel 8085 Programming, Wireshark, Scilab

ACHIEVEMENTS AND AWARDS

DEC 2018	Best Master's degree certificate for 2017/18 by Leibniz University Hannover
JUN 2017	Best Research Paper Award Nomination at JCDL 2017
2011 - 15	BITS Scholarship for Academic Excellence for the entire Bachelor's Degree

PUBLICATIONS

NOV 2019	WEBIS AT TREC 2019: DECISION TRACK A. Bondarenko, M. Fröbe, V. Kasturia, M. Völske, B. Stein and M. Hagen 28th International Text Retrieval Conference (TREC'19), Gaithersburg (Maryland, USA)
----------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

- JUN 2018 [RANKING ARCHIVED DOCUMENTS FOR STRUCTURED QUERIES ON SEMANTIC LAYERS](#)
Pavlos Fafalios, **Vaibhav Kasturia** and Wolfgang Nejdl
ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'18), Fort Worth (Texas, USA)
- NOV 2017 [BUILDING AND QUERYING SEMANTIC LAYERS FOR WEB ARCHIVES \(EXTENDED VERSION\)](#)
Pavlos Fafalios, Helge Holzmann, **Vaibhav Kasturia** and Wolfgang Nejdl
International Journal on Digital Libraries (IJDL)
- JUN 2017 [BUILDING AND QUERYING SEMANTIC LAYERS FOR WEB ARCHIVES](#)
Pavlos Fafalios, Helge Holzmann, **Vaibhav Kasturia** and Wolfgang Nejdl
ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'17), Toronto (Ontario, Canada)
- JUN 2017 [TOWARDS A RANKING MODEL FOR SEMANTIC LAYERS OVER DIGITAL ARCHIVES](#)
Pavlos Fafalios, **Vaibhav Kasturia** and Wolfgang Nejdl
ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'17), Toronto (Ontario, Canada)

TECHNICAL CERTIFICATIONS

- APR 2020 [COMPETITIVE PROGRAMMING \(CODING NINJAS\)](#)
MAR 2020 [INTRODUCTION TO THE BASH SHELL ON MAC OS AND LINUX \(PLURALSIGHT\)](#)
OCT 2019 [MASTER OBJECT ORIENTED DESIGN IN JAVA \(UDEMY\)](#)
SEP 2019 [COMPLETE PYTHON BOOTCAMP \(UDEMY\)](#)
MAY 2019 [IMPROVING DEEP NEURAL NETWORKS \(COURSERA\)](#)
MAY 2019 [STRUCTURING MACHINE LEARNING PROJECTS \(COURSERA\)](#)
MAY 2019 [NEURAL NETWORKS AND DEEP LEARNING \(COURSERA\)](#)

EXTRA-CURRICULAR ACTIVITIES

- OCT 2019 SUB-REVIEWER
ECIR 2020 Conference, Lisbon
- OCT 2019 SUB-REVIEWER
CHIIR 2020 Conference, Vancouver
- SEP 2016 STUDENT VOLUNTEER, ORGANIZING COMMITTEE
TPDL 2016 Conference, Hannover
- MAY 2016 STUDENT VOLUNTEER, ORGANIZING COMMITTEE
ACM WebSci'16 Conference, Hannover
- SEP 2012 - JUN 2013 GENERAL SECRETARY, STUDENT COUNCIL
BITS Pilani, Dubai

LANGUAGES

- ENGLISH: Native (C2) IELTS General (OCT 2019, OVERALL BAND: 8.0/9.0)
GERMAN: Advanced (C1) Goethe-Certificate C1 (JUL 2020, OVERALL SCORE: 74/100)
HINDI: Mother tongue

INTERESTS AND ACTIVITIES

Sketching, Swimming, Photography, Traveling