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
Twitter:	vbhvkas18
Github:	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

Query Understanding via Entity Linking

- **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 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