

# Philip Hausner

SOFTWARE ENGINEER · PYTHON ENTHUSIAST · INFORMATION RETRIEVAL · NATURAL LANGUAGE PROCESSING

Philipp-Otto-Runge Str. 10, 69126, Heidelberg, GERMANY

☎ (+49) 176 579 43 138 | ✉ Philip.E.Hausner@gmail.com | 🌐 philip-hausner.com | 📺 PhilipEHausner | 📷 philip-hausner

## Experience

### PROFESSIONAL

#### Quality Match

Heidelberg, DE

SOFTWARE ENGINEER

09/2021 – Now

- Designed cloud pipeline architecture to speed up processing of 60k images by 90% (from days to minutes)
- Lead developer for setting up a large scale MongoDB solution to enable a wide range of search scenarios for our customers
- Python developer in a major customer project including testing and code reviews
- Technologies:** Python, AWS (EC2, ECS, SQS), Docker, MongoDB, Go, SQL

#### Heidelberg University

Heidelberg, DE

RESEARCH ASSOCIATE

04/2020 – 09/2021

- Research project: "Dark Pattern Detection Project" at the chair of Prof. Dr. Michael Gertz
- Published at various scientific venues related to Information Retrieval and Natural Language Processing
- Developed a system to automatically detect cookie banners and investigate Dark Patterns in their context, achieving a detection rate of over 90%
- Implemented a Chrome extension to annotate web page elements
- Technologies:** Python, Selenium, Scikit-learn, BeautifulSoup, spaCy, Pandas, Pytorch, Docker, Javascript

#### Heidelberg University

Heidelberg, DE

STUDENT RESEARCH ASSISTANT

10/2016 – 03/2018

- Designed and implemented analysis tools for tropical cyclones for the use within the Waves2Weather research project
- Wrote a research paper regarding the extraction of tumble vortex core lines
- Technologies:** C++, ParaView, VTK, Eigen, CMake

### TEACHING

#### Institute of Computer Science, Heidelberg University

Heidelberg, DE

TEACHING ASSISTANT, LECTURE ASSISTANT & PROJECT SUPERVISOR

2016 – 2021

- Student project supervisor for practicals and theses, Database Systems Research Group (2020 to 2021)
- Teaching assistant: "Databases" (2017, 2019, 2020, 2021), "Text Analytics" (2020), "C++ Programming Course" (2019)
- Preparation and execution of weekly tutorial sessions in a broad field of topics as well as grading of student assignments
- Guidance of various student projects involving a variety of topics and technologies

## Education

#### Heidelberg University

Heidelberg, DE

M.SC. IN APPLIED COMPUTER SCIENCE, GERMAN GPA: 1.0 (WITH DISTINCTION), US GRADE: A

08/2018 – 03/2020

- Minor in Computational Linguistics
- Thesis: "Time-Centric Content-Exploration in Large Document Collections", Prof. Dr. Michael Gertz, grade: 1.0 (A+)
- Focus:** Natural Language Processing, Network Science

#### Kyoto University

Kyoto, JP

EXCHANGE YEAR, GRADUATE SCHOOL OF INFORMATICS

09/2017 – 04/2018

- Course selection focused on Computer Vision and Biosphere Informatics

#### Heidelberg University

Heidelberg, DE

B.SC. IN APPLIED COMPUTER SCIENCE, GERMAN GPA: 1.6, US GRADE: A/B

10/2013 – 08/2018

- Minor in Chemistry
- Thesis: "Visualization of Streamline Distributions in Uncertain 2D Vector Fields", Prof. Dr. Filip Sadlo, grade: 1.3 (A)
- Focus:** Flow Visualization, Image Processing

#### Mannheim University of Applied Sciences

Mannheim, DE

B.SC. IN BIOTECHNOLOGY, NOT FINISHED, GERMAN AVERAGE GPA: 1.9, US GRADE: B

10/2011 – 07/2013

## Skills and Competences

### PROGRAMMING LANGUAGES

- Python Advanced
- C++ Intermediate
- JavaScript Intermediate
- SQL Intermediate
- Go Basic
- Matlab Basic
- Java Basic

### TOOLS & FRAMEWORKS

- Git
- AWS (EC2, ECS, SQS)
- Docker
- MongoDB
- Linux (Debian/Ubuntu)
- NLP libraries (SpaCy, Gensim)
- Pandas, Numpy, Pytorch

### LANGUAGES

- German Native
- English Fluent
- Japanese Intermediate
- French Basic

### FIELDS OF INTEREST

- Natural Language Processing
- Information Retrieval

## Selected Projects

### Scale Existing Pipelines using AWS

Heidelberg, DE

QUALITY MATCH

10/2021 – 12/2021

- Part of a two developer team accelerated execution of pipelines by over 90% from days to a few hours
- Designed a flexible solution that enables the fast implementation of new pipeline steps for other developers
- Mainly responsible for developing the Python backend using boto3 to interact with AWS
- **Technologies:** Python, boto3, AWS (EC2, ECS, SQS), Git

### Integration of Large Scale MongoDB Solution

Heidelberg, DE

QUALITY MATCH

01/2022 – 02/2022

- Lead developer for setting up a MongoDB solution to administrate large amounts of customer data
- Collected and consolidated requirements from different parties, investigating several database solutions
- Integrated database with existing product to enable effective search of data for customers, accelerating the previous search speed by 50%
- **Technologies:** MongoDB, Python, PyMongo, FastAPI, Git

### Extracting News Articles Using Graph Embeddings

Heidelberg, DE

RESEARCH PROJECT, HEIDELBERG UNIVERSITY

01/2021 – 06/2021

- Successfully extracted a wide range of articles from 16 German news pages
- Achieved an average minimum edit distance of 51.1 characters to the ground truths texts
- Approach beats all baseline models by a significant margin
- **Technologies:** Python, Selenium, Javascript, Pytorch, Scikit-learn, Beautiful Soup

### HackerRank

WWW.HACKERRANK.COM/PHILIP\_E\_HAUSNER

Recurring

- Acquired and trained programming skills in a range of languages and difficulties
- Aimed at implementing solutions that are fast and low in memory usage

## Honors & Awards

10/2019 **Germany Scholarship**, Partially funded by Leonie-Wild Stiftung, 1 year

Heidelberg, DE

05/2019 **Best Use of Outside Data**, Datafest Germany 2019 (Hackathon), "All your players are belong to us"

Mannheim, DE

10/2018 **JASSO Scholarship**, Student Exchange Support Program, 6 months

Kyoto, JP

## Publications

### News Article Extraction Using Graph Embeddings

LWDA

P. HAUSNER, M. GERTZ

2021

### Dark Patterns in the Interaction with Cookie Banners

Workshop@CHI

P. HAUSNER, M. GERTZ

2021

### UniHD@CL-SciSumm 2020: Citation Extraction as Search

SDP@EMNLP

D. AUMILLER, S. ALMASIAN, P. HAUSNER, M. GERTZ

2020

### TiCCo: Time-Centric Content Exploration

CIKM

P. HAUSNER, D. AUMILLER, M. GERTZ

2020

### Time-centric Exploration of Court Documents

Text2Story@ECIR

P. HAUSNER, D. AUMILLER, M. GERTZ

2020

### Tumble-Vortex Core Line Extraction

SIBGRAPI WVIS

P. JUNG, P. HAUSNER, L. PILZ, J. STERN, C. EULER, M. RIEMER, F. SADLO

2017