

Dennis Aumiller

PHD STUDENT · INFORMATION RETRIEVAL · NATURAL LANGUAGE PROCESSING

Hans-Thoma-Str. 10, 69121, Heidelberg, GERMANY

☎ (+49) 176 817 47 622 | ✉ dennis.aumiller@gmx.de | 📷 dennlinger | 📱 dennlinger | 📺 dennis-aumiller | 🐦 d_aumiller

Experience

PROFESSIONAL

Heidelberg University

Heidelberg, DE

PHD IN COMPUTER SCIENCE, SUPERVISED BY PROF. DR. MICHAEL GERTZ

06/2019 – expected 06/2023

- Topic: "Context-Sensitive and Adaptive Text Exploration"
- Published at various scientific venues related to Natural Language Processing and Information Retrieval
- Special focus on multi-document representations and summarization
- Secondary focus on document retrieval and similarity embeddings with transformer models

Codefy

Heidelberg, DE

SOFTWARE ENGINEER (PART-TIME)

09/2019 – 01/2021

- Implemented product MVP for intuitive offline search, securing 200.000 € seed funding
- Conceptualized and implemented pipeline for semantic document parsing for various document types (PDF, Word, PPT,...), based on document heuristics, with focus on efficient and scalable processing
- Optimized indexing pipeline to achieve 30% faster insertion times
- Built prototype for unsupervised keyphrase extraction on legal documents
- Technologies: Python, PostgreSQL

SAP SE

Walldorf, DE

SOFTWARE ENGINEERING INTERN "PRODUCTS & INNOVATIONS"

06/2018 – 09/2018

- Responsible for optimizing Machine Learning operators for cloud-distributed frameworks by analyzing underlying algorithms
- Achieved 100x speedup over naive implementations with random algorithms and improved configuration for deployments
- Extensively evaluated different implementations across a wide tool stack
- Technologies: Docker, Kubernetes, GCP, Python, Apache Spark

Automation Laboratory, Heidelberg University

Mannheim, DE

STUDENT ASSISTANT

02/2015 – 03/2016

- Updated website to conform with corporate identity guidelines, and implemented a BiBTeX browser extension for Joomla!
- Technologies: PHP, HTML, CSS

Tut-All Software GmbH

Karlsruhe, DE

PART-TIME SOFTWARE ENGINEER

07/2014 – 01/2015

- Ported native software from Windows to Unix-based systems, involving system-level hooks on the X Window System
- Assisted with training of SVM-based OCR system for screen capturing to guide users in live tutorial sessions

REFLEXA-Werke Albrecht GmbH

Rettenbach, DE

PART-TIME SYSTEM ADMINISTRATOR

03/2013 – 03/2014

- Maintained a production ERP environment in a mid-sized company of more than 300 employees
- Updated and maintained product catalogue and production workflow for around 50 products

TEACHING

Institute of Computer Science, Heidelberg University

Heidelberg, DE

TEACHING ASSISTANT & LECTURE ASSISTANT & PROJECT SUPERVISOR

04/2016 – Present

- Student Project Supervisor, Database Systems Research Group (continuously since Summer 2019)
- Lecture Assistant: "Computer Graphics" (Winter 2016), "Databases" (Summer 2019, Summer 2020), Text Analytics (Winter 2020)
- Teaching Assistant: "Databases" (Summer 2016, Summer 2017), "Complex Network Analysis" (Winter 2018)
- Teacher for Freshman Introduction Course "Computer Science" (Winter 2016)
- Gained extensive public speaking experience by preparing and holding weekly tutorial sessions for students
- Graded weekly assignments and final exams
- Created assignments and final exams for lecture assistant positions
- Established regular code reviews and usage of git for student projects
- Technologies: Python, PostgreSQL, C++

Education

Heidelberg University

Heidelberg, DE

M.Sc. IN APPLIED COMPUTER SCIENCE, GERMAN GPA: 1.0 (EQUIV. OF 4.0 GPA; WITH DISTINCTION)

08/2017 – 05/2019

- Minor in Computational Linguistics
- Thesis Project: “Implementation of a Relational Document Hypergraph for Information Retrieval”, supervised by Prof. Dr. Michael Gertz, grade: 1.0 (A+)

University of Toronto

Toronto, CA

EXCHANGE YEAR, COMPUTER SCIENCE PROGRAM, GPA: 3.95

09/2017 – 04/2018

- Course selection focused on Machine Learning and Game Theory
- Research Project on Algorithmic Game Theory supervised by Prof. Dr. Nisarg Shah, grade: A

Heidelberg University

Heidelberg, DE

B.Sc. IN APPLIED COMPUTER SCIENCE, GERMAN GPA: 1.4 (EQUIV. OF 3.6 GPA)

10/2013 – 08/2017

- Minor in Computational Linguistics
- Thesis Project: “Mining Relationship Networks from University Websites“, supervised by Prof. Dr. Michael Gertz and Andreas Spitz, grade: 1.0 (A+)

Programming Side Projects

Stackoverflow

Heidelberg, DE

[HTTPS://STACKOVERFLOW.COM/USERS/3607203/DENNLINGER](https://stackoverflow.com/users/3607203/dennlinger)

06/2018 – Present

- Top 5% of active users every year since 2018
- Actively answering and moderating questions on various tags related to Python and Machine Learning
- Top all-time answerer for “huggingface-transformers”, top 30 all-time answerer for “PyTorch“, bronze batch in “Python“

Metaplx

Toronto, CA

AUTOMATED SEQUENCE LABELING FOR BIOMEDICAL TEXTS

02/2018 – 01/2020

- Created product MVP to assist with the automated extraction of relevant patient data from medical texts
- Managed ML part of the project, including design and implementation of a ML pipeline in PyTorch
- Built upon several state-of-the-art NLP architectures, including Seq2Seq and BERT variants
- Learned about annotation guidelines and importance of data preprocessing. Curated a dataset of approx. 70k samples
- Technologies: Python, PyTorch, React

Volunteer Activity

Comeniushaus Dormitory

Heidelberg, DE

SOCIAL TEAM MANAGER

10/2018 – 09/2020

- Responsible for organizing various student-run events
- Co-managing a team of six voluntary students

Youth Orchestra “Jugendstreichorchester inTONation e.V.“

Günzburg, DE

PRESIDENT

04/2014 – 11/2018

- Gained experience in organizing medium-sized events (ca. 200 attendees) for yearly concerts
- Learned about Management tasks by planning and discussing annual budgets

UAIG (Undergraduate Artificial Intelligence Group)

Toronto, CA

EXECUTIVE MEMEBER

01/2018 – 05/2018

- Organized a speaker series about AI targeted at first-/second-year students
- Participated in conception and sponsor acquisition for a new AI student conference (StartAI)

University of Toronto eSports Club

Toronto, CA

EXECUTIVE MEMEBER

09/2017 – 04/2018

- Hosted several offline events, including organization of location, catering and internet connection
- Partnered with game developer to host viewing party for eSports events

Honors & Awards

HACKATHONS

- 05/2019 **Best Use of Outside Data**, Datafest Germany 2019, “All your players are belong to us“ [Mannheim, DE](#)
- 04/2019 **1st Place**, Legal Tech Hackathon Universität Heidelberg, “AGBSpider“ [Heidelberg, DE](#)
- 03/2018 **1st Place**, BCB BioHacks 2018, “Human Body Atlas“ [Toronto, CA](#)
- 02/2018 **1st Place**, IEEE Ryerson Big Data Hackathon, see [Kaggle](#) [Toronto, CA](#)

SCHOLARSHIPS

- 10/2018 **Germany Scholarship**, Sponsored by Sovanta AG, For 1 year [Heidelberg, DE](#)
- 09/2017 **Baden-Württemberg Scholarship**, For 5 months [Toronto, CA](#)
- 10/2014 **Germany Scholarship**, Sponsored by SAP SE, For 2 years [Heidelberg, DE](#)

Publications

Topical Change Detection in Documents via Embeddings of Long Sequences

[arXiv Preprint](#)

D. AUMILLER*, S. ALMASIAN*, S. LACKNER, M. GERTZ

12/2020

- Segmentation of text into topically coherent segments using pretrained Transformer networks
- Introduced a cheap fine-tuning strategy, making the model easily adjustable for custom datasets

UniHD@CL-SciSumm 2020: Citation Extraction as Search

[SDP@EMNLP'20](#)

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

11/2020

- Participation in Shared Task on citation extraction from scientific papers
- Best non-neural solution since task inception
- Implemented search framework on top of BM25 + feature mix
- Compared to a neural re-ranking model based on BERT

TiCCo: Time-Centric Content Exploration

[CIKM'20](#)

P. HAUSNER*, D. AUMILLER*, M. GERTZ

10/2020

- Framework for automatically extracting timeline visualizations of large document collections
- Implementation of frontend for demonstration paper and equal contribution to manuscript

A Versatile Hypergraph Model for Document Collections

[SSDBM'20](#)

A. SPITZ, D. AUMILLER, S. BÁLINT, M. GERTZ

07/2020

- Presented a novel hypergraph representation for IR-specific applications
- Implemented and evaluated theoretical model, see [code repository](#), as well as manuscript

Time-centric Exploration of Court Documents

[Text2Story@ECIR'20](#)

P. HAUSNER, D. AUMILLER, M. GERTZ

04/2020

- Co-Supervision of underlying thesis project and help with manuscript

DNA accessibility of Chromosomes Quantified by Automated Image Analysis

[Scientific Reports](#)

M. WÜRTZ, D. AUMILLER ET AL.

09/2019

- Interdisciplinary research project with the German Cancer Research Center (DKFZ)
- Contributed an automatic image analysis toolkit for strands of nucleosomes
- Maintainer of the [code repository](#) and responsible for algorithmic descriptions

Competencies

PROGRAMMING LANGUAGES

- Python
- SQL
- Matlab
- JavaScript

TOOLS & FRAMEWORKS

- Huggingface Transformers
- PyTorch
- git
- PostgreSQL
- Apache Spark
- Docker

LANGUAGES

- English (Fluent), TOEFL iBT 115/120
- German (Native)

HOBBIES

- Viola
- Bouldering
- Running