Dennis **Aumiller**

Text Summarization · Natural Language Processing · Information Retrieval

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

☐ (+49) 176 817 47 622 | ✉ dennis.aumiller@gmx.de | ⌨ dennlinger | ▤ dennlinger | ▣ dennis-aumiller | 🐦 d_aumiller

Experience

Amazon
Berlin, DE

Applied Scientist Intern "Amazon Search", Supervised by Dr. Omar Zaidan — *08/2021 – 12/2021*

- Working on next query prediction for customer interactions on several marketplaces
- Using sequence-to-sequence networks to predict future customer interactions from their session history
- Proposed model improved coverage from around 55% to 100% of search queries across four target marketplaces

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 unsupervised semantic document parsing for diverse document types (PDF/Office)
- Optimized indexing pipeline to achieve 30% faster database ingestion times
- Built initial prototype for unsupervised keyphrase extraction on legal documents

SAP SE
Walldorf, DE

Software Engineering Intern "Products & Innovations", supervised by Dr. David Kernert — *06/2018 – 09/2018*

- Responsible for optimizing Machine Learning operators for cloud-distributed frameworks by analyzing underlying algorithms
- Achieved 100x speedup over existing operators with approximate random algorithms and improved default parameter configuration
- Extensively evaluated different implementations using scalable GCP clusters with Kubernetes and Apache Spark

Automation Laboratory, Heidelberg University
Mannheim, DE

Student Assistant — *02/2015 – 03/2016*

- Updated website to conform with corporate identity guidelines
- Implemented a BiBTeX browser extension for Joomla!

Tut-All Software GmbH
Karlsruhe, DE

Student Software Developer — *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

Education

Database Systems Research Group, 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

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", 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, Theoretical Computer Science, 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", grade: 1.0 (A+)

Teaching

Database Systems Research Group, Heidelberg University   *Heidelberg, DE*
THESIS AND PROJECT SUPERVISOR   *06/2019 – Present*

- Thesis supervisor for several Bachelor and Master's students
- Experience managing student reserach projects adjacent to my PhD topic
- Established regular code reviews and more streamlined usage of git for student projects within the group

Institute of Computer Science, Heidelberg University   *Heidelberg, DE*
LECTURE/TEACHING ASSISTANT   *06/2016 – Present*

- Lecture Assistant Graduate Courses: "Computer Graphics" (Winter 2016), Text Analytics (Winter 2020, 2022)
- Lecture Assistant Undergraduate Courses: "Databases" (Summer 2019, 2020, 2021, 2022)
- Teaching Assistant: "Databases" (Summer 2016, 2017), "Complex Network Analysis" (Winter 2018)
- Lecturer for Freshman Introduction Course "Computer Science" (Winter 2016)
- Gained xtensive public speaking experience by communicating scientific concepts to students
- Additionally designed assignments and final exams for lecture assistant positions

Programming Side Projects

Stackoverflow   *Heidelberg, DE*
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 2 user for "huggingface-transformers", top 30 for "PyTorch", top 50 for "NLP"; 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

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*

Selected Competencies

PROGRAMMING LANGUAGES
- Python
- SQL
- JavaScript
- Matlab

TOOLS & FRAMEWORKS
- Huggingface Transformers
- PyTorch
- git
- PostgreSQL

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

HOBBIES
- Viola
- Bouldering
- E-Sports
- Running

Volunteer Activity

Catsitters Heidelberg e.V.
VOLUNTEER CAT SITTER

Heidelberg, DE
02/2022 –present

- Volunteering to take care of cats during holiday seasons, thereby raising money for local animal shelters

Comeniushaus Dormitory
SOCIAL TEAM MANAGER

Heidelberg, DE
10/2018 – 09/2020

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

Youth Orchestra "Jugendstreichorchester inTONation e.V."
PRESIDENT

Günzburg, DE
04/2014 – 11/2018

- Planned and hosted annual concert with around 200 attendees
- Managing executive team of five people, as well as annual budget planning for orchestra

UAIG (Undergraduate Artificial Intelligence Group)
EXECUTIVE MEMBER

Toronto, CA
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
EXECUTIVE MEMBER

Toronto, CA
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

Publications

UniHD at TSAR-2022: Is Compute All you Need for Lexical Simplification?
D. AUMILLER, M. GERTZ

TSAR@EMNLP'22
12/2022

- 1st place submission across all three shared task languages
- Introducing a novel ensemble prompt-variation technique

EUR-Lex-Sum: A Multi-and Cross-lingual Dataset for Long-form Summarization
D. AUMILLER*, ASHISH CHOUHAN*, M. GERTZ

EMNLP'22
12/2022

- Associated code repository: `https://github.com/achouhan93/eur-lex-sum` and
  public dataset: `https://huggingface.co/datasets/dennlinger/eur-lex-sum`

Online DATEing: A Web Interface for Temporal Annotations
D. AUMILLER*, S. ALMASIAN*, D. POHL, M. GERTZ

SIGIR'22
07/2022

- Interactive web interface: `https://onlinedating.ifi.uni-heidelberg.de/`

Klexikon: A German Dataset for Joint Summarization and Simplification
D. AUMILLER, M. GERTZ

LREC'22
06/2022

- Associated code repository: `https://github.com/dennlinger/klexikon` and
  public dataset: `https://huggingface.co/datasets/dennlinger/klexikon`

Time for some German? Pre-Training a Transformer-based Temporal Tagger
S. ALMASIAN*, D. AUMILLER*, M. GERTZ

Text2Story@ECIR'22
04/2022

- Associated code repository: `https://github.com/satya77/Transformer_Temporal_Tagger`

Deep Learning und Legal Tech - Eine Bestandsaufnahme
MICHAEL GERTZ, D. AUMILLER

LTZ 01/22
02/2022

- (German) opinion piece on the usage of AI solutions in the Legal Tech space

BERT got a Date: Introducing Transformers to Temporal Tagging
S. ALMASIAN*, D. AUMILLER*, M. GERTZ

arXiv
09/2021

- Retracted due to inconsistent evaluation results, but remains relevant with novel datasets

Structural Text Segmentation of Legal Documents

D. Aumiller*, S. Almasian*, S. Lackner, M. Gertz

ICAIL'21
06/2021

- Associated code repository: `https://github.com/dennlinger/TopicalChange`

UniHD@CL-SciSumm 2020: Citation Extraction as Search

D. Aumiller*, S. Almasian*, P. Hausner*, M. Gertz

SDP@EMNLP'20
11/2020

- 3rd place solution in the shared task competition

TiCCo: Time-Centric Content Exploration

P. Hausner*, D. Aumiller*, M. Gertz

CIKM'20
10/2020

- Associated code repository: `https://github.com/PhilipEHausner/TiCCo`

A Versatile Hypergraph Model for Document Collections

A. Spitz, D. Aumiller, S. Bálint, M. Gertz

SSDBM'20
07/2020

- Assocciated code repository: `https://github.com/dennlinger/hypergraph-document-store`

Time-centric Exploration of Court Documents

P. Hausner, D. Aumiller, M. Gertz

Text2Story@ECIR'20
04/2020

- Associated code repository: `https://github.com/PhilipEHausner/TiCCo`

DNA accessibility of Chromatosomes Quantified by Automated Image Analysis

M. Würtz, D. Aumiller et al.

Scientific Reports
09/2019

- Associated code repository: `https://github.com/BMCV/quantAFM`