

Dennis Aumiller

TEXT SUMMARIZATION · NATURAL LANGUAGE PROCESSING · DATASET CURATION

1020 North Orkney Street, 19123 Philadelphia, USA

☎ (+1) 908 500 0413 | ✉ dennis.aumiller+cv@gmail.com | 📱 dennlinger | 📧 dennlinger | 🌐 dennis-aumiller | 🐦 d_aumiller

Experience

Cohere

Philadelphia, PA

MEMBER OF TECHNICAL STAFF

09/2023 – present

- Part of the Data and Evaluation efforts for Cohere's generative model

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 multiple marketplaces
- Using transformer-based networks to predict future customer interactions from their session history
- Improved long-tail performance, increasing query suggestion coverage from 55% to 100% of search queries

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

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!

Education

Data Science Group, Heidelberg University

Heidelberg, DE

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

06/2019 – expected 12/2023

- Topic: "A Unified Framework for Aspect-based Multi-document Summarization in Non-English Languages"
- Published at various scientific venues related to Natural Language Processing and Information Retrieval (see "Publications")
- Investigated limitations of current summarization systems, contributing new datasets to improve domain-specific evaluation
- Secondary focus on Natural Language Processing in non-English languages, particularly German

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

Data Science Group, Heidelberg University

Heidelberg, DE

THESIS AND PROJECT SUPERVISOR

06/2019 – 07/2023

- Thesis supervisor for several Bachelor and Master's students
- Experience managing student research projects adjacent to my PhD topic
- Established regular code reviews and mandatory pull requests for student projects within the group
- Full list of contributions online: <https://dbs.ifi.uni-heidelberg.de/team/aumiller/>

Institute of Computer Science, Heidelberg University

Heidelberg, DE

LECTURE/TEACHING ASSISTANT

06/2016 – 07/2023

- Lecture Assistant Graduate Courses: "Computer Graphics" (Winter 2016), "Text Analytics" (Winter 2020, 2022)
- Lecture Assistant Undergraduate Courses: "Databases" (Summer 2019, 2020, 2021, 2022, 2023)
- Teaching Assistant: "Databases" (Summer 2016, 2017), "Complex Network Analysis" (Winter 2018)
- Lecturer for Freshman Introduction Course "Computer Science" (Winter 2016)
- Gained extensive 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](https://stackoverflow.com/users/3607203/dennlinger)

06/2018 – Present

- Top 5% of all users, reaching more than 600k people with my answers to date
- Moderating questions on various tags related to Python and Machine Learning
- Top 2 user for the "huggingface-transformers" tag, top 30 for "pytorch", top 50 for "nlp"

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 backend for the project, including design and implementation of a training and evaluation pipeline in PyTorch
- Specific comparison of various architectures, including Seq2Seq and BERT variants
- Established consistent annotation guidelines and streamlined data pre-processing. 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
- Boulderling
- E-Sports
- Running

Volunteer Activity

Catsitters Heidelberg e.V.

VOLUNTEER CAT SITTER

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

Heidelberg, DE

02/2022 – 07/2023

Comeniushaus Dormitory

SOCIAL TEAM MANAGER

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

Heidelberg, DE

10/2018 – 09/2020

Youth Orchestra “Jugendstreicherorchester inTONation e.V.“

PRESIDENT

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

Günzburg, DE

04/2014 – 11/2018

UAIG (Undergraduate Artificial Intelligence Group)

EXECUTIVE MEMBER

- 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)

Toronto, CA

01/2018 – 05/2018

University of Toronto eSports Club

EXECUTIVE MEMBER

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

Toronto, CA

09/2017 – 04/2018

Publications

Evaluating Factual Consistency of Texts with Semantic Role Labeling

*SEM@ACL'23

J. FAN*, D. AUMILLER*, M. GERTZ

07/2023

- We propose an architecture extracting “fact tuples” from generated text segments, to investigate how factually accurate they are
- The system is statistically indistinguishable from previous SotA methods

On the State of German (Abstractive) Text Summarization

BTW'23

D. AUMILLER, J. FAN, M. GERTZ

03/2023

- We uncovered a number of inconsistencies in German summarization datasets (and subsequently, models)
- Associated code repository: <https://github.com/dennlinger/summaries>

UniHD at TSAR-2022: Is Compute All you Need for Lexical Simplification?

TSAR@EMNLP'22

D. AUMILLER, M. GERTZ

12/2022

- 1st place submission across all three shared task languages
- Introducing a novel ensemble prompt-variation technique on top of GPT-3

EUR-Lex-Sum: A Multi-and Cross-lingual Dataset for Long-form Summarization

EMNLP'22

D. AUMILLER*, ASHISH CHOUHAN*, M. GERTZ

12/2022

- Collected and filtered a multilingual resource of high-quality, human-written summaries from the European Union
- Public dataset: <https://huggingface.co/datasets/dennlinger/eur-lex-sum>

Online DATEing: A Web Interface for Temporal Annotations

SIGIR'22

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

07/2022

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

Klexikon: A German Dataset for Joint Summarization and Simplification

LREC'22

D. AUMILLER, M. GERTZ

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

Text2Story@ECIR'22

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

04/2022

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

Deep Learning und Legal Tech - Eine Bestandsaufnahme

L TZ 01/22

MICHAEL GERTZ, **D. AUMILLER**

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

arXiv

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

09/2021

- Retracted due to inconsistent evaluation results, but relevant due to introducing novel datasets and training techniques
- Associated code repository: https://github.com/satya77/Transformer_Temporal_Tagger

Structural Text Segmentation of Legal Documents

ICAIL'21

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

06/2021

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

UniHD@CL-SciSumm 2020: Citation Extraction as Search

SDP@EMNLP'20

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

11/2020

- 3rd place solution in the shared task competition

TiCCo: Time-Centric Content Exploration

CIKM'20

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

10/2020

- Demo paper introducing a generalized visual temporal retrieval framework, responsible for implementation and paper
- Associated code repository: <https://github.com/PhilipEHausner/TiCCo>

A Versatile Hypergraph Model for Document Collections

SSDBM'20

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

07/2020

- Lead contributor for the implementation of a theoretical graph retrieval framework on top of PostgreSQL
- Associated code repository: <https://github.com/dennlinger/hypergraph-document-store>

Time-centric Exploration of Court Documents

Text2Story@ECIR'20

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

04/2020

- Supervising the extension of a Master's thesis into publishable state
- Associated code repository: <https://github.com/PhilipEHausner/TiCCo>

DNA accessibility of Chromosomes Quantified by Automated Image Analysis

Scientific Reports

M. WÜRTZ, **D. AUMILLER** ET AL.

09/2019

- Primary point-of-contact for the underlying software component and description of software approach
- Associated code repository: <https://github.com/BMCV/quantAFM>

Academic Services

Reviewing Activities

Heidelberg University

UP-TO-DATE LIST AVAILABLE AT: [HTTPS://DBS.IFI.UNI-HEIDELBERG.DE/TEAM/AUMILLER/](https://DBS.IFI.UNI-HEIDELBERG.DE/TEAM/AUMILLER/)

- LIRAI@ACM Hypertext 2023 (PC member)
- Text2Story@ECIR 2023 (PC member, Recognized Reviewer Award)
- EDBT 2023 (external)
- ECIR 2022 & 2023 (external)
- BTW 2021 & 2023 (external)
- TSAR@EMNLP 2022 (shared task)
- ARR 2021 (September emergency reviewer)
- SIGMOD 2021 (external)
- SustaiNLP 2020 & 2021 (PC member)
- AI4Narratives@IJCAI-PRICAI 2020 (PC member)
- SDP@EMNLP 2020 (shared task)
- AAAI 2020 (external)
- EMNLP 2019 & 2020 (external)

Faculty Responsibility

Heidelberg University

EXAMINATION BOARD "COMPUTER SCIENCE"

10/2022-07/2023

- Reviewing exceptional student requests regarding examination regulations of the Computer Science faculty