



DAVID GRAUS

DATA SCIENTIST

+316 xxxx xxxx
david@graus.nu
www.graus.co

xxxxxxxx xx
xxxx xx, Amsterdam
The Netherlands

Birthdate: xxxx xx, 1985
Place of Birth: xxxxxx (NL)

I am results-oriented, communicative, enthusiastic, creative, outgoing, and an out-of-the-box thinker.

As a computer scientist I feel a responsibility in the current era where technology touches many aspects of our daily lives to participate in the public debate. I do so through public speaking, participating in panel discussions, and publications for the general public.

EXPERTISE

Information Retrieval, Natural Language Processing, Recommender Systems.



EXPERIENCE

Present
April 2018

DATA SCIENCE LEAD | *FD Mediagroep*, Amsterdam (NL)

I lead a team of four data scientists who bring AI-powered products and services to FD Mediagroep's channels. We work on segmentation, tagging, and personalization to offer a non-linear radio experience with audio segments that match our listener's interests (BNR *SMART Radio*). In addition, we work on abstractive personalized summarization of news articles in FD's *SMART Journalism* project.

September 2018
October 2016

DATA SCIENTIST | *Company.info (part of FD Mediagroep)*, Amsterdam (NL)

I worked on machine learning with company information. Projects included bringing an end-to-end custom-KB entity linking system into production (replacing human labeling), sentiment classification, and business sector classification.

2011
2008

EDITOR | *NTR*, Hilversum (NL)

Science editor for public broadcaster NTR, for radio and video news programs, and online science portal Wetenschap24. Work included news selection, preparing interviews, writing new items, video editing, and studio recording.

INTERNSHIPS AND VISITS

September 2015
June 2015

RESEARCH INTERN | *Microsoft Research*, Redmond, WA, USA

During my internship I worked on a large-scale analysis of Cortana user logs, resulting in a publication that received a best paper award at UMAP 2016, and the filing of a patent. Mentors: Paul N. Bennett, Ryen White, and Eric Horvitz

February 2014
January 2014

VISITING RESEARCHER | *University of Maryland*, College Park, MD, USA

I visited Prof. dr. Doug Oard's E-Discovery lab, where I followed a course on E-Discovery, and worked with the "Avocado" email corpus, which resulted in a SIGIR 2014 publication.



EDUCATION

June 2017
May 2012

PhD INFORMATION RETRIEVAL | *University of Amsterdam*, Amsterdam (NL)

PhD candidate at ILPS, a world-leading research group in Information Retrieval. My research focused on semantic search and entity retrieval, applied to digital traces. In June 2017, I defended my PhD thesis 'Entities of Interest – Discovery in Digital Traces.' Promotor: Prof. dr. Maarten de Rijke. Co-promotor: dr. Manos Tsagkias

May 2012
September 2009

MSc MEDIA TECHNOLOGY | *Leiden Institute of Advanced Computer Science*, Leiden University, Leiden (NL)

Thesis: 'Automatic Annotation of Cyttron Entries using the NCIthesaurus'

September 2008
September 2004

BA MEDIA STUDIES | *University of Amsterdam*, Amsterdam

Minor: American Studies



PROFESSIONAL ACTIVITIES

- Present
September 2018
- **COMMITTEE MEMBER** | *'Beroepenveldcommissie' School voor Journalistiek, Hogeschool Utrecht, Utrecht*
- Present
May 2017
- **CO-ORGANIZER** | *Recommender Systems Netherlands Meetup, Amsterdam*
Co-organizer of the Meetup since the 7th, host for the 8th meetup (with FD Mediagroep).
- 2015
2014
- **COMMUNICATIONS OFFICER** | *ILPS Research Group (UvA)*
To increase exposure of our research group I wrote articles of our group's publications in collaboration with the university's communication department. Some of our articles gained (inter)national press coverage on tech sites and in newspapers.
- 2015
2013
- **CHAIR** | *Informatics Institute PhD Council (UvA)*
The Ivi PhD Council promotes interaction between members of the institute's PhD community and communication between the PhD community and staff of the institute. I founded the council, held regular meetings with the institute management, and organized social and informational events for PhD candidates.
- 2014
2015
- **PUBLICITY CHAIR** | *ECIR 2014, Amsterdam, The Netherlands*
Responsible for all communication of the European Conference on Information Retrieval, i.e., social media, website, sending out calls for papers and calls for sponsors. Supervised the design of the conference materials, e.g., flyers, posters, and program booklet.
- ### TEACHING
- 2015
- **WEB SEARCH** | *MSc. Information Studies (UvA)*
Teaching assistant, responsible for lab sessions, project supervision, and grading.
- 2014
- **COMPLEX CRIME SCENES** | *MSc. Forensic Science (UvA)*
Co-lecturer, lectured on text-mining, social network analysis, and search engines. Co-responsible for lab sessions, assignments, and grading.
- 2013
- **SOCIAL NETWORK ANALYSIS** | *BSc. Information Studies (UvA)*
Teaching assistant, responsible for lab sessions and grading.
- ### PROGRAM COMMITTEE MEMBER
- 2018
- TheWebConf 2018, SIGIR 2018 (Short and Full papers)
- 2017
- SIGIR 2017 (Short and Full papers), CIKM 2017
- 2016
- SIGIR 2016 (Short Papers), DIR 2016, NewsIR 2016
- 2015
- CIKM 2015, SIGIR 2015 (Short Papers), ESAIR 2015, DIR 2015



SELECTION OF INVITED TALKS

- I enjoy speaking about my work and field of expertise. In the current era where technology touches many aspects of our daily lives, I also feel a responsibility as a computer scientist to participate in the public debate. The following talks are all aimed at the general audience.
- 2018
- **ZOEKEN, VINDEN, EN AANBEVELEN: PERSONALISATIE VS. PRIVACY** | *vogin-ip-lezing 2018: Zoeken & Vinden, Amsterdam, The Netherlands*
Lecture on recommender systems and (the need for) personalization.
- 2017
- **DE MACHT VAN DATA – HOE ALGORITMEN ONS LEVEN VORMGEVEN** | *De Balie, Amsterdam*
Opening address and participation in a panel on the impact of algorithms on our daily lives.
- 2016
- **BIG DATA & MACHINE LEARNING – MOGELIJKHEDEN & VALKUILEN** | *Koninklijke Marechaussee Intelligence Dag, Koning Willem III-kazerne, Apeldoorn (NL)*
Opening keynote on machine learning and algorithmic discrimination.
- 2015
- **ACHTER HET ALGORITME, WAT HET INTERNET VOOR JOU VERBERGT** | *Bèta Break (UvA), Amsterdam*
Participation in a panel discussion on algorithms and the filter bubble.



AWARDS

- 2016 ● **BEST STUDENT PAPER**
UMAP 2016 Best Student Paper Award for "Analyzing and Predicting Task Reminders."
- 2014 ● **ILPS MOST VALUABLE PLAYER 2014**
Award given to an ILPS member that has made significant contributions to the group.
- 2013 ● **CHALLENGE WINNER**
Doing Good by Linking Entities Challenge, WoLE Workshop at WWW 2013.



SELECTED PUBLICATIONS

PEER-REVIEWED SCHOLARLY ARTICLES

- 2018 ● David Graus, Daan Odijk, Maarten de Rijke. **The Birth of Collective Memories: Analyzing Emerging Entities in Text Streams.** *Journal of the Association for Information Science and Technology (JASIST)*
- 2016 ● David Graus, Paul N. Bennett, Ryen White, Eric Horvitz. **Analyzing and Predicting Task Reminders.** *Conference on User Modeling, Adaptation, and Personalization (UMAP)*
- 2016 ● David Graus, Manos Tsagkias, Wouter Weerkamp, Edgar Meij, Maarten de Rijke. **Dynamic Collective Entity Representations for Entity Ranking.** *International Conference on Web Search and Data Mining (WSDM)*
- 2014 ● David Graus, David van Dijk, Manos Tsagkias, Wouter Weerkamp, Maarten de Rijke. **Recipient Recommendation in Enterprises using Communication Graphs and Email Content.** *ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR)*

GENERAL PUBLIC

- 2016 ● David Graus, Maarten de Rijke. **Wij zijn racisten, daarom Google ook.** *NRC Handelsblad*
Opinion article in a daily Dutch national newspaper, on how algorithms aren't neutral, and why that is a good thing.



SOFTWARE SKILLS

- Coding/scripting ● Python, C#, LaTeX, JavaScript, Bash, HTML, CSS, SQL
- Tools/Libraries ● Elasticsearch, Git, Lucene, pandas, scikit-learn, matplotlib, NLTK, MongoDB, Redis, D3
- Software ● Adobe Creative Suite, Microsoft Office

OPEN SOURCE CONTRIBUTIONS

- 2018 ● **Textpipe** | <https://github.com/textpipe/textpipe>
Python package for converting raw text into clean, readable text, and metadata extraction.
- 2012 ● **Semanticizer** | <https://github.com/semanticize/semanticizer>
Machine learning-based entity linking framework in Python.



LANGUAGES

- Dutch ● Native language
- English ● Fluent in speaking, reading and writing, both technical and conversational
- French ● Conversationally fluent and reading and writing with high proficiency
- German ● Basic level of speaking, reading and writing



HOBBIES/INTERESTS

- I enjoy visiting **museums** and **galleries** (contemporary art and photography), **reading** (science fiction), **craft beer**, and **photography**. I like to dabble in **graphic design**, to **travel** the world, and **running**.