



+31 6xx xx xxxx  
xxxxx@xxxx.xx  
<https://graus.nu>

Date of birth: 1 July 1985  
Nationality: Dutch

# DAVID GRAUS

## DATA SCIENTIST

I am **strategic**, I quickly spot relevant patterns and issues, and separate signal from noise. I like **getting things done**: I am results-oriented, and love taking ideas from the academic world into working products. I am a **communicator**, a confident speaker and conversationalist. With a background in humanities and computer science, I like to bridge the two worlds. I bring **creative energy**: I think outside the box and come up with alternative ways to proceed.

## EXPERIENCE

- Present  
April 2018

**LEAD DATA SCIENTIST** | *FD Mediagroep, Amsterdam, NL*

I lead a team of four data scientists to bring AI-powered products to FD Mediagroep's channels. We work on the award-winning BNR SMART Radio and FD SMART Journalism projects, which revolve around personalization and summarization of radio and news. My responsibilities include hands-on work, coordinating the team, planning, setting the roadmap, and leading the research direction. With our team we delivered several products, such as a custom-built content-based recommender system, automated tagger for the journalist's CMS, and automated content-analysis framework.
- September 2018  
October 2016

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

As Company.info's first data scientist, I worked on machine learning with financial and company information. Projects included bringing an end-to-end custom-KB entity linking system into production which replaced human labeling, sentiment classification of news, and business sector classification.
- 2011  
2008

**SCIENCE EDITOR** | *NTR, Hilversum, NL*

Science editor for public broadcaster NTR, for radio 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*

At MSR I worked on a large-scale analysis of Cortana user logs, which resulted in a publication that received a best paper award, and a filed patent. I was honored to receive mentoring from 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 publication at SIGIR 2014.

## EDUCATION

- June 2017  
May 2012

**PHD INFORMATION RETRIEVAL** | *Informatics Institute, University of Amsterdam*

I did my PhD at the Information and Language Processing Systems (ILPS) group, 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** | *LIACS, Leiden Universiteit, Leiden, NL*

Thesis: 'Automatic Annotation of Cyttron Entries using the NCIthesaurus'
- September 2008  
September 2004

**BA MEDIA STUDIES** | *Faculty of Humanities, University of Amsterdam*

Minor: American Studies



## PROFESSIONAL ACTIVITIES

- Present  
September 2018 • **COMMITTEE MEMBER** | *'Beroepenveldcommissie' SvJ, Hogeschool Utrecht*  
As part of Utrecht's School of Journalism's committee that represents the industry, I provide my perspective 'from the field' on journalism education.
- Present  
January 2019 • **CO-ORGANIZER** | *DIR 2019, Amsterdam, NL*  
I co-organize DIR 2019, the Dutch Information Retrieval Workshop. DIR is a yearly conference that showcases the best work around Information Retrieval in The Netherlands. Responsibilities include everything from venue, attracting sponsors, website, program, and local organization.
- Present  
May 2017 • **CO-ORGANIZER** | *RecSys NL Meetup, The Netherlands*  
I am co-organizer of the Dutch RecSys Meetup since the 7th, host for the 8th meetup (with FD Mediagroep). This Meetup is a returning event where people from academia and industry present work on recommender systems and personalization. Responsibilities include inviting speakers and arranging venues.
- 2015  
2014 • **COMMUNICATIONS OFFICER** | *ILPS, University of Amsterdam*  
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 • **PHD COUNCIL CHAIR** | *Informatics Institute, University of Amsterdam*  
The Informatics Institute 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.
- 2015  
2014 • **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.

## PROGRAM COMMITTEE MEMBER

- 2019 • SIGIR 2019 (Short and Full papers)
- 2018 • TheWebConf 2018, SIGIR 2018 (Short and Full papers), DIR 2018
- 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



## TEACHING

### STUDENT SUPERVISION

- 2019 • **OBBERON BERLAGE**, MSc. Data Science, University of Amsterdam. Thesis internship at FD Mediagroep. Oberon successfully defended his thesis "*Improving automated segmentation of radio shows with audio embeddings*," and graduated with a 9/10.
- 2019 • **KLAUS LUX**, MSc. AI & MSc. Data Science, Radboud Universiteit Nijmegen. Internship at FD Mediagroep. Klaus worked on implementing a deep learning-based audio segmentation model.

### COURSES

- September 11, 2019 • **LECTURER** | *The ACM Summer School on Recommender Systems, Gothenburg, Sweden*  
I was invited to give a lecture and hands-on workshop at the RecSys Summerschool with Daan Odijk on "*Recommender Systems in the Media Industry – Relevance, Recency, Popularity, and Diversity*." Our hands-on session included iPython notebooks with a custom Wikinews dataset, and an assignment of implementing re-ranking strategies for related article finding.

- 2019 ● **INVITED LECTURER** | *BSc. Information Studies, University of Amsterdam*  
Collective Intelligentie. I gave a lecture on News Personalization and Ethics.
- 2015 ● **TEACHING ASSISTANT** | *MSc. Information Studies, University of Amsterdam*  
Web Search. Responsible for lab sessions, project supervision, and grading
- 2014 ● **CO-LECTURER** | *MSc. Forensic Science, University of Amsterdam*  
Complex Crime Scenes. Gave lectures on text-mining, social network analysis, and search engines. Co-responsible for lab sessions, assignments, and grading.
- 2013 ● **TEACHING ASSISTANT** | *BSc. Information Studies, University of Amsterdam*  
Social Network Analysis. Responsible for lab sessions and grading.



## SELECTED PUBLICATIONS

### GENERAL PUBLIC

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

### PEER-REVIEWED SCHOLARLY ARTICLES

- 2019 ● E. Sullivan, D. Bountouridis, J. Harambam, S. Najafian, F. Loecherbach, M. Makhortykh, D. Kelen, D. Wilkinson, **D. Graus**, N. Tintarev. **Reading News with a Purpose: Explaining User Profiles for Self-Actualization.** *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization (UMAP Adjunct)*
- 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)*

### IN THE PRESS

Below a selection of articles and media coverage about me and/or my work. For an exhaustive list, see: <https://graus.nu/press>

- June 4, 2019 ● **EUROPEAN SCIENCE-MEDIA HUB**, *A scientist's opinion: Interview with David Graus about AI-tools for journalists.* Interview in the context of a three-day Summer School AI and journalism on how we apply AI in Journalism at FD Mediagroep.
- May 10, 2019 ● **DE NATIONALE AI CURSUS**, *Geef niet het algoritme de schuld.* Interview in response to my op-ed on the alleged 'promotion' of anti-vaxx books by Bol.com's algorithms in Het Parool.
- November 26, 2018 ● **BNR**, *The case of: AI in Media.* BNR Podcast where I talk about our goals in the SMART Journalism project in the context of "applied AI" in the media industry.
- October 12, 2018 ● **DE VOLKSKRANT**, *Met deze eigen zoekmachine spit de politie schatten aan digitaal bewijs door.* Interview about the search technology developed by the Netherlands Forensic Institute, and its use in the high-profile "Mocromaffia" court cases, with my view on the role of such technology in the collection of digital evidence.
- October 2, 2018 ● **NRC**, *Waar blijft de Nederlandse robotjournalist?* Article on 'robo-journalism', where I mention our SMART Journalism project.
- July 10, 2018 ● **COMMUNICATIONS OF THE ACM**, *Dutch Police Fight Crime by Cracking PGP Phones.* Article about my work as an expert in digital forensics and search engines for Inez Weski on the "Ennetcom" case.

- November 12, 2017 ● **DENKTANK**, *Baas over brein*. Television show for youngsters on futuristic topics. In this episode: how companies influence us online through persuasion/behavior analysis.
- June 17, 2017 ● **NOS**, *David onderzocht de sporen die wij onbewust online achterlaten*. Article on my research and PhD thesis: *Entities of Interest – Discovery in Digital Traces*.

## SELECTION OF INVITED TALKS

I enjoy speaking about my work and field of expertise. I also feel a responsibility as a computer scientist to participate in the public debate and share my expertise. Below a selection of talks aimed at the general public. For an exhaustive overview of talks, see <https://graus.nu/talks>

- June 20, 2019 ● **EXPLAINABLE AI: SHOULD COMPANIES JUSTIFY ALGORITHMIC DECISIONS?** | *Academisch-Cultureel Centrum SPUI25, Amsterdam, The Netherlands*  
Talk on our "Reading News with a Purpose" work on explainable recommender systems, and participation in a panel session with prof. Marcel Worring (UvA) and Hinda Haned (Ahold).
- March 28, 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.
- June 18, 2017 ● **DE MACHT VAN DATA—HOE ALGORITMEN ONS LEVEN VORMGEVEN** | *De Balie, Amsterdam, The Netherlands*  
I gave the opening address on AI and machine learning, and participated in a panel on the impact of algorithms on our daily lives, at an evening at De Balie on the topic of "the power of data."
- November 22, 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.

## AWARDS

- 2019 ● **DUTCH INTERACTIVE AWARDS** For BNR SMART Radio we won two Dutch Interactive Awards: Gold in the category "Content," and silver in category "Disruptor."
- 2019 ● **ANNUAL MASTERS OF MEDIA AWARDS** Our news personalization efforts at FD Mediagroep won the AMMA 2019 Award in the category "Best Media Innovation."
- 2019 ● **MARCONI AWARD** Our BNR SMART Radio project won the first Marconi Online Award for innovative radio at het RadioGala van het Jaar.
- 2016 ● **BEST PAPER AWARD** UMAP 2016 Best Student Paper Award for "Analyzing and Predicting Task Reminders," the work that came out of my internship at Microsoft Research.
- 2014 ● **ILPS MVP 2014** Most Valuable Player Award given (by peer votes) 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 (TheWebConf) 2013.

## SOFTWARE SKILLS

- Coding/scripting ● Python, C#,  $\LaTeX$ , Clojure, JavaScript, bash, HTML, CSS, SQL,
- Tools/libraries ● scikit-learn, numpy, pandas, spaCy, AWS, git, elasticsearch, NLTK, MongoDB, Redis, D3
- Software ● Adobe Creative Suite, Microsoft Office

## OPEN SOURCE CONTRIBUTIONS

- present ● **TEXTPIPE** | <https://github.com/textpipe/textpipe>  
2018 Python package for converting raw text into clean, readable text, and metadata extraction. We use *textpipe* in production for SMART Journalism.

2016

**SEMANTICIZER** | <https://github.com/semanticize/semanticizer>

Machine learning-based entity linking (to Wikipedia) framework in Python. I contributed a novel global linking algorithm to the Semanticizer codebase.



## 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

● I enjoy visiting **museums** and **art 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** and **yoga**.