

Jan Zahálka

Data Scientist & CEO

+420 728 276 016

jan.zahalka@bohem.ai



Education

University of Amsterdam

Doctor (dr.) of Computer Science

May 2013 – July 2017

- Focus on multimedia analytics, i.e., the integration of interactive interfaces of visual analytics with advanced multimedia analysis methods from fields such as machine learning, computer vision, and information retrieval.
- Supervisor: prof. dr. Marcel Worring, director of Informatics Institute, University of Amsterdam.
- Fully funded, full-time research project.
- Applied research, with tangible software outputs in addition to scientific papers (Blackthorn, New Yorker Melange — see the Project section for details).
- Collaboration with Dutch forensic specialists from the Netherlands Forensic Institute, Dutch police and CLHC (Amsterdam Center for Forensic Science and Medicine).
- Thesis: *The Machine in Multimedia Analytics*.

Link: <http://dare.uva.nl/search?identifier=471f8832-69f1-481a-bd60-a056892f456a>.

Czech Technical University in Prague

Ing. (~MSc.) in Artificial Intelligence, graduated with honours

September 2010 – January 2013

- Faculty of Electrical Engineering, master programme Open Informatics
- Key topics: machine learning, optimization, advanced algorithmics, data visualization, game theory, multi-agent systems
- Scholarship for excellence granted every applicable semester.
- Thesis: *mRNA and miRNA Data Integration for Accurate Molecular Classification*.

Link: <https://cyber.felk.cvut.cz/theses/papers/313.pdf>

Delft University of Technology

Erasmus programme

February 2012 – July 2012

- Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS)
- Algorithmically-focused courses: Parallel and distributed algorithms, satisfiability and computational logic.
- Extra-field courses: technology and strategy (managerial course), Dutch language for beginners (Delftse methode),

University of Waterloo

Student exchange programme (roughly equivalent to Erasmus)

September 2011 – December 2011

- Faculty of Engineering
- Courses: Advanced topics in artificial intelligence (focus: multi-agent systems), computer security and privacy, data visualization (with focus on statistics).

Czech Technical University in Prague

Bc. (-BSc.) in Software Engineering, graduated with honours

September 2007 – June 2010

- Faculty of Electrical Engineering, bachelor programme Software Engineering and Management
- Key topics: programming language paradigms, software engineering, mathematics for computer science (linear algebra, discrete mathematics, probability and statistics), data mining.
- Scholarship for excellence granted every applicable semester.
- Thesis: *An Experimental Validation of the Occam's Razor as a Machine Learning Heuristic.*

Experience

bohem.ai

Data Scientist & CEO

May 2017 – present

- A multimedia analytics software company.
- Until December 2017, also a consultancy, in collaboration with Dutch AI companies (CliqOrange, Yellowroom, StyleCompass).

Czech Technical University in Prague

Software Analyst

June 2010 – September 2010

- Department of Cybenetics, Faculty of Electrical Engineering
- Analysis and design of SoSIReČR, a social network project (see the Projects section for details)

Deloitte

Application Analyst

December 2008 – April 2010

- Analysis, design, and development of applications improving the efficiency of the Internal Finance department of Deloitte in the Czech Republic.
- Improvement of design and coding skills, experience with professional environment in one of the leading companies in finance.

Honors & awards

2014

ACM Multimedia Grand Challenge 2014 1st Prize

Awarded for New Yorker Melange (see the Projects Section for details)

Projects

Blackthorn

Design & implementation
January 2016 – April 2017

- A fast, efficient C library for large-scale multimodal interactive learning.
- On a standard high-end workstation (16 cores, 64 GB RAM) handles one interaction round of interactive learning (i.e., training the model based on the relevant items indicated by user and producing the list of top relevant results) on 100 million images in approximately 1 second.
- Python wrapper enabling the use of Blackthorn as a Python module.

New Yorker Melange

Design & implementation
May 2014 – March 2015

- Enables broad exploration of large venue-related data (dataset of >1M images) in interactive time.
- "Give us a couple of examples of what you find a nice place, and we will tell you where people like you go in New York."
- Venue recommender, recommends users whose taste is similar to yours instead of recommending categorized venues upon query (e.g. Asian restaurants, galleries) like the major sites such as TripAdvisor do.
- Back-end written in Django, a Python Web framework, front-end visualization in D3 & HTML.
- Available here: <http://newyorkermelange.com/>

Phenotype prediction on myelodysplastic syndrome (MDS) data

Experimental design & execution
2012

- Conducted for the Intelligent Data Analysis (IDA) group of the Czech Technical University, in collaboration with the Institute of Hematology and Blood Transfusion (ÚHKT) in Prague.
- Bioinformatics project, underlying research for my master's thesis.
- Integrating matching mRNA and miRNA expression profiles of patients with/without MDS using a novel approach in order to improve phenotype prediction.

Analysis of the liver regeneration process

Experimental design & execution

2011

- Conducted for the Intelligent Data Analysis (IDA) group of the Czech Technical University, in collaboration with the Faculty of Medicine of Charles University in Hradec Králové.
- Bioinformatics project, expanding knowledge about the regenerative properties of liver.
- Identifying genes with notably elevated or inhibited activity after liver surgery based on advanced analysis of microarray data.

SoSIReČR

SW architecture design

2011

- Conducted for the Intelligent Data Analysis (IDA) group of the Czech Technical University, in collaboration with the Faculty of Mathematics and Physics of the Charles University in Prague.
- Analysis and design work on the project of novel social network for Czech IT experts from both academical and commercial spheres.
- Designing entities supported by the social network, their interaction and use cases of the system.

Skills

Areas of IT expertise

- Artificial intelligence
- Machine learning
- Multimedia analysis and analytics
- Software engineering
- High-performance computing and optimization
- Information visualization and visual analytics

Programming languages

- Python (incl. Django, numpy, scikit-learn)
- C/C++
- Java
- Matlab

- R
- SQL

Spoken languages

- Czech (native)
- English (advanced, fully professional, TOEFL computer-based test 112/120)
- Dutch (intermediate)
- German (intermediate reading and listening, beginner speaking and writing)
- Slovak (mutual intelligibility with Czech)

Publicity

- 2015 **Amsterdam Data Science magazine**
An article on New Yorker Melange (see Projects for details).
Link: http://new.amsterdamscience.org/wp-content/uploads/issue_01.pdf
(page 17)
- 2014 **Open Informatics (Czech Technical University in Prague) website**
Graduate interview for my master study programme (in Czech).
Link: <https://oi.fel.cvut.cz/cs/absolvent-jan-zahalka>

Presentations & Teaching

International conference presentations

- 2016 **ACM International Conference on Multimedia Retrieval (ICMR)**
New York, U.S.A.
Poster presentation of Interactive Multimodal Learning on 100 Million Images
(see Publications for details about this paper and the subsequent ones).
- 2015 **ACM Multimedia (MM)**
Brisbane, Australia
Oral and poster presentation of Analytic Quality: Evaluation of Performance
and Insight in Multimedia Collection Analysis.
- 2014 **IEEE Conference on Visual Analytics Science and Technology (VAST)**
Paris, France
Oral presentation of Towards Interactive, Intelligent, and Integrated
Multimedia Analytics.

2014 **ACM Multimedia (MM)**
Orlando, U.S.A.
Pitch-style presentation of New Yorker Melange: Interactive Brew of Personalized Venue Recommendations.

Workshop presentations

2016 **ICT.Open**
Amersfoort, Netherlands
Oral and poster presentation of Interactive Multimodal for Venue Recommendation.

2015 **Netherlands Conference on Computer Vision (NCCV)**
Lunteren, Netherlands
Oral and poster presentation of New Yorker Melange.

2015 **Landelijke Videodag @ Netherlands Forensic Institute**
The Hague, Netherlands
45-minute oral presentation of our multimedia analytics work for Dutch forensics, police, and security experts at the yearly security institution scientific gathering.

2014 **Amsterdam Center for Forensic Science and Medicine (CLHC)**
Amsterdam, Netherlands
Oral presentation of our multimedia analytics work at the opening event of CLHC.

Teaching

2014 – 2016 **Lab instructor, Information Visualization master course**
University of Amsterdam
Design and teaching of the lab content of the interdisciplinary infovis course for three consecutive years, 40–70 students each year.

Publications

Journal publications

- J. Zahálka, S. Rudinac, B. P. Jónsson, D. C. Koelma, and M. Worrying: *Blackthorn: Large-Scale Interactive Multimodal Learning*. IEEE Transactions on Multimedia (TMM), early access, 2017.
- M. Worrying, D. Koelma, and J. Zahálka: *Multimedia Pivot Tables for Multimedia Analytics on Image Collections*. IEEE TMM, 18(11), pp 2217–2227, November 2016.

- J. Zahálka, S. Rudinac, and M. Worrying: *Interactive Multimodal Learning for Venue Recommendation*. IEEE TMM, 17(12), pp 2235–2244, December 2015.
- J. Zahálka and F. Železný.: *An Experimental Test of Occam's Razor in Classification*. Machine Learning 82(3), pp 475–481, 2011.

Peer-reviewed publications in conference proceedings

- S. Rudinac, J. Zahálka, M. Worrying: *Discovering Geographic Regions in the City Using Social Multimedia and Open Data*. Multimedia Modelling (MMM), Lecture Notes in Computer Science (LNCS) vol. 10133, pp 148–159, January 2017.
- J. Zahálka, S. Rudinac, B. Þ. Jónsson, D. C. Koelma, and M. Worrying: *Interactive Multimodal Learning on 100 Million Images*. ACM International Conference on Multimedia Retrieval (ICMR), pp 333–337, June 2016.
- B. Þ. Jónsson, M. Worrying, J. Zahálka, L. Amsaleg, and S. Rudinac: *Ten Research Questions for Scalable Multimedia Analytics*. MMM, LNCS vol. 9517, pp. 290–302, January 2016.
- J. Zahálka, S. Rudinac, and M. Worrying: *Analytic Quality: Evaluation of Performance and Insight in Multimedia Collection Analysis*. ACM Multimedia (ACM MM), pp 231–240, October 2015.
- J. Zahálka, S. Rudinac, and M. Worrying: *New Yorker Melange: Interactive Brew of Personalized Venue Recommendations*. ACM MM, pp 205–208, November 2014.
- J. Zahálka and M. Worrying: *Towards Interactive, Intelligent, and Integrated Multimedia Analytics*. IEEE Conference on Visual Analytics Science and Technology (IEEE VAST), pp 3–12, November 2014.
- J. Kléma, J. Zahálka, M. Anděl, and Z. Krejčík: *Knowledge-Based Subtractive Integration of mRNA and miRNA Expression Profiles to Differentiate Myelodysplastic Syndrome*. BIOINFORMATICS 2014.

Scientific community service

2017	ACM Transactions on Multimedia Computing, Communications, and Applications, reviewer
2015, 2017	IEEE Transactions on Multimedia (TMM), reviewer
2015, 2017	International Workshop on Content-Based Multimedia Indexing (CBMI), reviewer
2016	ACM Multimedia (MM), demo session, reviewer

Hobbies

Computer games, hiking, piano, saxophone, European history