

# Jan Zahálka

Data Scientist & CEO

+420 728 276 016

[jan.zahalka@bohem.ai](mailto:jan.zahalka@bohem.ai)



---

## Vzdělání

### University of Amsterdam

Doctor (dr.) of Computer Science

květen 2013 – červenec 2017

- Zaměření na multimediální analytiku (multimedia analytics), tedy integraci interaktivních rozhraní visual analytics s pokročilými metodami analýzy multimediálních dat (strojové učení, počítačové vidění, získávání informací)
- Vedoucí: prof. dr. Marcel Worring, ředitel Informatics Institute, University of Amsterdam.
- Plně financovaná výzkumná pozice na plný úvazek.
- Aplikovaný výzkum s hmatatelnými softwarovými (Blackthorn, New Yorker Melange – pro více informací viz sekce Projekty) a vědeckými výsledky.
- Spolupráce s nizozemskými forenzními specialisty z Nizozemského forenzního institutu (Netherlands Forensic Institute), nizozemské policie a Amsterdamského centra forenzní vědy a medicíny (Amsterdam Center for Forensic Science and Medicine, CLHC).
- Dizertace: *The Machine in Multimedia Analytics*.

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

### České vysoké učení technické v Praze

Ing., obor Umělá inteligence, červený diplom

září 2010 – leden 2013

- Fakulta elektrotechnická (FEL), magisterský program Otevřená informatika.
- Klíčová témata: strojové učení, optimalizace, pokročilá algoritmizace, vizualizace dat, teorie her, multiagentní systémy.
- Prospěchové stipendium každý semestr, kdy jej bylo možné udělit.

- Diplomová práce: *mRNA and miRNA Data Integration for Accurate Molecular Classification*.  
Odkaz: <https://cyber.felk.cvut.cz/theses/papers/313.pdf>

## TU Delft

Program Erasmus

únor 2012 – červenec 2012

- Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS)
- Algoritmicky zaměřené předměty: Paralelní a distribuované algoritmy, splnitelnost a výpočetní logika
- Mimooborové (soft-skill) kurzy: technologie a strategie (manažersko-strategický kurz), nizozemština pro začátečníky (delftská metoda)

## University of Waterloo

Studentský výměnný program (zhruba ekvivalent Erasmu)

září 2011 – prosinec 2011

- Faculty of Engineering
- Předměty: pokročilá témata umělé inteligence (s akcentem na multiagentní systémy), počítačová bezpečnost a ochrana soukromí, vizualizace dat (s akcentem na statistiku).

## České vysoké učení technické v Praze

Bc., obor Softwarové inženýrství, červený diplom

září 2007 – červen 2010

- Fakulta elektrotechnická, bakalářský program Softwarové technologie a management
- Klíčová témata: programovací jazyky a jejich paradigmata, softwarové inženýrství, infromatická matematika (lineární algebra, diskrétní matematika, pravděpodobnost a statistika).
- Prospěchové stipendium každý semestr, kdy jej bylo možné udělit.
- Bakalářská práce: *An Experimental Validation of the Occam's Razor as a Machine Learning Heuristic*.

## Pracovní zkušenosti

### bohem.ai

Data Scientist & CEO

květen 2017 – současnost

- Software pro multimediální analytiku
- Do prosince 2017 také konzultantství, spolupráce s nizozemskými společnostmi zabývajícími se umělou inteligencí (CliqOrange, Yellowroom, StyleCompass).

### České vysoké učení technické v Praze

Softwarový analytik

červen 2010 – září 2010

- Katedra kybernetiky, Fakulta elektrotechnická
- Analýza a design SoSIReČR, projektu sociální sítě (pro více informací viz sekce Projekty)

### Deloitte

Application Analyst

prosinec 2008 – duben 2010

- Analýza, design, a implementace aplikací zlepšujících efektivitu oddělení Internal Finance v Deloittu Česká republika.
- Zlepšení programovacích a návrhářských schopností, zkušenost s profesionálním prostředím jedné z předních společností ve finančním sektoru.

## Ocenění

2014

**ACM Multimedia Grand Challenge 2014 1st Prize**

Udělena za New Yorker Melange (pro více informací viz sekce Projekty)

## Projekty

### Blackthorn

Design a implementace

leden 2016 – duben 2017

- Rychlá a efektivní knihovna v jazyce C pro interaktivní, multimodální strojové učení na velkých datech.
- Na standardním počítači vyšší kategorie (16 jader, 64 GB RAM), dokáže dokončit jedno kolo interaktivního učení (tedy natrénování modelu na základě relevantních tréninkových příkladů uvedených uživatelem a sestavení seznamu nejrelevantnějších výsledků) na 100 milionech obrázků za zhruba 1 sekundu.
- Python wrapper, který umožňuje použití Blackthornu jako modulu v jazyce Python.

### New Yorker Melange

Design a implementace

květen 2014 – březen 2015

- Umožňuje prozkoumávání velkého datového souboru dat pořízených v rozličných podnicích jako jsou asijské restaurace nebo galerie (více než 1 milion obrázků) v interaktivním čase.
- "Dejte nám pár příkladů podniků, které považujete za fajn, a my Vám řekneme, kam Newyorčani s podobným vkusem ve svém městě chodí."
- Doporučuje potenciálně zajímavé podniky na základě osobního vkusu uživatele, místo předkategorizovaných seznamů založených na celkové popularitě (jako je tomu například v případě TripAdvisoru).
- Back-end napsán v Django, webovém frameworku pro Python, front-end je napsán v D3 a HTML.
- Odkaz: <http://newyorkermelange.com/>

### Predikce fenotypu u myelodysplastického syndromu (MDS)

Experimentální design a provedení experimentů

2012

- Práce pro výzkumnou skupinu Intelligent Data Analysis (IDA) na ČVUT, ve spolupráci s Ústavem hematologie a krevní transfuze (ÚHKT) v Praze.
- Bioinformatický projekt, výzkum v rámci mé diplomové práce.
- Integrace expresních profilů mRNA a miRNA pacientů s/bez MDS za využití nového přístupu, vedoucí k přesnější predikci fenotypu.

## **Analýza procesu regenerace jater**

Experimentální design a provedení experimentů

2011

- Práce pro výzkumnou skupinu Intelligent Data Analysis (IDA) na ČVUT, ve spolupráci s Lékařskou fakultou Univerzity Karlovy v Hradci Králové.
- Bioinformatický projekt, rozšiřující znalost týkající se regenerativních vlastností jater.
- Identifikace genů s významně zvýšenou či naopak sníženou aktivitou po operaci jater založená na pokročilé analýze microarray dat.

## **SoSIReČR**

Návrh SW architektury

2011

- Práce pro výzkumnou skupinu Intelligent Data Analysis (IDA) na ČVUT, ve spolupráci s Matematicko-fyzikální fakultou Univerzity Karlovy v Praze.
- Analýza a design projektu nové sociální sítě pro české IT experty z akademických i komerčních kruhů.
- Návrh entit podporovaných touto sociální sítí, jejich interakce a use cases systému.

## **Znalosti**

### **IT - odborné znalosti**

- Umělá inteligence
- Strojové učení
- Multimediální analýza a analytika
- Softwarové inženýrství
- High-performance procesy and optimalizace
- Vizualizace informace a visual analytics

### **Programovací jazyky**

- Python (včetně Django, numpy, scikit-learn)
- C/C++
- Java
- Matlab
- R
- SQL

## Jazyky

- Čeština (mateřský jazyk)
- Angličtina (na plně profesionální úrovni, TOEFL computer-based test 112/120)
- Nizozemština (pokročilý)
- Němčina (pokročilé čtení a poslech, začátečník v mluvení a psaní)
- Slovenština (na základě vzájemné srozumitelnosti s češtinou)

## Publicita

- 2015      **Magazín Amsterdam Data Science**  
Článek o New Yorker Melange (pro více informací viz sekce Projekty).  
Odkaz: [http://new.amsterdamscience.org/wp-content/uploads/issue\\_01.pdf](http://new.amsterdamscience.org/wp-content/uploads/issue_01.pdf)  
(strana 17)
- 2014      **Web programu Otevřená informatika (ČVUT)**  
Absolventský rozhovor pro můj magisterský studijní program.  
Odkaz: <https://oi.fel.cvut.cz/cs/absolvent-jan-zahalka>

## Prezentace a výuka

### Prezentace na mezinárodních konferencích

- 2016      **ACM International Conference on Multimedia Retrieval (ICMR)**  
*New York, U.S.A.*  
Prezentace posteru Interactive Multimodal Learning on 100 Million Images  
(viz sekce Publikace pro více informací o tomto článku a člancích níže).
- 2015      **ACM Multimedia (MM)**  
*Brisbane, Austrálie*  
Ústní prezentace a prezentace posteru Analytic Quality: Evaluation of Performance and Insight in Multimedia Collection Analysis.
- 2014      **IEEE Conference on Visual Analytics Science and Technology (VAST)**  
*Paříž, Francie*  
Ústní prezentace článku Towards Interactive, Intelligent, and Integrated Multimedia Analytics.
- 2014      **ACM Multimedia (MM)**  
*Orlando, U.S.A.*  
Prezentace (pitch) projektu New Yorker Melange: Interactive Brew of Personalized Venue Recommendations.

## Prezentace na workshopech

- 2016      **ICT.Open**  
*Amersfoort, Nizozemsko*  
Ústní prezentace a prezentace posteru Interactive Multimodal for Venue Recommendation.
- 2015      **Netherlands Conference on Computer Vision (NCCV)**  
*Lunteren, Nizozemsko*  
Ústní prezentace a prezentace posteru New Yorker Melange.
- 2015      **Landelijke Videodag @ Netherlands Forensic Institute**  
*Haag, Nizozemsko*  
45minutová ústní prezentace našeho multimediálně analytického výzkumu pro nizozemské forezní, policejní a bezpečnostní odborníky na výročním vědeckém setkání nizozemských bezpečnostních institucí.
- 2014      **Amsterdam Center for Forensic Science and Medicine (CLHC)**  
*Amsterdam, Nizozemsko*  
Ústní prezentace našeho multimediálně analytického výzkumu na slavnostním otevření CLHC.

## Výuka

- 2014 – 2016      **Cvičení, magisterský předmět Information Visualization**  
*University of Amsterdam*  
Návrh a výuka cvičení mezioborového předmětu vizualizace dat tři roky po sobě, 40–70 studentů každý rok.

## Publikace

### Časopisecké články v impaktovaných časopisech

- J. Zahálka, S. Rudinac, B. P. Jónsson, D. C. Koelma, and M. Worring: *Blackthorn: Large-Scale Interactive Multimodal Learning*. IEEE Transactions on Multimedia (TMM), early access, 2017.
- M. Worring, 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. Worring: *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.

## Recenzované konferenční články

- 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. P. 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. P. 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.

## Služba vědecké komunitě

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

## Koníčky

Počítačové hry, turistika, klavír, saxofon, evropská historie