David Novak

Work address: Masaryk University, Botanicka 68a, 602 00 Brno, Czech Republic
Email: novak.david@gmail.com
Phone nr.: +420549495062
Homepage: disa.fi.muni.cz/david-novak/

Education

2004-2008
Faculty of Informatics, Masaryk University, Brno, Czech Republic
PhD Degree, dissertation topic: *Similarity Search on a Very Large Scale*
supervisor: prof. Pavel Zezula

1998-2004
Faculty of Informatics, Masaryk University, Brno, Czech Republic
Bachelor's and Master's Degree in Informatics

Experience

since 2009
researcher at Faculty of Informatics, Masaryk University, Brno, Czech Republic
Member of Laboratory of Data Intensive Systems and Applications: disa.fi.muni.cz
I have participated in more than 10 national and international research & application projects (selected):

- Center of Excellence on Multi-modal Data Interpretation on a Very Large Scale (national project)
- Searching in Large-Scale Multimedia Databases; Similarity Search with Constant Scalability; Visual Image Search on the Web (national projects from Czech Science Foundation)
- Multimedia Analysis; Efficient Search in Large-Scale Biometrical Data (Ministry of Interior of CR)
- SAPIR: Search In Audio Visual Content Using Peer-to-peer IR (EU FP6 IST-045128)
- DELOS: Network of Excellence on Digital Libraries (EU IST-507618 project)

Main research topics:

- centralized and distributed index structures for efficient and scalable similarity retrieval
- content-based image retrieval, peer-to-peer networks, subsequence matching, privacy-preserving outsourced similarity data management, advanced NoSQL databases

Selected practical results:

- M-Index, PPP-Codes: indexes for efficient similarity search in generic metric spaces
- M-Chord: distributed system for large-scale similarity search
- MESSIF: Metric Similarity Search Implementation Framework (Java library)
- system for content-based search in collection of 100 million digital images link

2003-2004
developer at Software Development Europe, Brno, Czech Republic

- project Sentinel: SSL telephony service monitoring (C++, Unix)

2002
researcher at State University of New York, Binghamton, NY, USA

- project Digital Data Embedding (Air Force Research Lab) - image steganography (C++)

Skills

Professional Interests and Experiences
algorithm design and development, similarity search, databases, NoSQL databases, graph databases, distributed data structures, management of Big Data, content-based retrieval, peer-to-peer structures

Technologies and Platforms
- good experience: Java, Bash, Linux, SQL, Infinispan, SVN, GIT, Maven, LaTeX
- experience: C++, Hadoop, Apache Giraph, XML, J2EE, Perl, JavaSript, PHP, MS Visual Studio

Languages
- Czech (mother tongue), English (fluent), Russian (communication), German (basics)

Activities

Teaching
- NoSQL Databases (preparation and teaching), 2014
- Database Systems, lectures for small groups, practices, 2005-2011
- Similarity Searching in Multimedia Data, assistent, 2010-2012

Other Research Activities
- consultant of several PhD students
- PC member: SISAP, LSDS-IR; reviewer: Elsevier IPM Jour., IEEE Trans. on MM, WISE conf.
- preparation of more than dozen research project proposals
- research fellowship in Max-Planck-Institute for Informatics, Saarbruecken, Germany, 2005

Awards
- Best Paper Award: Database and Expert Systems Applications (DEXA) 2014
- Dean's award for outstanding dissertation, 2009
- member of a group that received IBM Shared University Research Award, 2008

Personal Interests
- travelling, music, literature, sports, games

Publications

Selected publications:
- Novak, D., & Zezula, P. (2006). M-Chord: A Scalable Distributed Similarity Search Structure. In *Proceedings of InfoScale '06* (pp. 1–10). NY, USA: ACM Press. *Citations in GS: 121*
- Novak, D., Batko, M., & Zezula, P. (2009). Generic similarity search engine demonstrated by an image retrieval application. In *Proceedings ACM SIGIR '09*. New York, USA: ACM Press.
- Novak, D., Batko, M., & Zezula, P. (2011). Metric Index: An Efficient and Scalable Solution for Precise and Approximate Similarity Search. *Information Systems*, *36*(4), 721–733.
- Novak, D., & Zezula, P. (2014). Rank Aggregation of Candidate Sets for Efficient Similarity Search. In *Proceedings of DEXA 2014*. LNCS Vol. 8645, pp. 42-58. Springer. *(Best Paper Award)*

Full list of more than 25 papers: disa.fi.muni.cz/david-novak/publications/

Citation response in selected systems (including self citations, Summer 2014):
- Google Scholar: 500 citations, h-index: 11, link
- ACM Digital Library: 145 citations, link
- SCOPUS: 165 citations, link

References

prof. Pavel Zezula, Masaryk University, Brno, Czech Republic: zezula@fi.muni.cz, homepage
- supervisor of my PhD studies, head of the DISA lab

prof. Gerhard Weikum, Max-Planck-Institute for Informat., Germany: weikum@mpi-sb.mpg.de, homepage
- research fellowship in MPII, Saarbruecken, Germany; common EU project SAPIR